P-loop ATPases – Csn2 from *Streptococcus agalactiae* and the human liver ABC transporter BSEP

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Philipp Ellinger

aus

Solingen

Düsseldorf, Mai 2013

aus dem **Institut für Biochemie**

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. rer. nat. Lutz Schmitt

Koreferent: Prof. Dr. med. Ralf Kubitz

Tag der mündlichen Prüfung:

„Wer die Geduld verliert, verliert die Kraft."

Augustinus von Hippo

Contents

Abstract

Phosphate-binding loop (P-loop) containing proteins are ubiquitous proteins which fulfill diverse biochemical functions. These range from involvement in signal transduction, transport of substances across membranes to protein translation and DNA repair. The common ground of almost all these proteins is the P-loop. The P-loop (also called "Walker A" motif) is a glycine-rich loop responsible for binding the triphosphate moiety of nucleotides like ATP or GTP. Another highly conserved motif, the "Walker B" motif, is found together with the P-loop and is involved in nucleotide hydrolysis. 10-18 % of all sequenced genomes encode for P-loop harboring proteins. This doctoral thesis is dedicated to two P-loop proteins, the ATP-binding cassette (ABC) transporter bile salt export pump (BSEP, ABCB11) involved in the transport of bile salts across the membrane of hepatocytes in the human liver and Csn2, a protein involved in an inheritable, adaptive immune system of bacteria called clustered regularly interspaced short palindromic repeats/CRISPR-associated-system (CRISPR/Cas-system).

The first two chapters of this doctoral thesis deal with the Cas protein Csn2 from *Streptococcus agalactiae*. The CRISPR/Cas-system is an adaptive, bacterial immune system which relies on the integration of small fragments of foreign nucleic acids into a specific locus in the genome of bacteria, the CRISPR array. This array is subsequently transcribed and the resulting RNA is matured into small CRISPR-RNAs (crRNAs) and foreign nucleic acids are destroyed by an RNA-interference-like mechanism using the crRNAs as probe. To elucidate the function of Csn2 in the obscure integration mechanism of foreign nucleic acids into the CRISPR array, structural and biochemical techniqus were employed. First, the three-dimensional structure of Csn2, using X-ray crystallography, was solved to 2.0 Å resolution. Csn2 is a ring-shaped, stable tetramer with a conserved structural P-loop and "Walker B" motif, which are degenerated on the amino acid sequence level. Neither binding nor hydrolysis of nucleotides could be measured, but Csn2 showed a calcium-dependent binding of double-stranded DNA (dsDNA) in a sequence-unspecific manner. Further biochemical analysis revealed that dsDNA gets threaded into the inner hole of Csn2 on free DNA ends. This behavior is similar to the Ku protein involved in pro- and eukaryotic non-homologous end-joining (NHEJ) and therefore a model for the Csn2 role in foreign nucleic acid integration is proposed.

Chapters three to seven are dedicated to the human ABC transporter BSEP. This P-loop containing membrane transporter utilizes the energy of ATP hydrolysis to transport bile salts across membranes. Mutations within its gene can cause severe cholestatic diseases requiring the ultimate therapy of liver transplantation. To characterize human BSEP *in vitro*, a cloning and mutagenesis strategy for the instable *BSEP* cDNA was established. With this approach, an overexpression of human BSEP at high levels in the methylotrophic yeast *Pichia pastoris* was possible and allowed its purification in milligram quantities for the first time. For the purification, a set of over 100 different detergents was screened to identify detergents, which were able to solubilize BSEP out of the yeast membrane. BSEP was able to bind to nucleotides in detergent solution. Furthermore, a vesicular uptake assay relying on *P. pastoris* plasma membranes expressing human BSEP was established. The Michaelis-Menten constants for

bile salts were similar to published data. With the help of this method, a clinically relevant BSEP mutation, the BSEP G374S mutation could be investigated, showing a reduced transport. The mutation likely affected the substrate-binding site and the transport assay explained the patient's symptoms.

In summary, two complete different P-loop containing proteins could be characterized, showing the importance of this "small" structural element once again.

Summary in German

Phosphatbindende-Schleife (P-*Loop*) Proteine sind ubiquitär und an zahlreichen biochemischen Funktionen beteiligt. Diese reichen von Signaltransduktionsprozessen, dem Transport von Molekülen oder Ionen über biologische Membranen bis hin zur Translation von Proteinen und DNA Reparaturmechanismen. Ihre Gemeinsamkeit ist das Vorhandensein des P-*Loops*. Der P-*Loop* (auch „Walker A" Motiv genannt) ist eine Glycin-reiche Schleife, die für die Bindung des Triphosphats von Nukleotiden wie ATP oder GTP verantwortlich ist. Ein weiteres Motiv, das „Walker B" Motiv kommt gemeinsam mit dem P-*Loop* in demselben Protein vor und ist an der Hydrolyse dieser Nukleotide beteiligt. 10-18 % aller sequenzierten Genome codieren für P-*Loop* enthaltende Proteine. Diese Dissertation beschäftigt sich mit zwei P-*Loop* Proteinen, dem ATP-bindenden Kassette Transporter (ABC Transporter) <u>b</u>ile <u>s</u>alt <u>e</u>xport <u>p</u>ump (BSEP, ABCB11), der den Transport von Gallensalzen über die kanalikuläre Membran von Hepatozyten vermittelt, sowie mit Csn2. Csn2 ist Bestandteil des *<u>c</u>lustered <u>r</u>egularly <u>i</u>nterspaced <u>s</u>hort <u>p</u>alindromic <u>r</u>epeats/<u>C</u>RISPR-<u>as</u>sociated* Systems (CRISPR/Cas-System), einem bakteriellen, adaptiven Immunsystem.

Die ersten beiden Kapitel beschäftigen sich mit Csn2 aus *Streptococcus agalactiae*. Das CRISPR/Cas-System ist ein adaptives Immunsystem, das auf der Integration von kurzen, nicht-Wirt-Nukleinsäurefragmenten in einen speziellen genomischen Lokus, dem CRISPR Lokus, beruht. Dieser Lokus wird nachfolgend transkribiert und die erhaltende RNA weiter in kleine CRISPR-RNAs (crRNAs) prozessiert. Diese crRNAs dienen als Sonde zum Auffinden von fremder Nukleinsäure, welche dann in einem RNA-Interferenz ähnlichen Mechanismus zerstört wird. Csn2 ist an der Integration fremder Nukleinsäurefragmente in den CRISPR Lokus beteiligt. Der zugrunde liegende Mechanismus der Integration ist nur rudimentär verstanden. Um daher die Rolle von Csn2 im Integrationsmechanismus zu klären, wurden strukturelle und biochemische Methoden angewandt. Zunächst konnte mit Hilfe der Röntgenstrukturanalyse die dreidimensionale, räumliche Struktur von Csn2 mit einer Auflösung von 2.0 Å bestimmt werden. Csn2 besitzt eine „Ring"-ähnliche, tetramere Struktur mit einem Loch in der Mitte, sowie einen P-*Loop* und ein „Walker B" Motiv, die auf Aminosäureebene jedoch degeneriert sind. Es konnte weder die Bindung noch die Hydrolyse von Nukleotiden gezeigt werden, jedoch war Csn2 in der Lage doppelsträngige DNA Calcium-abhängig und Sequenz-unspezifisch zu binden. Weitere Analysen zeigten, dass Csn2 freie DNA Enden benötigt, die durch die freie Mitte von Csn2 eingefädelt werden können. Dieses Bindeverhalten ähnelt dem Bindeverhalten des Ku Proteins, einem Protein, das sowohl bei Pro- wie auch Eukaryonten am nicht-homologen *Endjoining* beteiligt ist. Daher wird ein ähnlicher Mechanismus für Csn2 bei der Integration von fremden Nukleinsäurefragmenten im CRISPR/Cas-System postuliert.

Kapitel drei bis sieben beschäftigen sich mit dem Leber-lokalisierten P-*Loop* ABC Transporter BSEP. BSEP nutzt die Energie der ATP-Hydrolyse zum Transport von Gallensalzen über biologische Membranen. Mutationen im *BSEP* Gen können zu schweren, cholestatischen Krankheiten führen, deren letzte Therapiemöglichkeit die Lebertransplantation ist. Um das humane BSEP *in vitro* zu

charakterisieren, wurde eine Klonierungs- und Mutagenesestrategie für die instabile cDNA von *BSEP* entwickelt und etabliert. Mit dieser Strategie war es möglich BSEP für das *Pichia pastoris* Überexpressionssystem zu klonieren und BSEP zum ersten Mal rekombinant in *P. pastoris* herzustellen. Für die Isolierung von BSEP aus *P. pastoris* Membranen wurden über 100 verschiedene Detergenzien getestet, um adequate Detergenzien für die weitere Aufreinigung zu identifizieren. In Detergenzlösung war BSEP in der Lage an Nukleotide zu binden. Somit konnten die Expression, Reinigung, sowie erste Bindestudien für humanes BSEP erfolgreich etabliert werden. Weiterhin wurde ein vesikulärer Transport-Funktionstest etabliert, der auf *P. pastoris* Plasmamembranen beruht, die BSEP enthalten. Die Michaelis-Menten Konstanten für zwei Gallensalze ähnelten den bereits publizierten Daten, die mit anderen Expressionssystemen erhalten wurden. Mit Hilfe dieses Systems, sowie der Klonierungs- und Mutagenesestrategie war es möglich eine klinisch relevante BSEP Mutation, BSEP G374S, zu charakterisieren. Der Transport von Gallensalzen durch das mutierte BSEP war drastisch reduziert, sodass die cholestatischen Symptome des Patienten mit dieser BSEP Mutation durch einen Transportdefekt zu erklären sind.

In dieser Arbeit wurden zwei Proteine charakterisiert, die als gemeinsames Motiv den P-*Loop* aufweisen, was die Bedeutung des P-*Loops* als strukturelles Element in vielen verschiedenen Proteinen mit diversen Funktionen unterstreicht.

1 Introduction

1.1 P-loop NTPases – An overview

Nucleotides such as adenosine triphosphate (ATP) or guanosine triphosphate (GTP) are protein ligands found to be important in all aspects of life ranging from the transport of molecules and ions across membranes to signal transduction, translation or DNA repair [1]. ATP is the universal energy "currency" of the cell. Molecules like glucose, which are taken up with food, get metabolized and the energy is stored in the triphosphate group of ATP [2]. Whenever an endergonic cellular process requires energy, ATP becomes hydrolyzed and the chemical potential energy stored in ATP gets liberated. The hydrolysis products of ATP, ADP and orthophosphate (P_i) are thermodynamically more stable and hydrolysis delivers a free energy that amounts to -30.5 kJ/mol [3]. A huge subset of proteins must be able to bind nucleotides like ATP or GTP to use them for their designated cellular processes. The binding of ligands to proteins is a crucial step for many aspects of cellular physiology. Molecular recognition is strictly dependent on the three-dimensional arrangement of the protein's ligand-binding site as well as the ligand itself. Distinct protein binding motifs have evolved to accurately bind nucleotides like ATP or GTP [4]. A highly conserved motif, in amino acid sequence and structure is the phosphate-binding loop, also called P-loop, which is the most abundant nucleoside triphosphate (NTP) binding motif found in eu- and prokaryotes as well as in archaea. [4]. Almost 10 – 18 % of every sequenced genome to date encodes proteins containing a P-loop [5]. The P-loop was originally identified by Sir John Walker and colleagues in different proteins in 1982 and is therefore also named the "Walker A" motif [6]. Its consensus sequence is referred to as GxxxxGK[S/T] (where x can be any amino acid), but can also slightly deviate [7]. Structurally, the P-loop has been found to be a glycine-rich loop. It accurately positions the triphosphate, more precisely the β- and γ-phosphate group of the nucleotide by interacting via hydrogen bonds of the main chain hydrogen atoms with the phosphate groups [8]. The lysine residues (K) are thought to neutralize the negative charge of the phosphate moiety and positions the nucleotide in a higher-energy conformation [9]. The following serine or threonine (S/T) participates directly or indirectly in the binding of a magnesium ion. P-loop containing

proteins also contain another highly conserved motif, the "Walker B" motif [6]. The Walker B motif consists of a conserved aspartate or glutamate, which is preceded by four hydrophobic amino acids (ΦΦΦD/E, where Φ can be any hydrophobic amino acid). The function of the Walker B motif is to bind a water-bridged magnesium ion, which is further coordinated by the β- and γ-phosphates of the nucleotide [9]. Typically, P-loop NTPases hydrolyze the triphosphate nucleoside to a diphosphate nucleoside by attacking the β-γ phosphodiester bond. P-loop containing proteins are α/β sandwich proteins (Figure 1). They are three layered, which denotes that they exhibit a parallel β-sheet flanked on both sides by α-helices. The P-loop forms a loop between a β-strand and an α-helix, whereas the Walker B motif is located C-terminally to the P-loop and forms a β-strand [10].
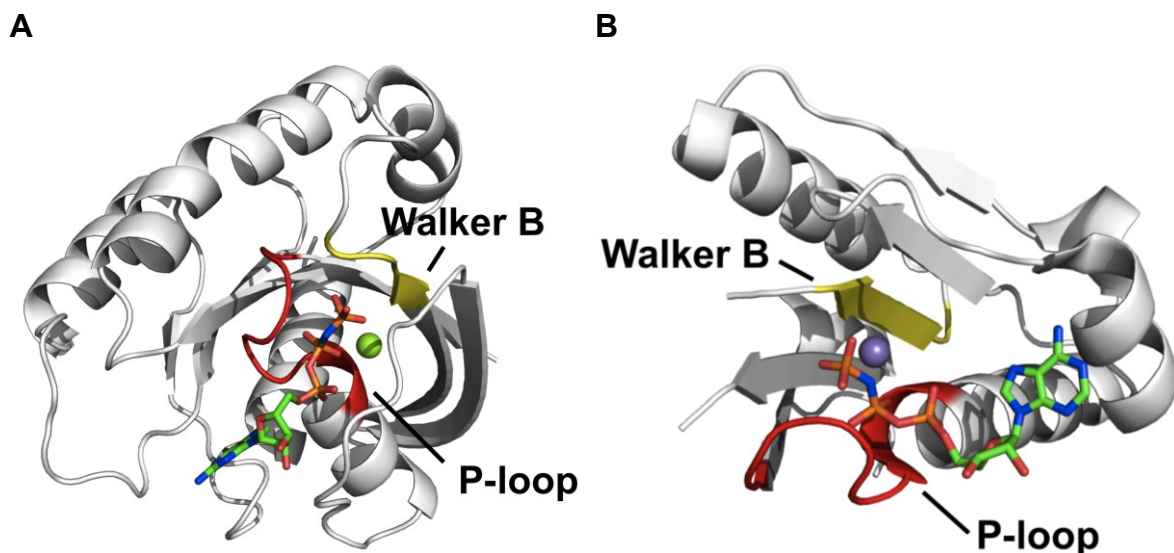
Figure. 1: General overview of two P-loop containing proteins.
The P-loop (Walker A motif) is shown in red, the Walker B motif in yellow, a magnesium ion as green sphere, a manganese ion as violet sphere and nucleotide analogs (phosphoaminophosphonic acid-guanylate ester (GNP) in A and phosphoaminophosphonic acid-adenylate ester (ANP) in B) are shown as sticks. **A)** The GTPase H-ras (PDB code: 5P21) is shown, which belongs to group I with direct vicinity of the P-loop and the Walker B motif. **B)** RecA from *Escherichia coli* (PDB code: 1XMS) belongs to group II and the P-loop and the Walker B motif are separated by an additional β-strand . For the sake of clarity only the P-loop containing domain of RecA is shown.

P-loop NTPases have been classified into two different groups [11,12]. Group I, the "KG"-group (kinase-GTPase-group), is defined by the direct neighborhood of the P-loop and the Walker B motif (Figure 1A). Group II is the "ASCE" class (additional strand, catalytic E) and contains other P-loop NTPases which are characterized by an additional β-strand between the two Walker motifs (Figure 1B).

Group II comprises the AAA+ helicases, ABC proteins, SF1/2 helicases, and RecA/F1 superfamilies of ATPases and they depend on a conserved glutamate in the Walker B motif for hydrolysis which acts as a base [11,12,13]. The widespread occurrence of P-loop proteins in cellular physiology is manifested in their vast diversity of function. These functions include nucleic acid metabolism by e.g. helicases or recombinatorial proteins as well as the transport across membranes by e.g. ABC transporters. This doctoral thesis characterizes two different P-loop containing proteins essential in the bile salt transport across a membrane in vertrebartes as well as in a nucleic acid based bacterial immunity system.

1.2 Clustered regularly interspaced short palindromic repeats (CRISPRs) – Bacterial immunity against foreign genetic elements

The first two chapters of this doctoral thesis deal with the P-loop containing protein Csn2 found in a recently discovered system of prokaryotic adaptive immunity called the CRISPR/Cas-system. This system is able to equip a prokaryotic organism with a set of small DNA fragments of viral or plasmid origin to fight invading foreign DNA of the same origin by an RNA interference-like mechanism. The system depends on two different levels, the DNA level, which is represented by clustered regularly interspaced short palindromic repeats (CRISPRs), as well as on the protein level by Cas proteins (CRISPR-associated proteins), which bind and cleave nucleic acids. How these different levels cooperate to achieve the survival of the prokaryote is described in more detail below.

1.2.1 A brief history of the CRISPR/Cas-system

Microorganisms have evolved several strategies to survive in the fight against viruses. These include blocking the adsorption of phages to host cell receptors by masking or mutating them or blocking DNA injection into the host cell by proteins [14,15]. After foreign DNA has invaded the cell, the next level of defense is the well-known restriction-modification system (R-M system). To date four different types of restriction enzymes (type I – IV) have been discovered [16]. The host DNA is usually methylated by a methyltransferase allowing the cell to discriminate between its own

and foreign DNA and is thus protected against cleavage by type I – III restriction enzymes, whereas type IV restriction enzymes cleave foreign modified DNA. A further mechanism of a microbial cell is to abort infection by undergoing cell death to protect the microbial community [17]. The latest discovered line of defense is the CRISPR/Cas-system. This inheritable, adaptive prokaryotic immune system was first recognized in 1987 by Ishino and coworkers when they found sets of 29 nucleotide (nt) long repeats interspaced by short (32 nt), non-repetitive sequences (spacers) downstream of the *iap* gene responsible for isozyme conversion of alkaline phosphatase in *Escherichia coli* [18]. In the last decade, more of those arrays where discovered in sequencing projects and in 2002, the term CRISPR/Cas was coined by Jansen [19]. In 2005, it was recognized that the spacer DNA in the CRISPR arrays was derived from viral or plasmid DNA [20,21,22] and finally in 2007, Barrangou [23] and coworkers experimentally proved that the CRISPR/Cas-system was able to provide bacterial immunity against viruses. Since then important advances to unravel the molecular mechanism of action of the CRISPR/Cas-system have been achieved.

1.2.2 Architecture and mode of action of the CRISPR/Cas-system

CRISPR/Cas-systems consist of the CRISPR array and a set of distinct *cas* genes upstream of the CRISPR array located in the genome of microorganisms (Figure 2). The CRISPR array is defined by conserved DNA repeats with a size between 20 to 50 nt and spacer sequences with approximately the same length of viral origin or plasmid DNA connecting them [24,25,26]. The number of spacer-repeat units as well as the length of the repeat sequence varies between different organisms [27]. Located upstream of the CRISPR array, an A-T base pair rich DNA region, called the CRISPR "leader", functions as a promotor [28]. A set of diverse *cas* genes is located even further upstream. These *cas* genes encode proteins and between four to ten are found within the CRISPR/Cas-operon in different organisms [29]. Because of gene and locus organization, CRISPR/Cas-systems have been categorized into three basic types with further subtypes (type I A-F, type II A-B and type III A-B) (Figure 2) [30].
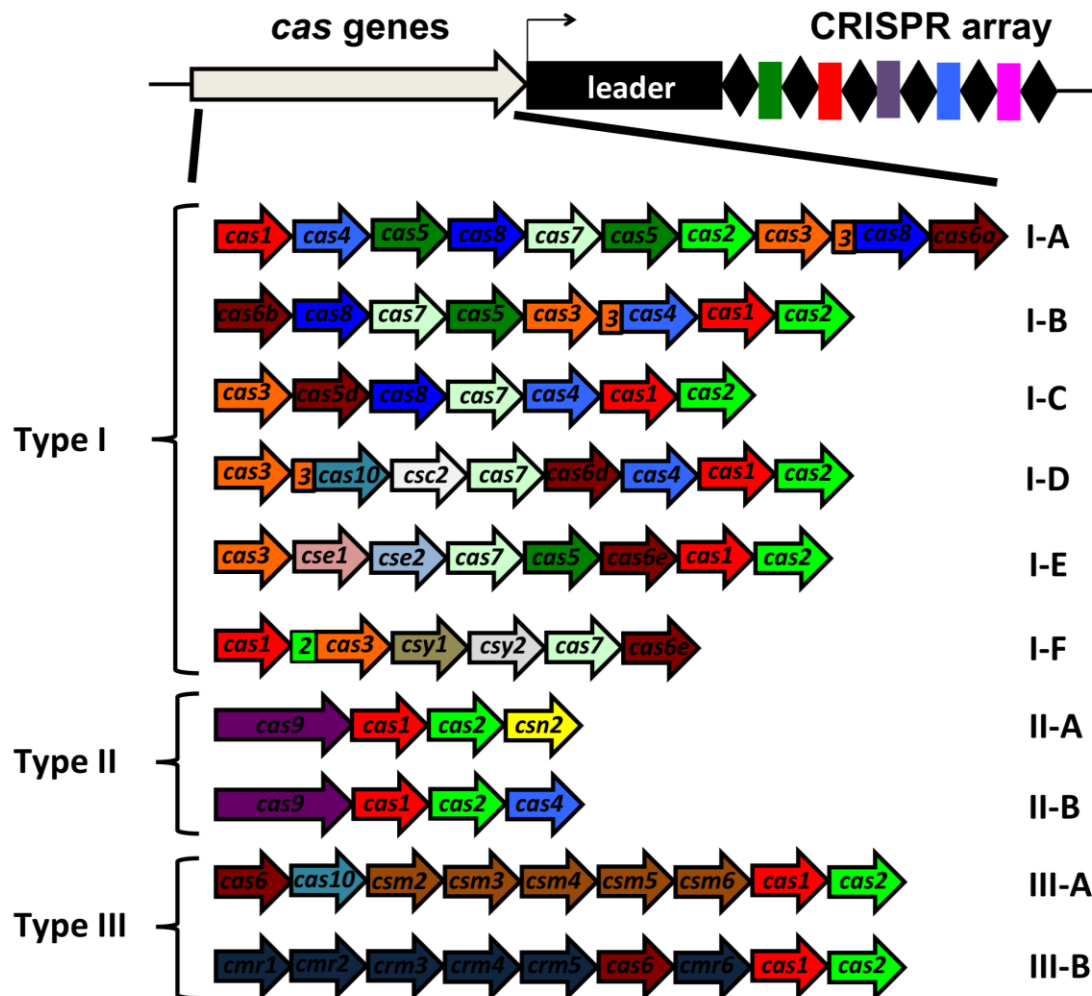
Figure 2: Overview of the CRISPR/Cas-system structure and the different *cas* gene operons.
Cas genes are shown as arrows, the leader sequence as black box, repeats as black diamonds and spacers as colored rectangles. *Cas* genes have been classified into three main types with further sub-classifications.

The Cas1 and Cas2 proteins are hallmark proteins. These nucleases are found in every *cas* gene operon, whereas the other Cas proteins are subtype specific. Although they differ in their Cas protein endowment, the global mechanism underlying the CRISPR/Cas-system is conserved. The mechanism can be divided into three main stages (Figure 3). The first stage is the acquisition stage (I), where foreign DNA is taken up and gets integrated into the CRISPR array as spacers. The second stage (II) is the expression of the CRISPR array. Thereby, crRNA gets transcribed, processed and is used as probe to detect foreign DNA. Foreign DNA is subsequently degraded by Cas proteins in the third stage (III), the interference stage. All three stages are explained in more detail below.

Figure 3: CRISPR/Cas mode of action.
The CRISPR/Cas-system acts in three steps: I) spacer acquisition, II) expression and maturation and III) interference. In type I and type II systems the selection of protospacers from foreign genetic elements depends on the PAM-motif, but how the PAM is recognized is still obscure. Cas1 and Cas2 are hallmark proteins of the CRISPR/Cas-system and are very likely involved i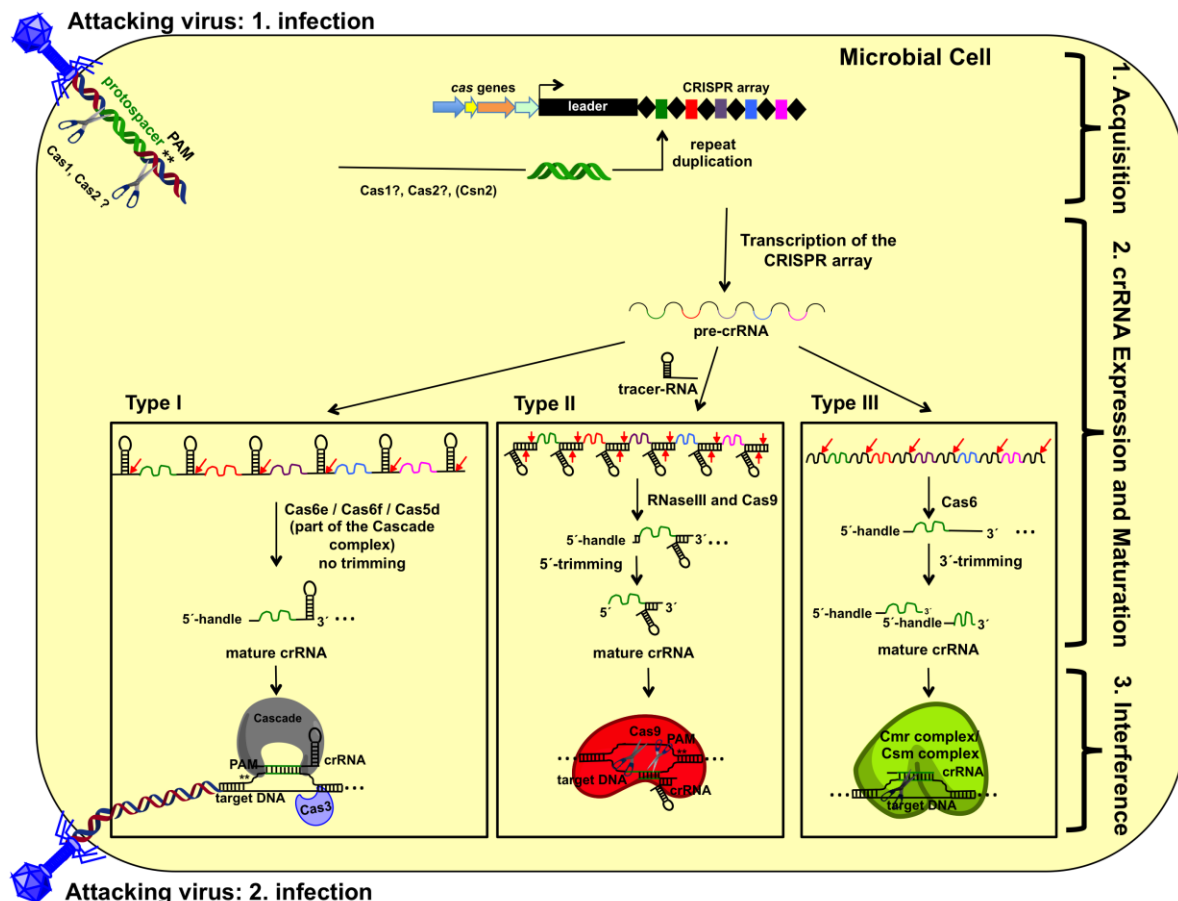n spacer integration, as well as Csn2 in type II systems. In the second step, the CRISPR array gets transcribed, which produces a long primary CRISPR transcript (the pre-crRNA). The pre-crRNA gets cleaved into smaller, mature crRNAs by different proteins depending on the type. In type I this is performed by the Cas6e, Cas6f or Cas5d protein depending on the subtype within the Cascade complex. In type II a small *trans*-activating-RNA (tracrRNA), RNaseIII and Cas9 are involved and in type III this maturation is performed by Cas6. In the interference step, the crRNAs are loaded on an effector complex (type I: Cascade, type II Cas9 and type III Csm or Cmr complex) and are used as probes to detect foreign nucleic acids and perform the digestion of the foreign nucleic acid.

1.2.3 The spacer acquisition step – Csn2 as a hallmark for type II systems

The first step in gaining immunity against foreign genetic elements is the integration of a short piece of this foreign nucleic acid into the CRISPR array. This piece is called the protospacer and is identified by a protospacer-adjacent motif (PAM) in the foreign nucleic acid fragment. PAMs are short DNA sequences between 2 to 5 nt in the foreign DNA and vary in sequence and location (upstream or downstream of the protospacer sequence) among the different CRISPR/Cas-subtypes [31]. The overall mechanism of spacer acquisition has still to be unraveled, but different studies

suggest an involvement of certain Cas proteins like Cas1 and Cas2. These two proteins are the hallmark proteins for CRISPR/Cas-systems, because they occur in all known systems. Both are conserved, metal-dependent nucleases which participate in spacer integration [32,33,34]. Overexpression of Cas1 and Cas2 from CRISPR types I-E and I-F in *E. coli* BL21 (DE3) resulted in spacer integration, but only when co-expressed . All newly acquired spacers were dependent on a PAM sequence and it was suggested that Cas1 and/or Cas2 recognize this motif [35]. Knock-out studies in type II-A systems revealed that spacer integration is dependent on the subtype-specific Cas protein Csn2 [23,36]. Csn2 is a calcium-dependent DNA binding protein [37,38] and its knock-out completely abolished spacer acquisition in *Streptococcus thermophilus*. Elucidation of the Csn2 mode of action could give an answer to spacer integration at least for type II-A systems or for the CRISPR/Cas-system in general, if orthologs of Csn2 existed in other subtypes. New protospacers are integrated next to the leader and it has been shown that the first repeat is duplicated during the integration event [23,35]. How the leader sequence is recognized for spacer acquisition still remains obscure, but the first 60 nt are essential for spacer integration as shown by mutational studies [35]. Furthermore, the first repeat next to the leader serves as a template for repeat duplication during spacer integration via an unknown mechanism [35].

1.2.4 CRISPR RNA expression and maturation

The next critical step in achieving immunity in CRISPR/Cas-systems is the expression and maturation of the CRISPR array. Once a cell is equipped with protospacer DNA, the whole CRISPR array is transcribed into a long RNA leading to a premature CRISPR-RNA (pre-crRNA). Transcription is initiated at the leader end, which serves as promotor [28,39,40]. After transcription, the pre-crRNA is further processed into small RNAs, the mature crRNAs, by nucleolytic cleavage of Cas proteins. The crRNAs can then be used as a probe to target foreign nucleic acids. The mechanism for pre-crRNA maturation differs with regard to the different types of CRISPR/Cas-systems. In type I systems, the pre-crRNA is processed by a multi-protein complex called CRISPR-associated complex for antiviral defence (Cascade) [41,42]. Cascade binds to the pre-crRNA and a Cas protein, which is a member of the complex (e.g. Cas5d, Cas6e or Cas6f depending on the type I subtype) cleaves

the pre-crRNA within the repeat sequence at the 3´-end [43,44,45,46,47]. The repeat sequence is partially palindromic and results in a stable stem loop which is recognized by the Cas protein and cleaved 3´ within its base. The final mature crRNA contains an 8 nt 5´-handle, the spacer sequence and, a 21 nt stem loop at its 3´-end [41,44]. The Cas6 protein processes pre-crRNA in type III systems [48,49]. In this system, the repeat sequence is unstructured, thus does not contain a stem loop. Cas6 cleaves within the unstructured repeat sequence, which leads to a 69 nt long crRNA including a 8 nt 5´-handle [50,51]. This crRNA gets further trimmed at the 3´-end, resulting in its complete loss of the 3´-repeat sequence. Trimming also results in two different long crRNA (e.g. 39 nt vs. 45 nt in *Pyrococcus furiosus*) [52,53].

In contrast to type I and III systems, it was shown for type II systems that pre-crRNA maturation is dependent on the multidomain Cas protein Cas9, the non-Cas protein RNaseIII and a small RNA, the *trans*-activating-RNA (tracrRNA) [36,54,55]. This RNA was found in *Streptococcus pyogenes* using deep sequencing and it could be shown that this 25 nt long RNA was complementary to the repeat sequence. It is encoded upstream of the CRISPR locus and binds to the repeat sequence in the pre-crRNA. The RNA duplex then is processed by RNaseIII, leaving a 15 nt overhang at the 5´-end and 21 nt overhang at the 3´-end. The duplex RNA is further trimmed at the 5´-end leading to the complete loss of the 5´-handle and 6 nt of the spacer sequence by an unknown mechanism.

1.2.5 CRISPR/Cas-interference – Destroying foreign nucleic acids

The last stage in CRISPR/Cas action is the interference step. Large ribonucleoprotein complexes target the invading nucleic acids and destroy them. Therefore, the crRNA is base-paired with the complementary nucleic acid. In type I this is performed by the Cascade complex, which is still loaded with the crRNA after maturation [56]. The complex scans target DNA for a PAM sequence, which then is bound [57]. The first 8 nt of the crRNA are critical for binding to the complementary target sequence, the so-called "seed" sequence. Point mutations within this sequence in the target prevent base pairing and destruction [57]. Then, the endonuclease Cas3 is recruited and nicks the DNA which reduces the affinity of Cascade and DNA is further degraded by Cas3 [58].

In contrast to type I and III systems, in type II systems the multidomain protein Cas9 is loaded with the duplex of crRNA and tracrRNA [54,59]. The PAM as well as base pairing are necessary for nucleic acid degradation. Cas9 cleaves the foreign DNA at the base-paired as well as the non-paired strand and leaves a blunt-end cut in the invading DNA [54,55]. For type III systems the degradation is similar to type I and is achieved by the Cmr or Csm complexes, where Cmr was shown to target RNA and the Csm complex DNA [52,53,60,61]. In contrast to type I systems, the identification of the target sequence is independent of the PAM motif. This suggests a fundamental difference in spacer acquisition and target destruction mechanisms between the different systems.

1.2.6 The human pathogen *Streptococcus agalactiae* harbors a P-loop containing *csn2* gene

Streptococcus agalactiae is a Gram-positive bacterium and the leading cause for neonatal infections, but can also infect older people [62,63]. Despite this potential, *S. agalactiae* is also part of the commensal microbiota in gastrointestinal and genitourinary tracts in healthy people. Eight genome sequences from human isolates have been determined [64,65]. CRISPR/Cas-systems have also been found within those genomes and a recent study of CRISPR spacer content from 351 *S. agalactiae* isolates showed its high dynamics and diversity of spacer acquisition [66]. Furthermore, spacer sequences from 10 patient isolates underlined this finding (unpublished results). They all varied in their spacer content as well as in their length showing that the CRISPR/Cas-system is active in *S. agalactiae* (Figure 4A). It has been shown that the repeat sequence is always linked to a specific CRISPR type and PAM motif [31]. A PAM consensus sequence is -NGG- in *Streptococcus spec*. This is linked to a type II-A CRISPR/Cas-system, a system with four Cas proteins (Figure 4B), the ubiquitous Cas1 and Cas2 nucleases, the multidomain protein Cas9 and Csn2. Cas1 and Cas2 play a role in spacer acquisition, whereas Cas9 is involved in crRNA maturation and interference. Csn2 is indispensable for spacer acquisition step as shown by mutational analysis [23] and therefore, much attention has been paid to this protein to unravel the acquisition mechanism.

A
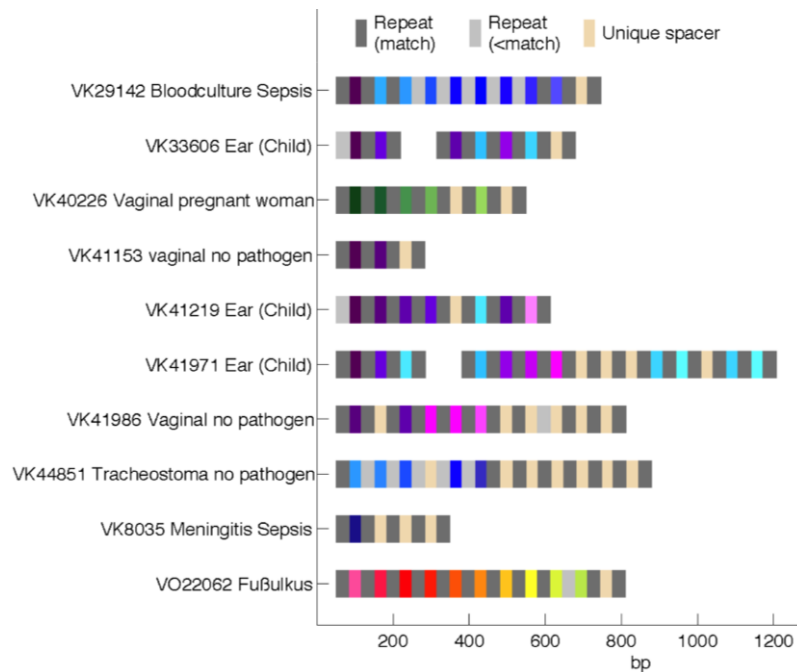
B

Type II-A CRISPR/Cas-system of *S. agalactiae*
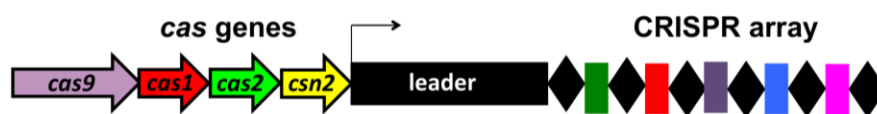
Consensus repeat sequence

5´- GTTTTAGAGCTATGCTGTTTTGAAT<u>GG</u>TCCCAAAAC-3´

Figure 4: CRISPR/Cas activity and structure of *S. agalactiae*.
A) Analysis of the spacer content of ten isolates of *S. agalactiae* from patients. Grey rectangles represent repeat sequences with the same sequence, colored obstacles represent spacer sequences from foreign origin. Each color stands for a different sequence. **B)** structure of the type II-A CRISPR/Cas-system of *S. agalactiae* and the consensus repeat sequence. Underlined is the PAM sequence (-NGG-, N= A,T,C,G) found in the repeat showing the link between repeat, PAM and type.

As no obviously conserved protein domain could be detetcted on the sequence level, a 3D homology search using the Csn2 sequence from *S. agalactiae* resulted in protein structure hits all reminiscent of P-loop containing proteins involved in DNA metabolism or even membrane transport. In chapter 1 of this thesis, the crystal structure of Csn2 was solved at 2.0 Å resolution and confirmed the existence of a sequence degenerated P-loop, but still displayed the structural P-loop. Other proteins also show this characteristic; they contain a structural P-loop, but no sequence

homology to the P-loop consenus sequence [67,68]. Csn2 is a calcium-dependent DNA binding protein, binding exclusively to dsDNA and showing no binding or hydrolysis of nucleotides [37,38,69,70]. It forms into a stable tetramer with a positively charged inner hole of a diameter of ~30 Å, often found in DNA binding proteins. Chapters 1 and 2 try to derive a structure-function analysis to gain insight into CRISPR spacer acquisition.

1.3 Transport across biological membranes

P-loop containing domains are not only found in soluble proteins, but are also present as part of membrane proteins, involved e.g. in the transport of substances across membranes [4]. Biological membranes are a prerequisite of life and are responsible for forming the envelope of a cell and shielding the interior from the outside [71]. The nature of a biological membrane is to provide a selective-permeable barrier letting pass required molecules and leaving harmful ones outside [72,73]. Small non-polar substances can easily penetrate membranes, whereas polar substances must penetrate the membrane with the help of membrane proteins. Prokaryotes contain only a plasma membrane (Gram-positive) or additionally an outer membrane (Gram-negative) [74]. Furthermore, biological membranes partition eukaryotic cells into different compartments possessing very specialized functions [1]. All these compartments like the nucleus, mitochondria, lysosymes, the Golgi apparatus or the endoplasmatic reticulum are defined by a unique composition of the membrane. In addition, biological membranes provide energy for the cell due to electrochemical gradients and building blocks for biosynthesis and play a role in signal transduction. Membrane proteins can account for up to 80 % of the composition of a biological membrane (e.g. lysosomal membranes) [75]. One of their main functions is to facilitate the transport across the membrane for large, polar molecules and ions by transmembrane (TM) proteins [1]. In general, membrane transport proteins can be classified into five classes with different subclasses according to the transport classification system of the International Union of Biochemistry and Molecular Biology (IUBMB, http://www.tcdb.org) (Figure. 5) [76]. Class I constitutes channels and pores and class II embraces electrochemical potential-driven transporters. The third class (class III) comprises primary active

transporters, class IV represents group translocaters while the last class, class V, comprises the transmembrane electron carriers.
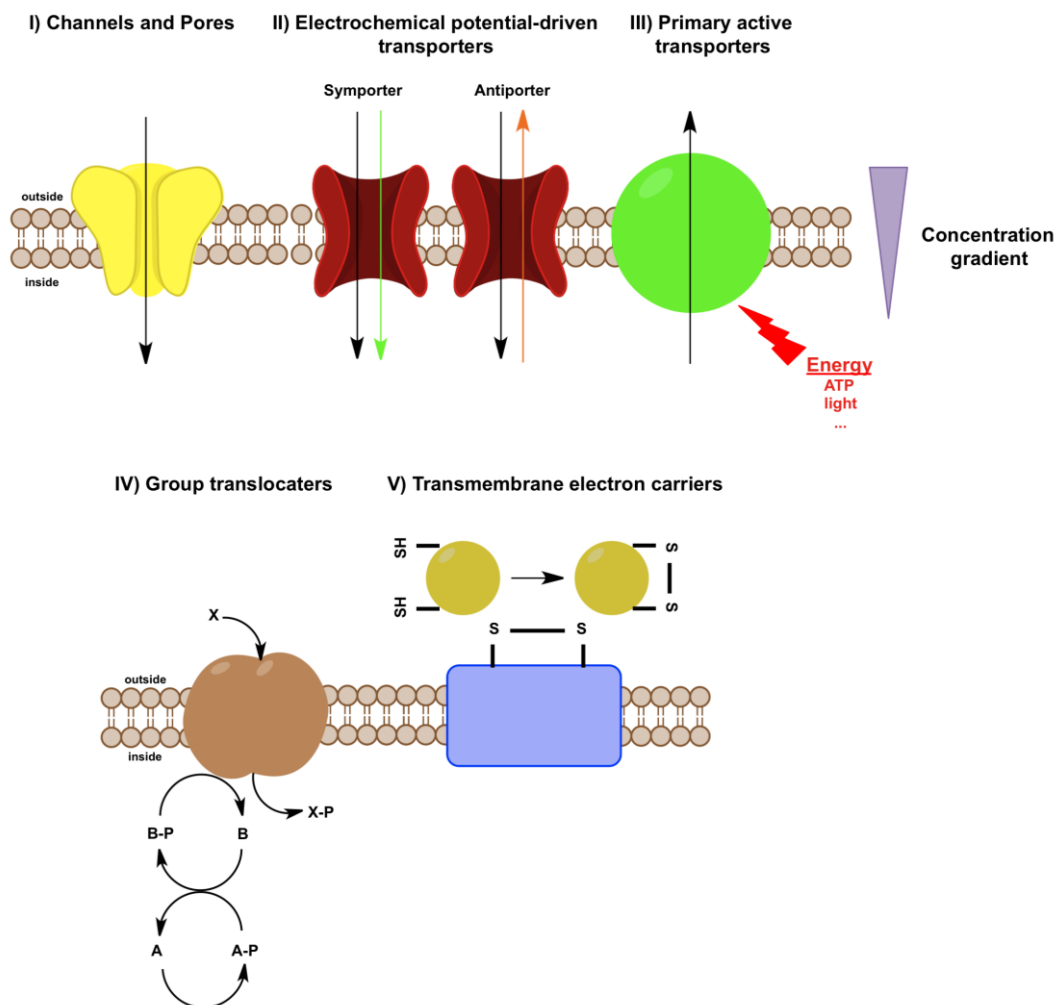
Figure 5: The five main classes of transmembrane proteins according to the IUBMB classification.
I) Channels and pores mediate the facilitated diffusion of substances along their electrochemical gradient, whereas II) electrochemical potential-driven transporters (also called secondary-active transporters) are either sym- or antiporter. They couple the transport of a substance against its electrochemical potential (green and orange arrow, respectively) to the electrochemical downhill transport of another molecule (black arrow, e.g. H^+). Class III transporters are primary active transporters using an energy source like ATP or light to transport against a concentration gradient. Group translocators (group IV) are sugar-transporting proteins. The sugar gets phosphorylated (P=phosphate group) while passing the membrane. The phosphorylated sugar cannot pass the membrane again and a concentration gradient is formed. The last class (class V) represents transmembrane electron carriers, which are membrane embedded electron providers.

Chapters 4 to 7 of this doctoral thesis are focused on the human ATP-binding cassette transporter bile salt export pump (BSEP), a P-loop containing primary active transporter, which uses binding and hydrolysis of ATP to transport bile salts across membranes of hepatocytes in the liver. The next sections give a general introduction into ABC transporters and their mode of action.

1.4 P-loop ATPases involved in membrane transport – the ATP binding cassette transporter (ABC transporter) superfamily

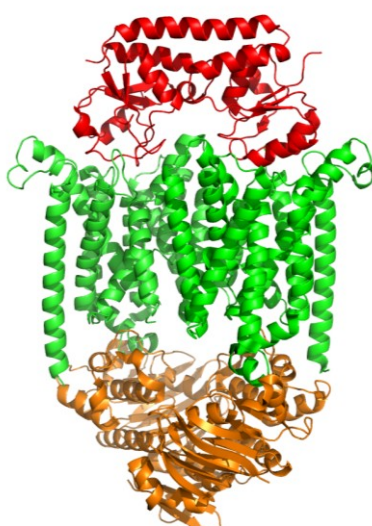ATP binding cassette (ABC) transporters belong to the large superfamily of ABC proteins and constitute one of the largest superfamilies among primary active transporters [77]. ABC transporters occur in all three kingdoms of life, namely archaea, bacteria and eukarya. Common to all is their modular architecture consisting of two transmembrane domains (TMDs) and two P-loop containing ABC modules called the nucleotide-binding domains (NBDs) (Figure 6). They all couple the uphill transport of a substrate or "allocrite" [78], which is chemically not altered, against a concentration gradient by the binding and hydrolysis of ATP. ABC transporters were first identified in the mid 1980s [79] with the identification and cloning of the human P-glycoprotein (P-gp, ABCB1, MDR1) [80,81,82] found to be responsible for drug-resistance in cancer cells as well as with the maltose and histidine permease systems in bacteria [83,84]. ABC transporters can be divided into two classes. Exporters, transporting allocrites out of a cell or into organelles and importers taking up e.g. nutrients into the cell [85]. The latter only occur in prokaryotes and archaea and are further subdivided into classes I and II (Figure 6 and below) [86]. Recently, a new class of import ABC transporters was discovered, the energy coupling factor (ECF) type ABC transporters [87,88]. These transporters consist of two NBDs (EcfA and EcfA´) and the transmembrane proteins EcfT and the S-component (EcfS) [89,90]. EcfS is responsible for the substrate binding, whereas EcfA-EcfA´-EcfT build up the energizing module. ECF transporters are widespread in Gram-positive bacteria and mediate the uptake of vitamins and metal ions [91]. To date ABC transporters with import function have not been found in eukaryotes. Although ABC transporters transport diverse and very heterogeneous allocrites (from small ions to whole proteins) and contribute to pivotal processes, they all contain the same blueprint with varying module alignment (Figure 6) [92]. In prokaryotes, the transporter modules are encoded mostly as single genes and are arranged in different ways, e.g. the transporter is built up of the same TMDs and NBDs or all modules differ. As a fifth "module", prokaryotic importers also depend on a substrate binding protein (SBP) [93]. This protein binds the allocrite, e.g. nutrients like sugars or vitamins, and delivers them to the transporter. In Gram-negative bacteria, these are soluble proteins diffusing in the periplasmic space where they bind their

13

substrates, whereas in Gram-positive the SBPs are anchored to the membrane via a lipid anchor or a transmembrane-helix [94]. In exporter-encoded genes all modules can either be found in one gene leading to a full-size transporter or there is also the possibility of half-size transporters, where only one TMD and one NBD are encoded on a single gene. These half-size transporters can either homo- or hetero-dimerize to form the functional unit of an ABC transporter [95].
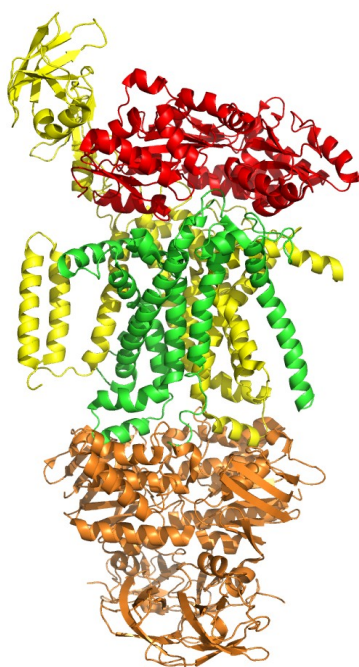
Type I Importer **Type II Importer** **Exporter**
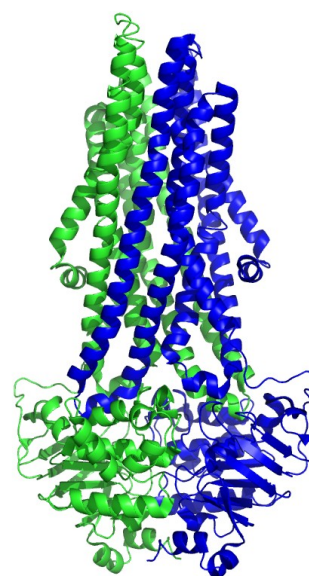
Figure. 6: Modular architecture of the three ABC transporter classes.
The three classes are exemplified by three crystal structures. Each color displays a separate gene. Type I importers are represented by the maltose transporter MalEFGK$_2$ (PDB code: 2R6G) [96]. The NBDs (MalK) are shown in orange, the different TMDs MalF and MalG in yellow and green, respectively, and the SBP MalE in red. Type II importer are represented by BtuCDF (PDB code: 2QI9) with the NBDs BtuD in orange, the TMDs BtuC in green and the SBP BtuF in red [97]. Sav1866 (PDB code: 2HYD) exemplifies the exporters [98]. This transporter is a homodimeric half-size transporter, each half consisting of one NBD and TMD are shown in green and blue, respectively.

How the binding of the allocrite in the TMDs, ATP binding and hydrolysis in the NBDs and the subsequent translocation of the allocrite is coordinated in concert will be discussed in more detail in the following sections.

1.4.1 The transmembrane domains (TMDs) – the tunnel through the membrane

The TMDs of ABC transporters form helical bundles which transverse the membrane and provide the translocation pathway for allocrites. Achievements of structural biology of whole ABC transporters over the last decade have provided a first insight into the nature of the TMDs [99]. To date three distinct TMD ABC transporter folds can be distinguished, type I ABC importer, type II ABC importer and ABC exporter folds [86]. Type I ABC transporter folds contain a core of five α-helices per TMD (Figure 7). Four α-helices (α-helices 2-5) exhibit an "up-down" topology lining the pore, whereas one α-helix (α-helix 1) wraps around the other four and lines the membrane.

This fold was first observed in the ModABC transporter, as well as in the MetI and MalFGK$_2$ transporter [96,100,101]. Type II ABC importer folds where first recognized for the vitamin B$_{12}$ importer BtuCD from *E. coli* and Hi1471 from *Haemophilus influenzae* [102,103]. Within one TMD 10 α-helices are found in total and nine of them pack around α-helix 2 like a barrel in a complex manner (Figure 7). The last fold represents the exporter fold. Until today this has been detected in all solved structures of ABC exporters, first in the Sav1866 transporter and later in MsbA, P-gp, ABCB10 (PDB code: 4AYW) and TM287-TM288 [98,104,105,106]. It shows a 6 x 6 topology, which means that each helical bundle contains six transmembrane α-helices. It is assumed that this fold is exclusively found within exporters. Strikingly, this fold contains a so-called "domain swapping". One α-helical bundle is formed by α-helices 1 and 2 from one TMD and α-helices 3-6 from the other TMD in e.g. Sav1866 [98].

Type I Importer fold
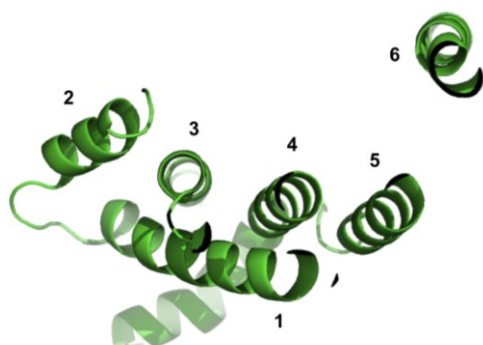
Type II Importer fold
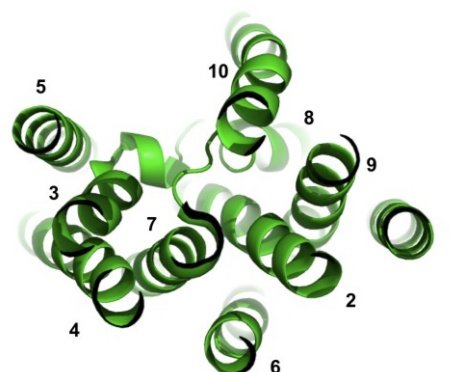
Exporter fold

Figure 7: TMD folds of ABC transporters.
The three different ABC transporter folds within one TMD (importer type I: MalG (PDB code: 2R6G), importer type II: BtuC (PDB code: 2QI9), exporter: Sav1866 (PDB code: 2HYD) are depicted. Shown is a cross-section of one TMD with a top view. Picture adopted from [99].

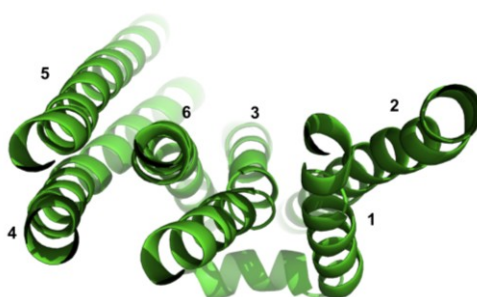Usually it is assumed that the TMDs determine the substrate specificity by providing a high affinity allocrite binding pocket within the TMDs in a certain conformation [107], but from all known structures, only two were solved with bound alllocrites or inhibitors. The maltose permease binds maltose only at one TMD (MalF) via ten residues using aromatic interactions as well as hydrogen bonds, but not at the other TMD (MalG) [96]. The crystal structure of P-gp from *Mus musculus* was solved with two different inhibitors [105]. They both bind in a cavity formed within the TMDs in the membrane with a size of ~6000 Å^3. They are bound by non-covalent interactions like electrostatic and van der Waals interactions, which might explain the polyspecificity of multidrug transporters [108]. Furthermore, the size of the cavity can theoretically accommodate more than one substrate or is filled with cholesterol to infill the cavity when an allocrite is bound regarding to the "cholesterol fill-in" theory. This model tries to explain how an ABC transporter can transport various distinct allocrites with the same binding site by the use of cholesterol [109]. Given the flexibility for allocrite binding there are also hints that some ABC importers might not have an allocrite-binding site at all, but are just dependent on the substrate binding protein (SBP) delivering the allocrite. Thus the TMDs just present a site for short-term passage during the allocrite's pathway [110]. Common to all TMDs is that they have intracytosolic loops (ICLs). They differ in length depending on their function as im- or exporters. ICLs of exporters are ~20-25 Å distant from the membrane interface because the TMDs of exporters have extended α-helices compared to the importer TMDs. The ICLs are short α-helices forming the interface between the TMDs and the NBDs and translate the binding and hydrolysis of ATP to allocrite transport and vice

versa, thus they are also called coupling helices (CHs) [111] and are explained in more detail in 1.4.3.

1.4.2 The nucleotide-binding domain (NBD) – the heartbeat of an ABC transporter

Nucleotide-binding domains are the cytosolic localized modules of ABC transporters containing the P-loop. They are responsible for powering allocrite transport by the binding and hydrolysis of ATP, thus they act as the motors of the transporter [112,113]. In contrast to the TMDs, NBDs are highly conserved within all three domains of life in respect to sequence features as well as structural shape. Structural and biochemical studies of isolated NBDs have greatly contributed to the understanding of their mechanism [114,115,116,117,118]. NBDs are assembled of two domains, a catalytic subdomain (CSD) and a helical subdomain (HSD), which leads to an L-shaped appearance (Figure 8). The two domains are hinged by two loops, the Q-loop and the Pro-loop. The CSD is derived from a RecA-like and an ABCβ subdomain. This CSD domain contains the aforementioned P-loop (Walker A) and Walker B motif. Furthermore, NBDs also contain conserved loops like the D-loop (-SAL<u>D</u>-) the H-loop and the A-loop.
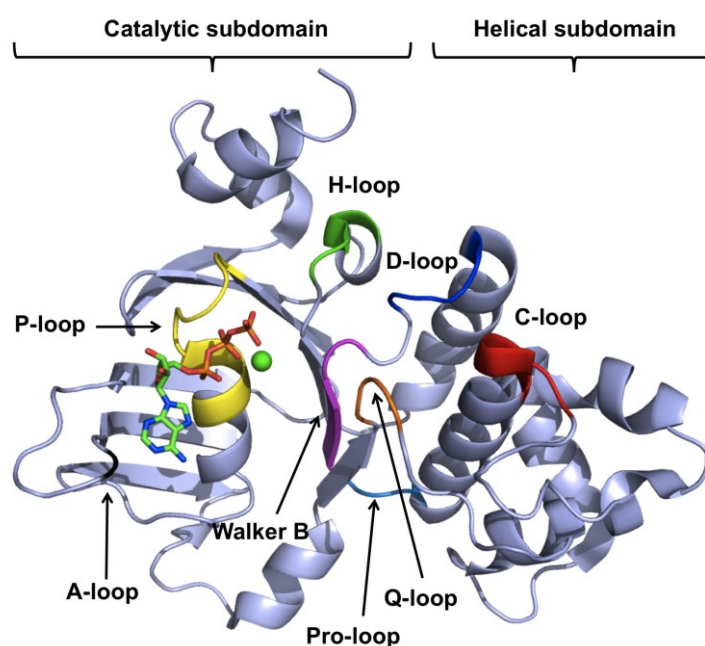
Figure 8: The NBD of an ABC transporter.
Shown is the NBD of the haemolysin B (HlyB) ABC transporter (PDB code: 1XEF) [117]. The P-loop is shown in yellow, the Walker B motif in magenta, the H-loop in green, the D-loop in blue, the C-loop in red, the Q-loop in orange, the Pro-loop in marine, the A-loop in black and ATP as sticks and a magnesium ion as green sphere.

The signature motif or C-loop (-LSGGQ-), which is the diagnostic feature or hallmark for an ABC protein is located in the HSD. These motifs complete the nucleotide-binding site. The ATP-binding sites are formed through dimerization of two NBD monomers in a head to tail arrangement and are located within the interface (Figure 9). Thus, the dimer contains two composite ATP-binding sites. The P-loop coordinates the β- and γ-phosphate groups of the ATP molecule, whereas the adenine ring stacks against the A-loop via π-π interactions. The C-loop of the opposing NBD monomer contacts the γ-phosphate group via the serine. The aspartate of the Walker B motif coordinates a magnesium ion through a water molecule and the glutamate acts as base by abstracting a proton of a water molecule, which then attacks the γ-phosphate group [13]. Together, the D-loop of the opposite NBD monomer, the H-loop and the Walker B motif coordinate the nucleophilic water molecule for an in-line attack on the γ-phosphate group. The H-loop acts as a "linchpin" in this catalytic triangle holding all residues together [117]. The Q-loop contributes to magnesium coordination and is further involved in NBD:TMD interface communication (see below).
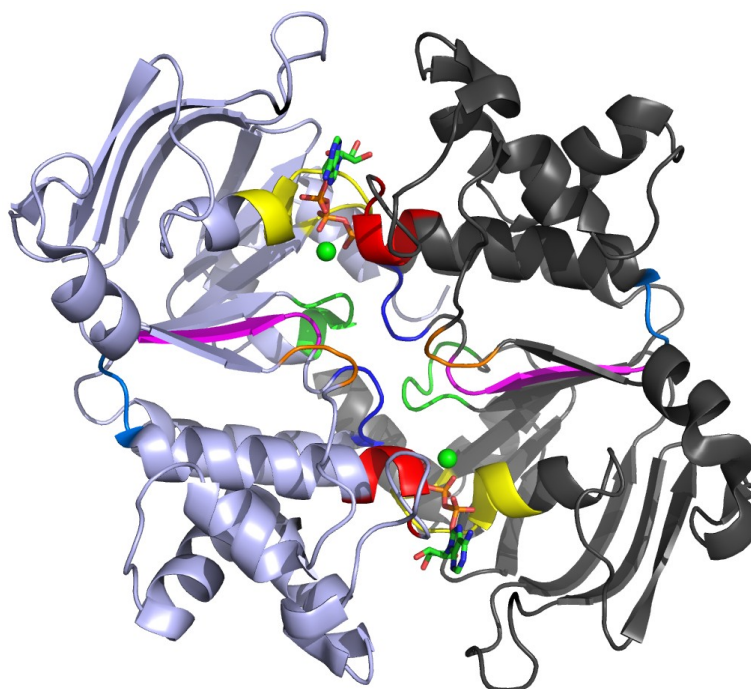
Figure 9: Dimer of the HlyB NBD.
The coloring is the same as in figure 8. ATP is sandwiched between the interface of two NBD monomers (violet and gray) by the P-loop, the Walker B of one NBD and the C-loop of the opposite NBD.

1.4.3 Coupling ATP binding and hydrolysis to transport of allocrites

To trigger the transport of allocrites, the TMDs, where the allocrite is bound and the NBDs, which fuel the transport, have to communicate with each other. This is achieved via an interface between the TMDs and the NBDs. As mentioned above, TMDs have intracytosolic loops (ICLs), the coupling helices, which are lined parallel to the membrane plane and interact within a groove in the NBDs (Figure 10). ABC exporters exhibit two of those loops (CH1 and CH2) per TMD and CH1 from TMD1 contacts NBD1, whereas the second coupling helix (CH2) interacts with the opposing NBD2 due to domain swapping [98,106]. In total, each NBD interacts with two coupling helices, one from each TMD. For ABC importers, only one CH per TMD is found which interacts with the directly apposed NBD [96,102]. The coupling helix of importers contains a diagnostic sequence, the EAA box, which is highly conserved [119]. The coupling helices interact with amino acid residues surrounding the Q-loop in e.g. Sav1866 or BtuCD (Figure 10, red surface). Furthermore, a new conserved region in exporters, the X-loop (-TEVGERG-), was found. It interacts with both coupling helices in each NBD and directly precedes the C-loop and thereby may play a role in sensing ATP binding and hydrolysis to the TMDs [98].

A   **B**
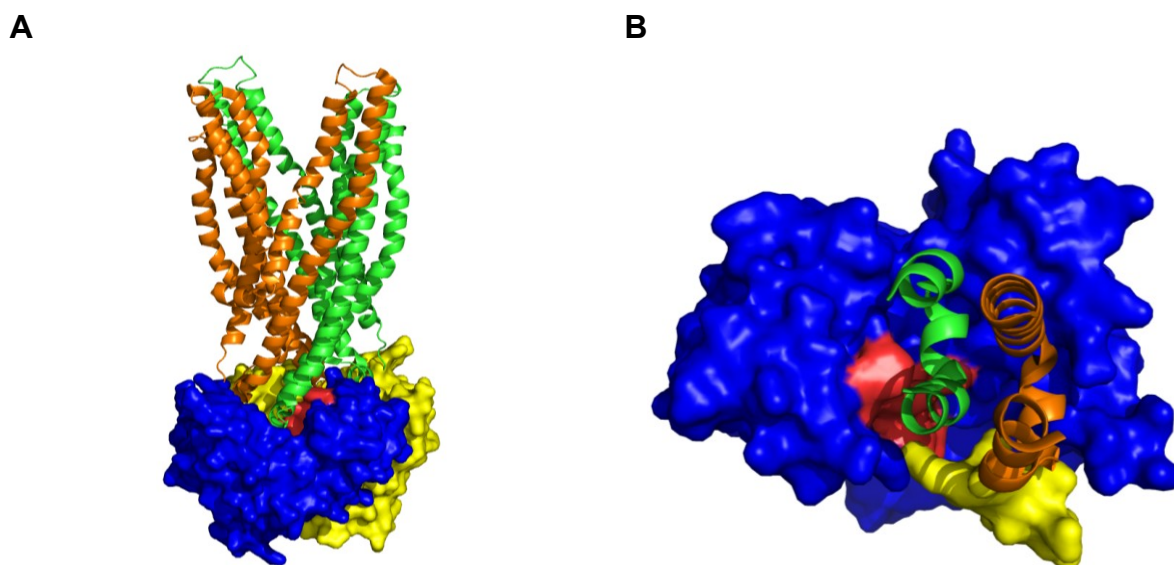
Fig. 10: NBD-TMD communication.
A) The Sav1866 exporter is shown (PDB code: 2HYD) with the TMDs colored green and orange and the NBDs blue and yellow. The NBDs are shown as surface. In red the Q-loop is shown. **B)** One NBD of the ABC transporter Sav1866 is shown (top view) in a blue surface representation with the Q-loop marked in red. The orange α-helix is from the apposed TMD, whereas the green one is from the opposite TMD. Marked in yellow is the X-loop of the opposing NBD. Both α-helices interact with the X-loop and the green α-helix also with the Q-loop.

1.4.4 The ATP-switch model

A widely accepted model for the transport mechanism of ABC transporters is the ATP-switch model (Figure 11) [120,121]. Based on the alternating access model for membrane proteins by Jardetzky established in 1966 [122], ABC transporters also alternate between two different conformations, the inward-facing conformation with a high-affinity binding site (in case of exporters, for importers it is vice versa) and an outward-facing conformation with a low-affinity binding site. The switch model is divided into four steps:

Step I: Allocrite binding to the high-affinity site within the TMDs induces conformational changes that are transmitted via the CHs to the NBDs and initiates the transport by binding ATP and closing the space between the NBDs.

Step II: The switch to the fully closed NBD dimer presents the power stroke, which inverts the TMD configuration from the inward-facing to the outward-facing conformation. In this conformation, the affinity for the allocrite is reduced and the allocrite exits the binding site.

Step III: ATP gets hydrolyzed and breaks up the closed NBD dimer.

Step IV: ADP and P_i are released from the NBDs and the transporter is reset.
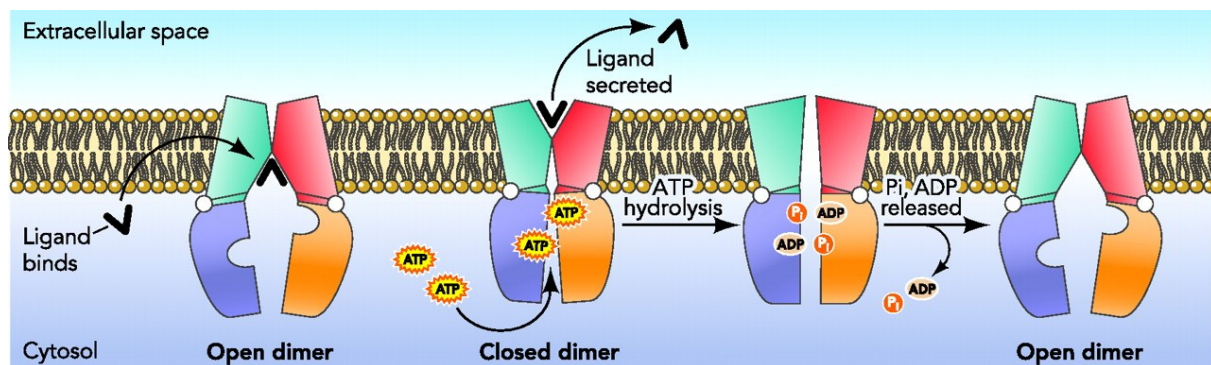
Figure 11: Scheme of the ATP switch model. Picture taken from [107].

1.5 The ABC of bile formation in the liver

In humans (*Homo sapiens*) 48 genes are identified, which encode for functional ABC proteins and ABC transporters [123,124]. Furthermore, 22 pseudogenes of ABC transporters are encoded in the human genome. A large portion of those is transcriptionally active, but no functional protein can be detected [125]. Because of

sequence similarities, domain organization and phylogenetic analysis, human ABC proteins and transporters can be subdivided into seven families (A-G) [123,124]. Twelve of these transporters are located in the liver, where they transport a huge variety of different allocrites.

The liver is a key organ in the body of vertebrates and performs a diversity of different functions. Main functions are the biosynthesis of amino acids, carbohydrates and lipids, the secretion of proteins e.g. albumin or blood-clotting proteins, the detoxification of the body by elimination of endo- and exogenic xenobiotics and the formation of bile [126]. The liver has a weight of 1.5 – 2 kg and 80 % of the liver volume is formed by hepatocytes. Hepatocytes are polarized cells with a sinusoidal (basolateral) membrane and a canalicular (apical) membrane (Figure 12A and 13) [127]. The sinusoidal membrane lines the blood stream and nutrients and xenobiotics are taken up from the blood across this membrane. Canalicular membranes form small bile ducts, called bile canaliculi and are the place of bile component secretion [128]. Bile is essential for the digestion of fat and for the absorption of lipids and fat-soluble vitamins like vitamin A and K in the small intestine [129,130]. Bile is a mixture of bile salts, phosphatidylcholine and cholesterol. Bile salts are synthesized in the hepatocytes from cholesterol as starting substance [131,132]. The synthesis of a primary bile salt like cholic acid or chenodeoxycholic acid consists of four steps including 17 enzymes: I) the initiation of synthesis by 7α-hydroxylation of sterol precursors, II) further modifications to the ring structure, III) oxidation and shortening of the side chain and IV) conjugation of the bile acid with the amino acids glycine or taurine, which lowers the pKa compared to the unconjugated form. After its synthesis bile salts are transported into the lumen of the canaliculi [128]. There, together with lipids and cholesterol mixed micelles are formed, which constitute bile. These mixed micelles dampen the detergent effect of the amphipathic bile salts and act as acceptors for poorly soluble xenobiotics as well as they prevent the formation of cholesterol crystals [133]. Bile is further stored in the gall bladder and is hormone-dependently released into the duodenum after food ingestion [134]. In the small intestine, primary bile salts are further converted to secondary bile salts by bacteria. Furthermore, bile salts are again taken up from the enterocytes and recycled back to the liver. This mechanism is called the enterohepatic circulation, which displays a very efficient recycling system (Figure 12B) [135]. Only 0.2 – 0.6 g of the 3 – 4 g human bile salt pool are excreted via the feces per day [136]. This amount is

resynthesized in the hepatocytes and bile salts circulate six to ten times a day between liver and intestine.

A

B
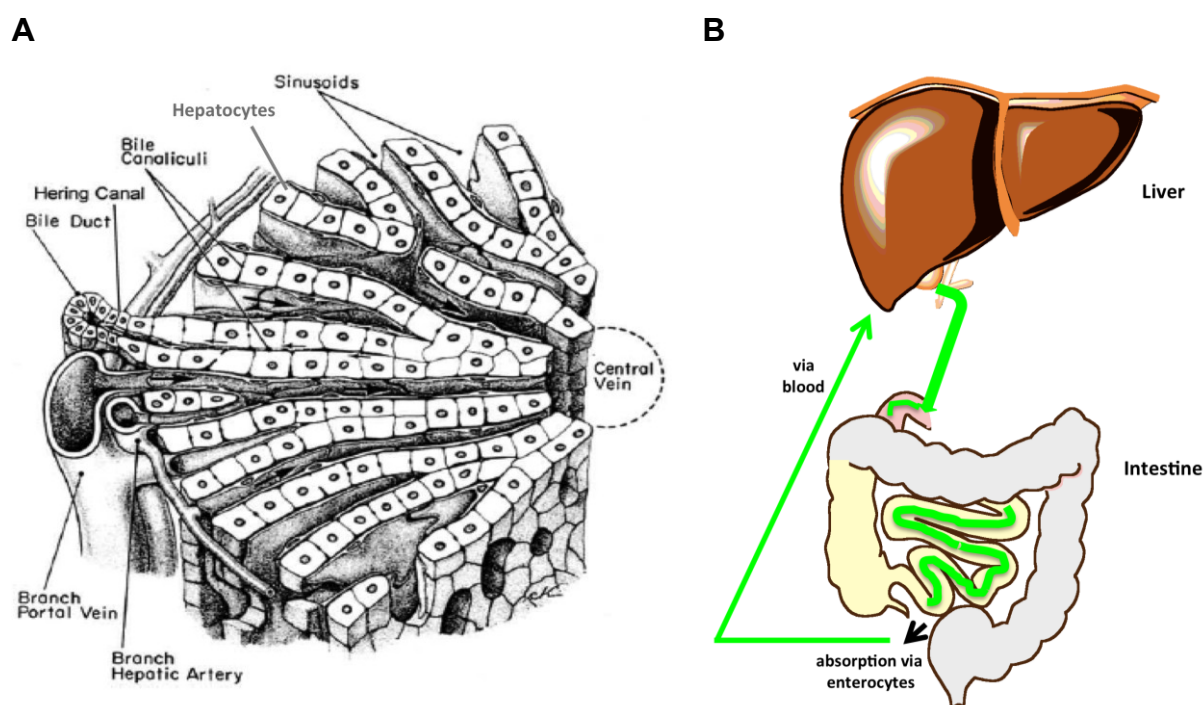
Figure 12: Liver lobule structure and the enterohepatic circulation.
A) A schematic overview about a liver lobule. Shown are hepatocytes lining bile canaliculi on one side and the sinusoidal blood on the other side. Picture adopted from [137] **B)** A diagram illustrating the enterohepatic circulation (bile flow is represented in green) is shown. Bile salts are released from the gall bladder into the duodenum and reabsorbed by enterocytes in the terminal ileum and then transported back to the liver via the blood.

Hepatocytes comprise a specialized set of transporters, especially ABC transporters, which are responsible for bile secretion and detoxification (Figure 13) [128,138]. The canalicular membrane contains ABC transporters often referred to as "vacuum cleaners" like P-gp (MDR1, ABCB1) and ABCG2 (BCRP), which transport a variety of hydrophobic, structurally not related molecules and can confer multidrug resistance in e.g. chemotherapy [139,140]. Divalent bile salts and organic anions conjugated to glutathione are substrates of ABCC2 (MRP2) [141]. As high concentrations of bile salts are cytotoxic, hepatocytes harbor a salvage system. This is built up by the two ABC transporters ABCC3 (MRP3) and ABCC4 (MRP4) as well as by the heterodimeric solute carrier (SLC) family members OSTα/OSTβ, all located in the sinusoidal membrane [142,143,144,145,146]. ABCC1, 5 and 6 (MRP1, 5 and 6) also reside in this membrane and transport different glutathione conjugates, folates and cyclic nucleotides [147]. Bile salts are transported by members of the SLC family

from the blood into the cytosol predominantly in a sodium-dependent manner through the sodium-taurocholate cotransporting polypeptide (NTCP) across the sinusoidal membrane [148].
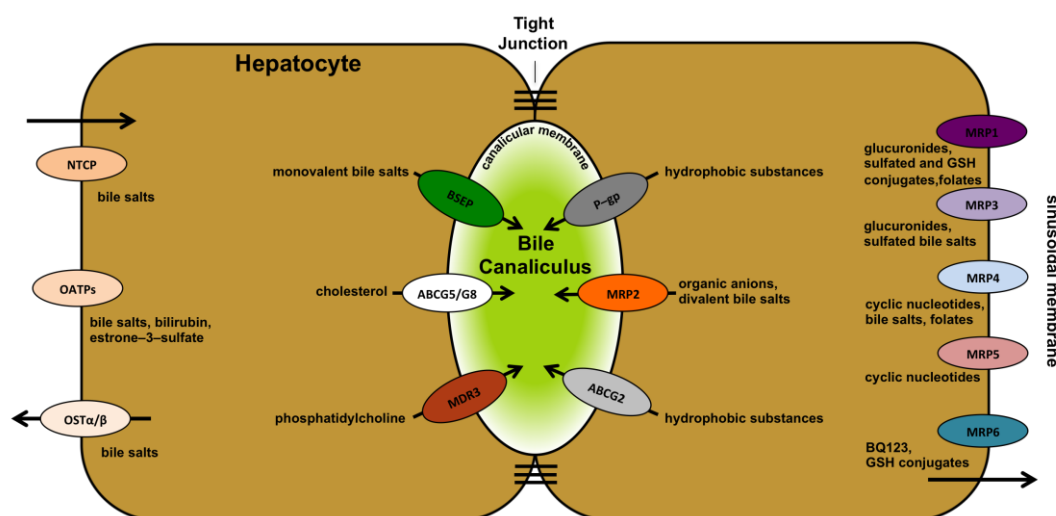
Figure 13: Localization of transporters in the hepatocytes.
Bile salts are taken up at the sinusoidal (basolateral) membrane through the sodium-taurocholate cotransporting peptide (NTCP) in a sodium-dependent manner and to a lesser extent through a sodium-independent transport by organic anion transporting proteins (OATPs). They are then further shuttled to the canalicular membrane and transported via the bile salt export pump (BSEP) into the canaliculus. Multidrug resistance protein 3 (MDR3) and ABCG5/G8 complete the bile formation by flopping phosphatidylcholine (PC) from the inner to the outer leaflet as well as transporting cholesterol. Bile salts, PC and cholesterol form mixed micelles, which constitute the basis of bile. P-glycoprotein (P-gp, MDR1) and the breast cancer resistance protein (ABCG2) transport a variety of hydrophobic substances into the bile and confer multidrug resistance (MDR). Furthermore, a substanial number of multidrug-related proteins (MRPs) are localized in the sinusoidal membrane, except MRP2. They transport a broad range of organic anions and conjugated substances. They also participate in MDR and some of them, e.g. MRP4 as well as the organic solute transporter (OST) act as salvage system for too high bile salt concentrations within the cell. Picture taken from chapter 3.

To a lesser extent, a sodium-independent transport, which involves organic anion transporting polypeptides (OATPs) is used [149]. Once bile salts have entered the cytosol, they are shuttled across the cell to the canalicular membrane, where the ABC transporter BSEP transports them into the canaliculus. Multidrug resistance protein 3 (MDR3, ABCB4) is an ABC transporter responsible for the flopping of phosphatidylcholine from the inner to the outer leaflet of the membrane. The bile forming machinery is completed with the transport of cholesterol through the heterodimeric ABC transporter ABCG5/G8 [128]. Chapter 3 of this doctoral thesis gives a deeper insight into the history and characteristics of BSEP and MDR3.

BSEP was discovered in 1995 when screening a pig liver cDNA library and named sister of P-glycoprotein (s-P-gp) because of its high sequence identity (~50 %) to P-gp [128].

It had been presumed that BSEP was also involved in multidrug resistance, but in 1998, it was proven that BSEP is a bile salt transporter responsible for bile salt transport in the liver [150]. In 2002, two studies regarding the first cloning of human BSEP and its characterization in an insect cell system were published [151,152]. BSEP is a full-size transporter with a size of 1321 aa and a molecular weight of roughly 160-180 kDa depending on its glycolsylation state. It is exclusively localized in the canalicular membrane and transports primary and secondary bile salts conjugated to taurine or glycine (e.g. taurocholate, taurodeoxycholate or glycochenodeoxycholate). High concentrations of free bile salts within the cell have detrimental effects on hepatocytes [153,154]. These undergo apoptosis or develop to cancer cells. This accumulation is called cholestasis, a reduction of bile flow. One cause can be mutations within the *BSEP* gene. These can lead to progressive familial intrahepatic cholestasis type 2 (PFIC2) [155], a disease which is characterized by an almost abolished transport of bile salts and therefore leads to cytotoxic effects due to accumulation of bile salts within the cell or benign recurrent intrahepatic cholestasis type 2 (BRIC2) [156], a milder form of cholestasis. The latter is characterized by a residual transport and influenced by the way of life. There are some therapies to make the bile salt pool more hydrophilic [157], e.g. the application of ursodeoxycholate, a more hydrophilic bile salt or surgical therapies like partial external biliary diversion (PEBD), to disrupt the enterohepatic ciruclation. By these therapy, bile gets excreted externally and the enterohepatic circulation and the bile salt concentration are lowered. The ultimate therapy for PFIC2 and BRIC2 is liver transplantation. Other effects like drug-induced cholestasis (DIC) or intrahepatic cholestasis of pregnancy (ICP) are also linked to different BSEP mutations or single nucleotide polymorphisms (SNPs) [158,159]. BSEP is inhibited either by drugs or hormones and a cholestasis is manifested. Until now, BSEP has only been characterized in heterologous overexpression systems based on cell culture systems. To characterize BSEP in its isolated form and to make a first step towards its structural elucidation, adequate amounts are required, which cannot be achieved with cell culture systems. Thus, other systems have to be used. In chapters 4 and 5, the instable human *BSEP* cDNA was cloned for two yeast-based expression systems and BSEP could be overexpressed to high amounts and subsequently be purified for biochemical characterizations. To understand the impact of mutations leading to PFIC2 or BRIC2, plasma membranes of BSEP-expressing yeasts were prepared to

characterize clinically relevant mutants in a vesicular transport assay. A first clinically-relevant mutant could be investigated with this established system, as shown in chapters 6 and 7.

2. Aims and Objectives

Nucleotides like ATP and GTP serve diverse functions like providing energy for biochemical processes, signal transduction or involvement in protein translation. As diverse as these functions are, as diverse are the structures of the nucleotide binding proteins. But, one common ground of almost all these proteins is that they harbor the same structural motif for binding nucleotides, the P-loop. The aim of this doctoral thesis is to characterize two P-loop containing proteins involved in different processes.

The first aim is to assign a function to the Cas protein Csn2 from *S. agalactiae*. This protein is involved in the bacterial immune system CRISPR/Cas and has a crucial role in acquiring immunity. It is involved in the integration of small DNA fragments into the host genome although the mechanism of this step is obscure. Thus, to unravel this step and the function of Csn2 as well as the involvement of the P-loop, one aim is the elucidation of the three-dimensional structure of Csn2 using X-ray crystallography. Furthermore, biochemical assays, including DNA binding assays, should further support the structural findings and lead to a model for spacer integration in CRISPR/Cas-systems.

The second aim of this doctoral thesis is the investigation of a P-loop protein involved in membrane transport, the human ABC transporter bile salt export pump (BSEP). This ABC transporter is the main bile salt transporter in the liver of vertebrates and mutations within its gene can lead to severe cholestatic diseases. To study the transporter in its isolated form *in vitro* to prevent disturbing side effects, one objective is the heterologous overexpression of human BSEP in yeast systems, sufficient for purification. Biochemical studies should gain insight into mechanistic coupling between transport and ATP-hydrolysis, dependent on the P-loop. One obstacle is the instability of the *BSEP* cDNA, which hampers its cloning for overexpression systems. Therefore, a general cloning and mutagenesis strategy should be developed to circumvent this problem leading to a rapid and flexible cloning of *BSEP*. Beside overexpression and purification, a further objective is the introduction of clinically relevant mutations into BSEP and their charaterization *in vitro*.

26

Chapter 1

The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*

Published in: *Journal of Structural Biology*

Impact factor: 3.406

Own Proportion

 to this work: 70 %

Cloning, expression and purification of Csn2

Size exclusion chromatography and blue native PAGE analysis

Crystallization and structure determination of Csn2

Writing of the manuscript

The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*

Philipp Ellinger [a], Zihni Arslan [b], Reinhild Wurm [b], Britta Tschapek [a], Colin MacKenzie [c], Klaus Pfeffer [c], Santosh Panjikar [e,1], Rolf Wagner [b], Lutz Schmitt [a], Holger Gohlke [d,*], Ümit Pul [b,*], Sander H.J. Smits [a,*]

[a] Institute of Biochemistry, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany
[b] Institute of Molecular Biology of Bacteria, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany
[c] Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany
[d] Institute of Pharmaceutical and Medicinal Chemistry, Heinrich Heine University, Universitätsstr. 1, 40225 Düsseldorf, Germany
[e] EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

A B S T R A C T

The prokaryotic immune system, CRISPR, confers an adaptive and inheritable defense mechanism against invasion by mobile genetic elements. Guided by small CRISPR RNAs (crRNAs), a diverse family of CRISPR-associated (Cas) proteins mediates the targeting and inactivation of foreign DNA. Here, we demonstrate that Csn2, a Cas protein likely involved in spacer integration, forms a tetramer in solution and structurally possesses a ring-like structure. Furthermore, co-purified Ca^{2+} was found important for the DNA binding property of Csn2, which contains a helicase fold, with highly conserved DxD and RR motifs found throughout Csn2 proteins. We could verify that Csn2 binds ds-DNA. In addition molecular dynamics simulations suggested a Csn2 conformation that can "sit" on the DNA helix and binds DNA in a groove on the outside of the ring.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Microorganisms have developed several mechanisms to defend invasions by foreign nucleic acids. In contrast to the well-known defense strategies such as the restriction-modification system or inhibition of phage adsorption, the recently discovered CRISPR system (clustered regularly interspaced short palindromic repeats) functions as an inheritable and adaptive immune system of prokaryotes (Al-Attar et al., 2011; Karginov and Hannon, 2010; Marraffini and Sontheimer, 2010a). More than 90% of archaea and nearly 40% of sequenced bacteria are equipped with CRISPR (Grissa et al., 2007), consisting of one or more CRISPR cassettes and a group of CRISPR-associated (Cas) proteins (Haft et al., 2005; Jansen et al., 2002). The CRISPR cassette consists of short, often palindromic, 28–40 bp DNA repeat sequences, separated by non-identical spacer sequences of similar length. In general, the CRISPR cassettes represent the acquired memory of immunity, enabling recognition of the invader nucleic acid. More than 40 Cas protein families are known (Haft et al., 2005). According to the recent "polythetic classification" of the different CRISPR–Cas modules, three different CRISPR types (Type I, II, and III) exist, which differ in the Cas protein composition and defense mechanisms (Makarova et al., 2011).

CRISPR defense can be dissected in three steps, (i) CRISPR adaptation, (ii) CRISPR expression/processing and (iii) CRISPR interference. The first step (immunization or adaptation stage) describes the capturing of new spacers that originate from the cleavage of invading DNA to short pieces; these spacers are subsequently integrated into the CRISPR array of the host genome (Barrangou et al., 2007). The exact mechanism of this integration step is unknown. Two universal Cas proteins Cas1 and Cas2, which are common to the three CRISPR types, are likely involved in this process (Deveau et al., 2010; van der Oost et al., 2009). Both proteins have been identified as nucleases (DNase/RNase activity for Cas1 and RNase activity for Cas2) (Babu et al., 2011b; Beloglazova et al., 2008; Samai et al., 2010; Wiedenheft et al., 2009). Short motifs within the invading DNA, typically a few nucleotides in length and referred to as proto-spacer-adjacent motifs (PAMs), seem to be

---

involved in recognition/selection of new spacers (Mojica et al., 2009). In addition, the PAM sequences are required to prevent a CRISPR autoimmunity reaction (Marraffini and Sontheimer, 2010b).

The second step comprises the transcription of the CRISPR array into the precursor CRISPR-RNA (pre-crRNA) and expression of Cas proteins; the latter mediate processing of pre-crRNA to active crRNAs. In type I CRISPR systems, this processing is carried out by a multiprotein complex termed Cascade (Brouns et al., 2008). The CRISPR Nmeni-subtype, one of eight subtypes, belongs to the type II CRISPR systems and processes the pre-crRNA with the help of a trans-encoded tracrRNA and RNaseIII and the multidomain protein Cas9 (known as Csn1) (Deltcheva et al., 2011).

The third step is the inactivation of invading foreign DNA, or in some cases RNA, which is also performed by the Cascade complex and Cas3 in type I (Brouns et al., 2008) or the Cmr/Csm complex (Hale et al., 2009) in type III. The foreign DNA is targeted with the help of the crRNA and inactivated by digestion.

In the Nmeni CRISPR subfamily, the Cas9 protein is involved in neutralizing invading DNA (Garneau et al., 2010), whereas the endonucleases (or nuclease) Cas1 and Cas2 as well as the subfamily specific Csn2 protein are involved in spacer integration, although their precise roles in this process are still obscure (Sapranauskas et al., 2011). Most functional information of these proteins comes from knockout studies. For example, deletion of Cas9 in *Streptococcus thermophilus* abolishes the resistance against foreign DNA. Knockdown of Csn2 still confers the ability to target and inactivate invading, foreign DNA with the help of existing spacers (Barrangou et al., 2007; Garneau et al., 2010; Sapranauskas et al., 2011). This led to the assumption that Csn2 proteins are involved in spacer integration. Recently, Csn2 from *E. faecalis* was solved at 2.7 Å resolution, revealing a tetrameric ring-like structure. Furthermore it was shown that Csn2 is capable of binding ds-DNA, and it was proposed that the DNA binds through the center of the ring (Nam et al., 2011).

Here, we report the X-ray structure of Csn2 of *S. agalactiae* ATCC13813 at 2.0 Å resolution and reveal that Csn2 is a stable tetramer that can undergo significant conformational changes. Furthermore, we show that Csn2 binds ds-DNA. Csn2 could have two binding modes: (I) "sitting" on the DNA and (II) the DNA passes through the center of the ring.

2. Materials and methods

2.1. Recombinant protein expression vectors

Full length Csn2 was subcloned into pET28b (Novagen) on a NcoI/XhoI fragment generated by PCR from genomic *Streptococcus agalactiae* ATCC13813 DNA using the following primers: For 5′-TTCCTAGGAGATTCCCATGGTCAAGATTAATTTTCCAAT-3′ and Rev 5′-TATTACTGTGCTTTTACTACTCGAGTACCATATTTTCGCC-3′. The coding sequence was introduced in frame with the start codon and a C-terminal hexahistidine purification tag encoded by the vector. The Csn2 mutants $E_{24}A$, $K_{132}A$, $Y_{29}A$, $Y_{29}W$ and $R_{198}A$, $R_{199}A$ were prepared using the QuikChange XL mutagenesis kit (Stratagene). All sequences were verified by in-house sequencing at the Biological Medical Research Centre (BMFZ) Heinrich Heine University, Düsseldorf.

2.2. Recombinant protein expression

pET28b-Csn2 transformed *Escherichia coli* BL21(DE3) pLysS cells (Novagen) were grown in LB medium at a temperature of 37 °C to an OD_{600} of ~0.6 and then protein expression was induced by adding β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Following induction, cells were grown at 30 °C for 5 h and harvested by centrifugation at 7500g for 15 min and 4 °C, flash frozen in liquid nitrogen and stored at −80 °C.

2.3. Purification of Csn2

Cells were resuspended in lysis buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 2 mM β-mercaptoethanol (β-ME) and lysed using a cell disruptor (Constant Systems). The lysate was clarified by centrifugation at 40,000g in a Ti45 rotor (Beckmann) for 1 h at 4 °C. Imidazole was added to the Csn2 supernatant to a final concentration of 20 mM before application to a 5 ml HiTrap Chelating column (GE Healthcare) loaded with nickel, which had been pre-equilibrated with 5 column volumes (CV) of running buffer (50 mM Tris–HCl, pH 8.0, 150 mM NaCl, 2 mM β-ME, 20 mM imidazole). After application of the supernatant to the column, it was washed with 5 CV of running buffer and elution was performed by a linear-gradient with running buffer containing 300 mM imidazole. Pooled fractions were loaded onto a HiLoad Superdex 200 prep grade 16/60 column (GE Healthcare) equilibrated with 1.3 CV of running buffer (10 mM Tris–HCl, pH 8.0, 50 mM NaCl, 1 mM DTT). Fractions containing protein were pooled, concentrated using an *Amicon* Ultra-15 filter (Millipore) with a cut-off of 10 kDa and stored at 4 °C.

2.4. Expression and purification of selenomethionine-labelled Csn2

For selenomethionine substitution, *E. coli* B834(DE3) cells were grown in M9 minimal medium supplemented with 50 μg/ml of L-selenomethionine (Molecular Dimensions). Expression and purification were identical with those for native Csn2.

2.5. Native page of Csn2

The oligomeric state of Csn2 and its mutants was analyzed by native gel electrophoresis using a 4–16% NativePAGE™ Novex Bis-Tris gradient gel (Invitrogen), which was stained with Coomassie brilliant blue.

2.6. Multiple angle light scattering

To determine the oligomeric size of Csn2, a multiple angle light scattering (MALS) setup consisting of miniDAWN Treos/optiLAB rex (Wyatt Technologies) connected to an Äkta Purifier (GE Healthcare) using a Superdex 200 10/300 analytical size exclusion column was used. The flow rate was set to 0.2 ml/min and ultraviolet (UV) detection was monitored at 280 nm. Light scattering was detected at angles of 0°, 90°, and 107° and the obtained values were averaged and evaluated using the program ASTRA (Wyatt Technologies). For light scattering experiments different NaCl concentration were used in SEC running buffer (50 mM, 250 mM and 500 mM).

2.7. Crystallization of Csn2

Crystallization trials were carried out at 4 °C. Crystals of Csn2 (at 25 mg/ml) were grown by mixing protein solution with reservoir solution containing 0.1 M Hepes, pH 6.8–7.2, 9–12% (w/v) PEG6000 and 0.05 M phenol in a 1:1 ratio. Crystals normally grew in 7–10 days. Suitable crystals were cryo-protected using crystallization buffer with 30% (v/v) ethylene glycol and then cryo-cooled in liquid nitrogen. Selenomethionine derivatised crystals were obtained and treated using the same method.

2.8. Data collection and structure determination

Data sets from native crystals were collected at the BM30A beamline (ESRF, Grenoble, France) at a wavelength of 0.979 Å at 100 K. For selenomethionine-substituted crystals beamline X12 (EMBL, DESY Synchrotron, Hamburg, Germany) was used and anomalous diffraction data were collected at the selenium absorption edge; $E = 12.667$ keV and $\lambda = 0.977$ Å at 100 K. All data sets were processed using the XDS software packages (Kabsch, 1993). The structure was solved by single-wavelength anomalous dispersion (SAD) from a selenomethionine derivative with a resolution of 3.4 Å using the program Autorickshaw (Panjikar et al., 2005; Panjikar et al., 2009). The SeMet model was then used to phase the native data set at a resolution of 2.0 Å. Automatic model building was performed with the program ARP/warp (Langer et al., 2008) and the quality of the electron density allowed also automatic placement of water molecules which was manually checked on correct assignment. Iterative cycles of model refinement using TLS groups were carried out using the program REFMAC5 (Murshudov et al., 1997) from the CCP4 package and manual adjustments between refinement was done with the program Coot (Emsley and Cowtan, 2004). Ramachandran analysis was done using MolProbity (Chen et al., 2010).

Data collection and Refinement statistics are listed in Table 1. Images of the models were prepared using MacPyMOL (www.pymol.org).

2.9. DNA-binding analyses

DNA-binding activity of Csn2 was analyzed by incubation of 2 nM ^{32}P-labeled 256 bp P1 DNA fragment (Pul et al., 2007) or IGLB DNA fragment (Pul et al., 2010) with indicated amounts of Csn2 in a buffer containing 50 mM Tris–HCl, pH 7.4, 70 mM KCl, 15 mM NaCl and 10 mM β-mercaptoethanol for 20 min at 37 °C. The complexes were separated on a native 5% (w/v) Tris/Glycine polyacrylamid gels (Nam et al., 2011).

2.10. Molecular dynamics simulations

Molecular dynamics simulations were performed with the AMBER 11 suite of programs (Case et al., 2005) together with the force field as described by Cornell et al. (1995) using modifications suggested by Simmerling et al. (2002). Two different MD simulations were performed: (I) the tetrameric Csn2 crystal structure is used as a starting structure. Henceforth, this structure will be referred to as "tetrameric structure". The simulation length is 100 ns; (II) A dimer composed of two α/β domains in the orientation observed in the crystal structure is used as a starting structure. This structure was obtained from chains A and B of the crystal structure by excising residues 70–137 and capping the resulting C- and N-terminal ends with N-methylamine and acetate, respectively. Henceforth, this structure will be referred to as "dimeric structure". The simulation length is 200 ns. In all cases, the starting structure was placed into an octahedral periodic box of TIP3P water molecules (Jorgensen et al., 1983). The distance between the edges of the water box and the closest atom of the protein was at least 11 Å, resulting in system sizes of 51,000 ("dimeric structure") and 160,000 ("tetrameric structure") atoms. The system was minimized by 50 steps of steepest descent minimization followed by 450 steps of conjugate gradient minimization. The particle mesh Ewald (PME) method (Darden et al., 1993) was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using SHAKE (Ryckaert et al., 1977). The time-step for all MD simulations was 2 fs, with a direct-space, non-bonded cutoff of 8 Å. Applying harmonic restraints with force constants of 5 kcal mol^{-1} Å^{-2} to

Table 1
Crystal parameters and data collection statistics are derived from XDS. Refinement statistics were obtained from REFMAC5. Ramachandran analysis was performed using MolProbity.

	Native	Selenomethionine
Crystal parameters at 100 K		
Space group	C2	C2
Unit cell (Å) a, b, c	75.3, 83.3, 110.4	74.9, 82.3, 109.4
β°	109.4°	108.9°
Data collection and processing		
Wavelength (Å)	0.979981	0.977530
Resolution (Å)	50.00–2.00 (2.05–2.00)	50.00–3.41 (3.50–3.41)
Mean redundancy	4.5 (4.2)	3.7 (3.2)
Unique reflections	43206 (3103)	16,726 (1232)
Completeness (%)	99.4 (97.9)	99.2 (97.9)
I/σ	36.51 (7.39)	14.83 (6.70)
R_{merge}[a]	2.6 (17.9)	7.9 (17.4)
Wilson B-factor (Å^2)	33.33	31.15
Refinement		
R_F[b] (%)	20.4	
R_{free}[c] (%)	22.8	
rmsd Bond lengths (Å)	0.0077	
rmsd Bond angles (°)	1.0372	
Average B-factors (Å^2)		
Protein	35.62	
Ca^{2+} ions	33.30	
Ramachandran plot		
Molprobity Score	1.43	
Favored (%)	94.4	
Allowed (%)	3.4	
Outliers (%)	2.2	
Monomers/ASU	2	
Protein residues	433	
Others	3 ethylene glycol, 5 calcium ions, 296 water	

[a] R_{merge} is defined as $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.
[b] R_F as $R_f = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$.
[c] R_{free} is calculated as R_F but for 5% randomly chosen reflections that were omitted from all refinement steps. Data in brackets corresponds to the highest resolution shell.

all solute atoms, canonical ensemble (NVT)-MD was carried out for 50 ps, during which the system was heated from 100 K to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD was used for 150 ps to adjust the solvent density. Finally, the force constants of the harmonic restraints on solute atom positions were gradually reduced to zero during 100 ps of NVT-MD. The following NVT-MD at 300 K with a time constant of 10 ps for heat-bath coupling was used for analysis, with conformations extracted every 20 ps.

For analyzing the trajectories, conformations were superimposed with respect to C_α atoms of one of the α/β dimers. This resulted in an almost perfect overlay of the dimers in all cases. The superimposition was also used for monitoring the movement of the second α/β dimer with respect to the first one; for this, root mean-square deviations with respect to the tetrameric starting structure were determined for C_α atoms of the second α/β dimer.

2.11. Detection of transient pockets

Transient pockets were detected in conformations extracted from MD simulations of the dimeric structure at 200 ps intervals by means of the PocketAnalyzer^{PCA} program (Craig et al., 2011). PocketAnalyzer^{PCA} implements a pocket identification strategy similar to the one proposed by Hendlich et al. (1997). After building a cubic grid around the protein, only those cubes that are not

covered by a protein atom are further considered. Each of these "empty cubes" must meet the following three criteria to become part of a binding pocket: (I) A minimal degree of buriedness (dob), computed as the number of rays extending from the cube center through the center of all six cube surfaces as well as along all cube diagonals that hit atoms of the surrounding protein at a distance ⩽10 Å. (II) To exclude shallow indentations on the protein surface, a minimal number of neighbors (mnb) is required for each empty cube to become part of the binding pocket. (III) The minimal cluster size (mnc) gives the number of neighbored grid points that are required for a binding pocket to be recognized as such. In this study, these parameters were set to dob = 11, mnb = 15, and mnc = 50 at a grid spacing of 1.0 Å. Snapshots with pockets in the interface between the two α/β dimers were then identified by visual inspection.

3. Results and discussion

3.1. Expression and purification of Csn2

Csn2 from S. agalactiae ATCC13813 was cloned, overexpressed, and subsequently purified from E. coli BL21(DE3) pLysS cells, via immobilized metal–ion affinity chromatography (IMAC) and size exclusion chromatography (SEC). We obtained 15–20 mg of pure protein per liter of culture. The protein was a tetramer in 10 mM Tris–HCl, pH 8.0, 50 mM NaCl, and 1 mM DTT, as indicated by native PAGE with subsequent staining with coomassie brilliant blue and SEC studies combined with multi angle light scattering (SEC-MALS) experiments (Fig. 1A, low salt run (L)). This technique allows for an accurate molecular mass determination of proteins in solution (for a detailed description of the method see (Slotboom et al., 2008)).

The tetramer of Csn2 appeared to be stable in the range of pH 7.0–9.0. At lower pH values like pH 6.0, Csn2 precipitated (Fig. 1B lane 1–4). Changing the ionic strength by increasing the NaCl concentration up to 500 mM resulted in a shift in the retention time of the Csn2 peak shifting with increasing ionic strength towards a lower molecular weight in SEC studies. However, SEC-MALS revealed that Csn2 still remained a tetramer (Fig. 1A (M,H)), which was also observed in native PAGE analysis using the same samples (Fig. 1B). The shift in retention time can be a result of an interaction with the column matrix, due to the increased

strength of hydrophobic interaction in high salt, or due to a salt dependent large conformational change.

3.2. The structure of Csn2

We solved the structure of Csn2 at 2.0 Å resolution using single wavelength anomalous dispersion (SAD) with a selenomethionine derivative (see Table 1 for data collection and structure refinement statistics). The two molecules of Csn2 occupy in the asymmetric unit, and they further assemble into a tetrameric ring structure with a symmetry related dimer. The protein crystallized in the monoclinic space group C_2 with the following unit cell parameters: $a = 75.3$ Å, $b = 83.3$ Å, $c = 110.4$ Å with an angle of $\beta = 109.40°$ and a solvent content of 59%. Since the structure of both protomers is almost identical (except a slightly different orientation of domain I and II towards each other), the overall structure is described for protomer A only. In promoter B however, one loop region, residues 44–48, showed poor density for the side chains, which were not modelled into the final structure. The overall dimension is roughly 70×70 Å in width and 55 Å in length. Csn2 consists of two distinct domains. Domain I (referred to as the "head" domain) contains the N- and C-terminal portion of the protein (amino acid 1–62 and 140–222), whereas domain II (referred to as the "leg" domain) is composed of amino acids 79–139. The head domain is composed of a six-stranded parallel β-sheet core (β-strands 5, 6, 7, 3, 8 and 9 at the C-terminus, see Fig. 2A); only β-strand 9 runs antiparallel. This core sheet is flanked by one α-helix (H1) and an anti-parallel β-sheet (β-strand 1, 2 and 4) on the one side and two α-helices on the other side (H6 and H8), leading to an overall α/β fold (Fig. 2A). Van der Waals interactions between those helices and the core β-sheet are leading to a compact and stable hydrophobic core. A long flexible linker ($V_{63}–S_{73}$) connects the head domain with the leg domain starting with a small 3_{10}-helix (H2). The leg domain consists of two long α-helices orientated almost perpendicular towards each other (H3 and H4). From the last helix a very long linker ($E_{117}–D_{143}$) reverses back to the head domain. This linker possesses a random coil structure except for a small α-helix (H5), located almost in its center ($T_{126}–G_{135}$). Overall, the Csn2 dimer structure has a "crab-like" shape with the leg domain of one protomer being slightly twisted with respect to the leg domain of the other protomer (Fig. 2A). The dimer interface is solely formed via interactions between the head of both protomers (Fig. 2A). The tetramer of Csn2 as determined in solution (see above) can be observed
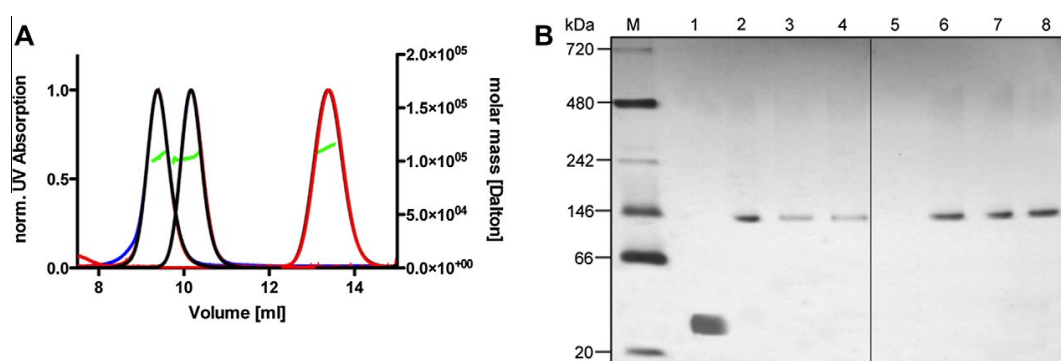
Fig.1. Oligomeric status of Csn2. (A) The elution profile from a Superdex 200 10/300 column at three different sodium chloride concentrations is shown. The lowest NaCl concentration is 50 mM (L), followed by 250 mM (M) and 500 mM NaCl (H). Blue line: normalized UV-Signal; red line: normalized LS-(90°)-Signal; black line: refractive index signal. The green line indicates the corresponding molar mass. Csn2 has a mass of 104.4 ± 0.2 kDa in the peak fraction indicating a homogenous, tetrameric species in solution. (B) Native PAGE (BN PAGE, 4–16% gradient gel) analysis of Csn2 in different NaCl concentration and pH (see Section 2). Lane 1 Csn2 treated with SDS as control showing the protomeric species, lane 2, 50 mM NaCl, lane 3, 250 mM NaCl, lane 4, 500 mM NaCl, lane 5, pH 6, lane 6, pH 7, lane 7, pH 8 and lane 8, pH 9. As can be seen from the gel, also the pH-screen revealed the tetramer except for pH 6 because the protein precipitated caused by its low isoelectric point.
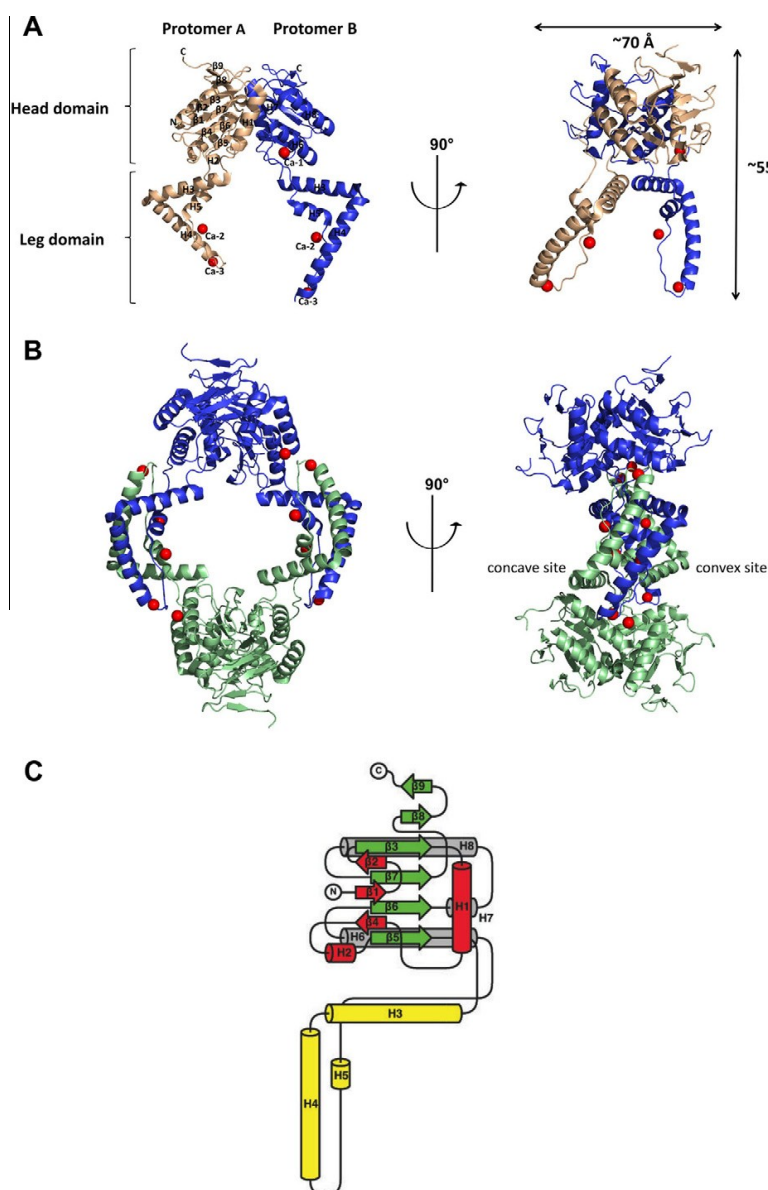
Fig.2. Overall structure of Csn2. (A) Left panel, front view of the Csn2 dimer, which is observed in the asymmetric unit, protomer A is depicted in wheat, protomer 2 in blue. Calcium ions are shown as red spheres. Right panel, side view of the dimer. (B) Left panel, the tetrameric ring structure of Csn2 is shown, which is built by symmetry (see Supplementary information), one dimer is depicted in blue, the other in green. Right panel, side view of the tetramer. Ca^{2+} ions are depicted in red spheres. (C) Topology model of a Csn2 protomer.

within the crystal packing of Csn2. Via a twofold axis a tetramer can be created, the dimers of which tightly interact via H3–4 in the leg domain. These interactions are inducing the stable tetramer as observed in solution. This results in a ring-shaped structure with a hole in the center with a diameter of ∼30 Å (Fig. 2B).

In total, 5 strong electron densities were observed in the dimer (a total of 10 in the tetramer), which were identified as Ca^{2+} ions based on a difference map (F_o–F_c map) analysis using Mg^{2+}-ions resulting in "positive" peaks (green mesh) and Mn^{2+}-ions producing "negative" peaks (red mesh). Only Ca^{2+} ions resulted in an

evident electron density map (see Fig. S2). Further proofs that verify that calcium–ions are bound to Csn2 are distances between the metal center and the coordinating ligands (see Table S1). For validation of the structure a target distance is used and the reported bond distances should be within 0.5 Å of this target sequence (Harding, 2001). We used the target distance of 2.36 Å for calcium (Harding, 2006), and as can be seen from Table S2 all distances are in this range, further verifying that calcium–ions are bound to Csn2. Taken together this suggests that the bound atoms are indeed Ca^{2+} ions.

Two Ca^{2+} ions are located in protomer A and three in protomer B. Since Ca^{2+} ions were not externally added, they were co-purified from *E. coli* cells. Two of the Ca^{2+} ions (Ca-2 and Ca-3) are located at identical positions in the leg domains of both protomers (Fig. S1B), while the third Ca^{2+} ion (Ca-1) is only observed in the head domain of protomer B (Fig. S1A). Due to the lack of this Ca^{2+} ion, the head domain of protomer A is orientated slightly different when compared to protomer B (Fig. 2A, right panel). This slightly different orientation breaks an extra twofold axis within the crystal packing, therefore the symmetry breaks from C_{222} to a lower C_2 symmetry group. The role of these Ca^{2+} ions is to stabilise the orientation of the leg domain. Where Ca-1 and Ca-2 play a role in the orientation of the leg domain Ca-3 is playing a role in the turn at the bottom of the leg domain, which by binding of the Ca-3 ion makes a sharp U-turn like structure.

The importance of this Ca^{2+} ion for stabilizing the conformation of Csn2 explains the slightly twisted conformation, when one is lacking. Adding EDTA or EGTA in the buffer to capture Ca^{2+} ions did not lead to a destabilizing effect (concentrations tested up to 20 mM); the tetrameric conformation was still observed in SEC-MALS (Fig. S8), and no oligomers or aggregates appeared.

A side view of the Csn2 tetramer highlights a different opening width on either side of the hole associated with a different curvature of the leg regions: while the concave side shows an opening of ~35 Å, the one of the convex side is ~45 Å (Fig. 2B, right panel). This difference might indicate a specific role for each side of the tetramer; however, it might also be induced by the lack of one Ca^{2+} ion in protomer B.

Recently, the structure of Csn2 from *E. faecalis* was solved at 2.7 Å resolution revealing a similar tetrameric architecture (Nam et al., 2011). Both proteins exhibit a protein sequence identity of 57% (76% similarity). The four protomers form a ring like structure with a central hole in the center. In this structure, Ca^{2+} ions were also observed. However, they had been added externally prior to crystallization. In contrast to our Csn2 structure, the leg domain of each protomer has the same orientation leading to a symmetrical structure. In general, the overall structures as well as the positions of the secondary structure elements are the same in both the *S. agalactiae* and *E. faecalis* structures. However, in contrast to our analysis, the addition of EDTA or EGTA to the Csn2 protein from *E. faecalis* leads to the formation of higher oligomers, or even aggregates, thus indicating a less stable tetramer when depleted of Ca^{2+} ions.

3.3. The head domain of Csn2 exhibits a helicase fold

To shed light on the function of Csn2, we performed a structural comparison search using the DALI server (Holm and Rosenstrom, 2010). The comparison reveals that the head domain of Csn2 contains a helicase fold (Fig. S2), similar to the one observed in the chromosome segregation SMC protein (Lowe et al., 2001), the replicative helicase-primase of bacteriophage T7 (Sawaya et al., 1999), as well as the helicase domain observed in nucleotide-binding domains of ABC proteins such as RAD50 (Lammens et al., 2011) or ABC transporters (Hollenstein et al., 2007). On a sequence level, these proteins are only 6–13% identical; in contrast, the high structural similarity is highlighted by an overall RSMD of 3–4 Å over 220 Cα atoms (Table S2).

Generally, helicases catalyze the separation of duplex polynucleotides into single strands in an ATP-dependent manner (Hall and Matson, 1999). Helicases contain some characteristic sequence motifs such as the classical Walker 'A' (phosphate binding loop or P-loop) and Walker 'B' (involved in Mg^{2+}-binding through an aspartic acid) motifs (Walker et al., 1982). Furthermore, Caruthers and McKay classified different helicase superfamilies based on sequence homology (Caruthers and McKay, 2002). According to this,

Csn2 belongs to the DnaB-related helicases, a small subgroup involved in DNA replication. However, in Csn2, the sequence corresponding to the Walker 'A' motif, $_{23}IEDVSVYS_{30}$, deviates from the consensus sequence GxxGxGKS/T (where x can be any amino acid) (Walker et al., 1982). Here especially the lysine residue has been shown to be important which is replaced by a tyrosine at position 29 in Csn2. A similar degeneration is observed for the Walker 'B' motif, where the sequence is $_{164}LLVFV_{168}$ in Csn2 instead of the consensus sequence hhhhD (where h is a hydrophobic amino acid). So the aspartate residue important for the Mg^{2+} coordination has been mutated to a valine. This degeneration of Walker 'A' and Walker 'B' is unusual, but observed in other helicase domains, too. For instance, the Mfd (Mfd-N) protein from *E. coli*, which is involved in transcription-coupled repair, contains degenerated Walker 'A' and Walker 'B' motifs (Assenmacher et al., 2006).

Helicases typically use the energy of ATP hydrolysis to unwind DNA. We could not detect hydrolysis or helicase activity with Csn2 using ATP and GTP (data not shown). Therefore, we concluded that Csn2 is not exhibiting the putative helicase activity thought to be present in the CRISPR/Cas system (Makarova et al., 2006). When aligning Csn2 protein sequences, a highly conserved DxD motif at the C-terminal part was identified (Fig. S4). This motif is conserved in DNA and RNA polymerases or in toprim domains found in bacterial primases and topoisomerases (Frick and Richardson, 2001). Structurally, the DxD motif in Csn2 is located at the top of the head domain (Fig. 3): D_{210} and D_{212} reside in a loop connecting β8 and β9 and are in close proximity to E_{24}. These three residues can maybe build up a acidic triad, which binds metal ions like the two-metal ion mechanism of polymerases (Steitz, 1999). Unfortunately, the mutation of any of the amino acids of the $_{210}DxD_{212}$ and E_{24} motif resulted in inclusion body formation, which could not be refolded (data not shown). Therefore, we were not able to assess the exact role of any of these amino acids *in vitro*.

A flexible loop containing a RR-motif ($_{199}R–R_{200}$) is located near the DxD motif (Fig. 3). In polymerases, such as the polymerase domain (PolDom) of the NHEJ DNA polymerase of *M. tuberculosis*, an arginine residue is also present (Brissett et al., 2011). This arginine interacts with one aspartate residue of the above-mentioned acidic triad, resulting in what is called the "pre-ternary" complex. The binding of either an Mg^{2+} ion or the nucleotide starting the polymerase reaction interrupts this interaction. In Csn2, R_{200} interacts with E_{24} of the acidic triad, which therefore cannot participate in Mg^{2+} binding, suggesting that our structure may also represent a "pre-ternary" complex. Such a role of R_{200} is likely conserved, because all Csn2 proteins analyzed contain at least one arginine residue at this position (Fig. S4). Interestingly, the RR-motif is forming an overall positively charged groove on the top of the head (Fig. S7). Although assayed, a defined polymerase activity was not observed during our experiments (data not shown).

3.4. Csn2 binds ds-DNA

Csn2 is capable of binding double stranded (ds)-DNA (Fig. 4). With increasing concentration of Csn2 higher Csn2-DNA-complexes are formed (Fig. 4A, lane 7). External addition of Ca^{2+} is not required, indicating that the co-purified Ca^{2+} ions are sufficient for function (Fig. 4B, lane 7). These Ca^{2+} ions are essential because removing them by EGTA resulted in a loss of ds-DNA binding (Fig. 4B, lanes 5–6), showing their importance for binding.

The RR-motif forms a groove on the top of the head domain as described above (Fig. 3). Furthermore, the RR-motif is located in a loop connecting β7 and β8 (close to the degenerated Walker 'A' and 'B' motifs). It is thus tempting to speculate that the RR-motif is interacting with the ds-DNA template. To further characterize this RR-motif, we mutated both R residues of the RR-motif into alanine
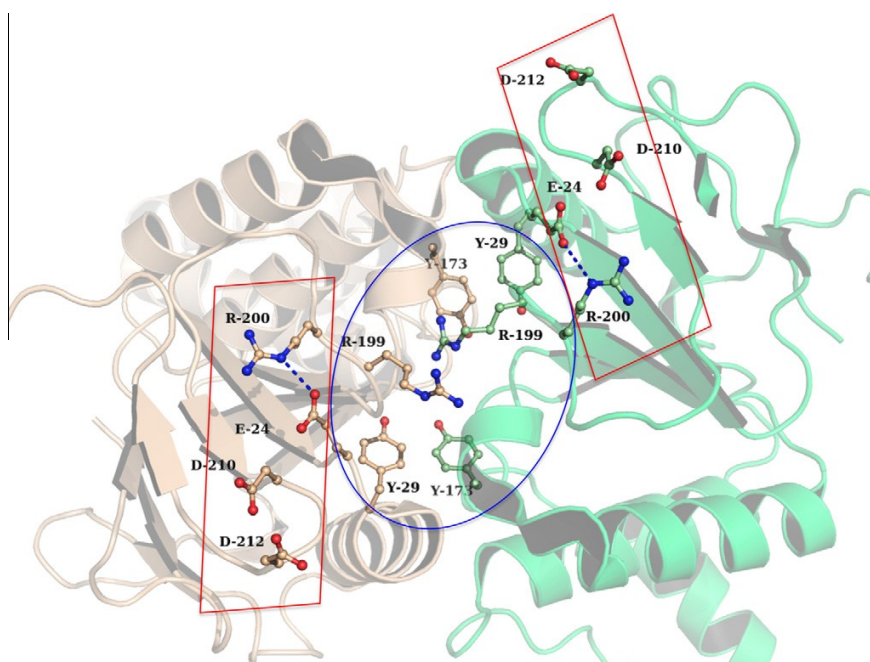
Fig.3. Top view onto the Csn2 head region. The head domain of protomer A is colored in gray, protomer B in green. Important residues are shown in ball and sticks. D_{210}, D_{212} and E_{24} are thought to form up an "acidic triad". Also, two tyrosines (Y_{29} and Y_{173}) of each protomer are shown.
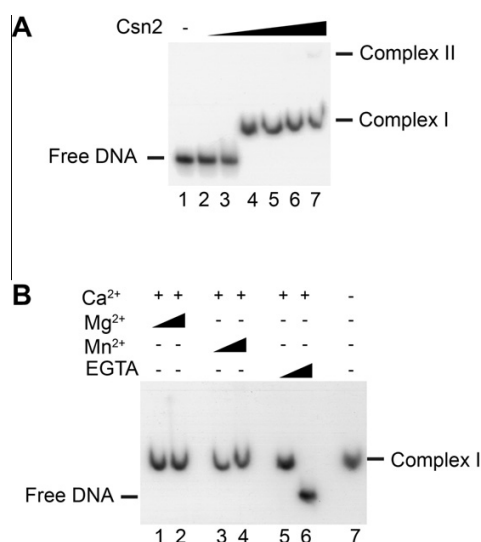
Fig.4. DNA binding to Csn2. (A) 2 nM radiolabeled 256 bp ds-DNA was incubated with increasing amounts of Csn2 protein: 0 µM (lane 1), 0.1 µM (lane 2), 0.2 µM (lane 3), 0.5 µM (lane 4), 1 µM (lane 5), 2 µM (lane 6) and 4 µM (lane 7) Csn2 in presence of 20 mM Ca^{2+} at 37 °C for 20 min and separated on a 5% native PAA. (B) 1 µM Csn2 were incubated with 2 nM ds-DNA in presence of Mg^{2+}, Mn^{2+} or EGTA: Lanes 1, 2: 10 mM or 20 mM Mg^{2+}, lanes 3, 4: 10 mM or 20 mM Mn^{2+}, lanes 5,6: 10 mM or 20 mM EGTA, lane 7: without addition of ions or EGTA.

Fig.5. Mutational analysis of Csn2 in DNA binding activity of Csn2. 2 nM radiolabeled 449 bp ds-DNA was incubated with 4 µM of indicated Csn2 variants at 37 °C for 20 min without an addition of external Ca^{2+} and separated on a 5% native PAA.

and tested DNA binding. The mutant of Csn2 is still able to bind ds-DNA, like wildtype Csn2. Furthermore, the mutation of Y_{29} into either alanine or tryptophan does not seem to interfere with DNA binding either (Fig. 5). This implies that the groove found at the top of the head is not participating in DNA binding of Csn2 and, thus, likely serves another purpose.

K_{132} is another residue that is conserved among all Csn2 proteins. K_{132} is located in the leg region, in the middle of the small helix H5 (for an alignment see Fig. S4). Interestingly, when mutated to alanine, the DNA binding activity is lost completely (Fig. 5). Nam et al. proposed that the DNA binds through the center of the ring of the tetramer (Nam et al., 2011). At first view, this suggestion is at variance with the fact that Csn2 exists as a stable tetramer in solution (see above), without separation into dimers, which should be necessary to embrace circular enclosed genomic DNA through the hole. It is possible however, that prior to the

binding of Csn2 free DNA ends need to be generated, for example by the nuclease activity of Cas1. Indeed, the nuclease activity of Cas1 has been shown to be essential for uptake of new spacer sequences in *E. coli* (Yosef et al., 2012).

The gel shift assays demonstrate that Csn2 binds to ds-DNA in absence of Ca^{2+} chelating agent EGTA. Mutations of Y29 or the RR-motif to alanine do not interfere with the DNA-binding activity of Csn2, indicating that the grove on the head domain is not involved in DNA interaction. The different mobility of the Y29W mutant compared to the wildtype protein is possibly due to a different conformation of the Csn2-DNA complexes or to unspecific occupation of multiple binding sites on the DNA (Fig. 5, lanes 2, 4 and 6). Very low quantities of such low mobility complexes can also be observed with the wildtype Csn2 at high protein concentration (Fig. 4A, lane 7). In contrast, replacement of the highly conserved K132 residue by alanine completely abolished the DNA-binding activity (Fig. 5, lane 5). The loss of the DNA-binding activity of this mutant was not caused through limitation of the co-purified Ca^{2+} ions in the preparation, since addition of external Ca^{2+} did not restore its binding to the DNA (data not shown). In summary, our study supports the conclusion that DNA binding site is not located in the head domain, but somewhere near the hole. To corroborate this notion, we performed molecular dynamics calculations.

3.5. Large conformational changes observed by molecular dynamics simulations reveal a potential ds-DNA binding site

An all-atom molecular dynamics (MD) simulation of 100 ns length of the tetrameric structure suggests an alternative potential binding site for ds-DNA. The simulation reveals a drastic movement ("closing") of the head region of one dimer with respect to the head region of the other dimer (Fig. 6). This is demonstrated by root mean square deviations (rmsd) with respect to the crystal structure of up to 35 Å observed for one head region, when aligning the structures with respect to the other head region. In contrast, conformational changes within each of the head regions and the legs are moderate (rmsd < 4.5 Å; Fig. S5). Within the initial 70 ns of the simulation, the head regions approach each other repeatedly (rmsd up to 25 Å; Fig. 6) but always return to an "open" configuration

similar to the one in the crystal structure (rmsd ≈ 5 Å for all amino acids) before undergoing the drastic closing in the final 30 ns of the simulation. The periodic movements thus reveal an inherent tendency of the tetrameric structure to return to the open configuration. The movements are also reflected in the breaking and reformation of salt bridges between K_{137} and E_{121} in the hinge region (Fig. 7); these salt bridges largely get lost when the head regions approach each other (blue and green curves in Fig. 7C). This breaking may contribute to the conformational strain that fosters an open configuration in the absence of ds-DNA. In contrast, salt bridges between K_{91} and $E_{112'}$ stay intact throughout the closing movement (red and magenta curves in Fig. 7C), that way contributing to the stability of the tetramerization interface and, in particular, the orientation of H4 and H4' on the convex and concave leg sides. This is also corroborated by the location of a hinge axis as identified by DynDom (Hayward and Berendsen, 1998), which passes through the centers of the two pairs of helices H4 and H4' (Fig. 6B).

The closing occurs by bending motions between H3 and H4 (Fig. 7B and D) and within a region located in between the head and the leg domains, as demonstrated by varying orientations of H4' and H6 (Fig. 7B and E). The convex and concave sides of the tetramer behave differently in this case, both with respect to the direction of the change and, in particular, its magnitude. As such, the main motion is enabled by an increase of the splaying angle between H3 and H4 and a decrease of this angle between H4' and H6, for helices located on the concave side (Fig. 7B; blue and green curves in Fig. 7D and E). In contrast, the orientation between these helices on the convex side, which is presumably involved in ds-DNA binding (see below), changes only moderately (Fig. 7B; red and magenta curves in Fig. 7D and E). Both sides of the leg domain are also asymmetric with respect to the closing motion itself: The closing only occurs in the direction of the convex side; in contrast, no overstretching of the tetramer in the direction of the concave side has been observed during the MD simulation (data not shown).

By the closing movement, the 3_{10} helices H2 on the convex side of the legs get as close together as 33 Å, starting from 45 Å in the crystal structure (Fig. 6B). In contrast, the width of the hole does not change, as measured between the helices H5 on opposite sides
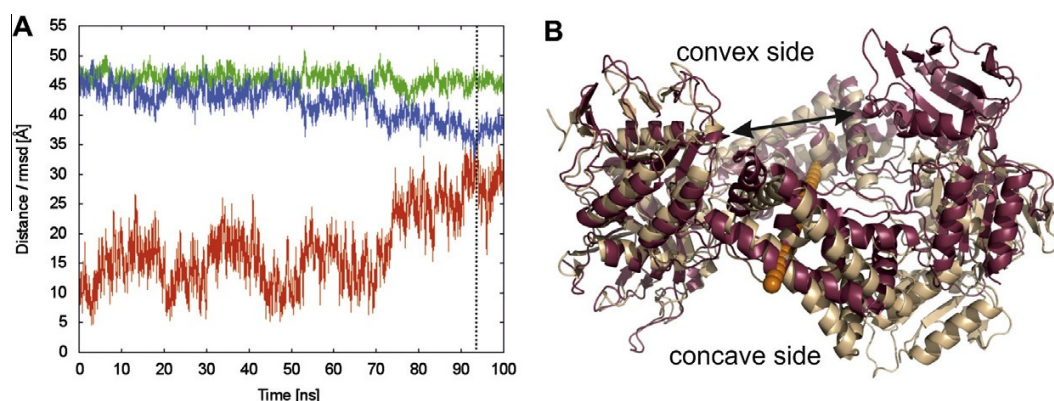
Fig. 6. Closing motion of tetrameric Csn2 observed by MD simulation. (A) Red curve: Rmsd of Cα atoms of one head region of the tetrameric Csn2 structure with respect to the crystal structure; the structures were aligned with respect to the other head region. Blue curve: Distance between the average coordinates of the Cα atoms of the two 3_{10} helices located in the head domains of the tetrameric Csn2 structure on the convex side. See arrow in (B). Green curve: Distance between the average coordinates of the Cα atoms of the two pairs of H5 helices; each pair of helices is located in one of the legs of the tetrameric structure. The dashed line indicates the time point at which the conformation shown in (B) is extracted. (B) Side view of an overlay of the tetrameric Csn2 structure obtained by X-ray crystallography (wheat) with a conformation extracted from the MD trajectory after 94 ns that shows the most pronounced closing movement (raspberry); the structures were aligned with respect to one head region. The arrow indicates the 3_{10} helices between which the distance shown in (A) is measured. The line of orange spheres indicates the hinge axis as identified by DynDom (Hayward and Berendsen, 1998).
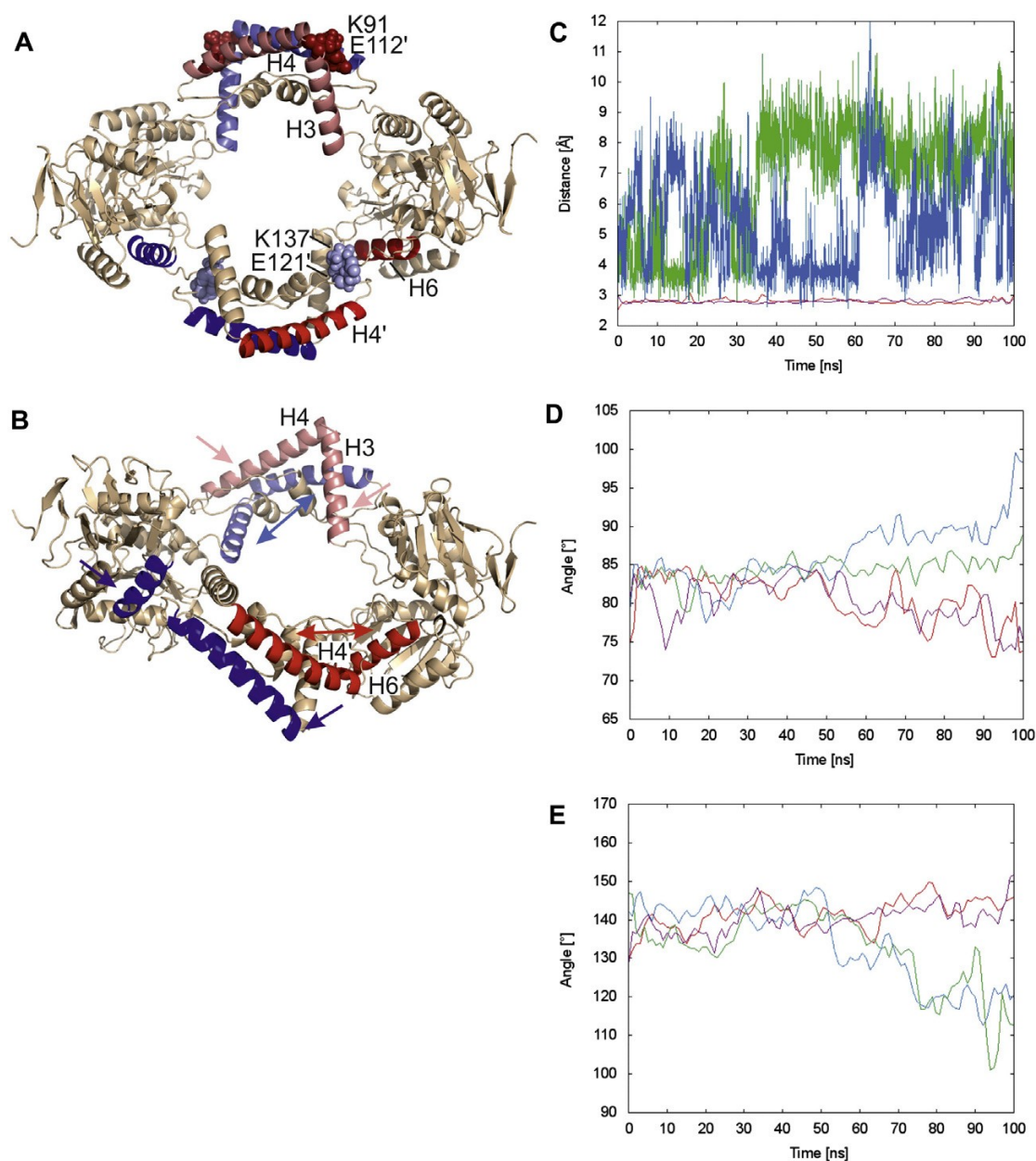
Fig.7. Structural parameters of the tetrameric Csn2 structure during the MD simulation. (A) Structural parameters analyzed in panels (C–E). The tetramer is oriented with the convex side to the viewer. Helix pairs are depicted in identical colors, as are pairs or residues that form a salt bridge. Redish colors mark those objects whose parameters vary little in panels (C–E) whereas bluish colors mark those objects whose parameters vary strongly in panels (C–E). Unprimed labels mark residues and helices that belong to one dimer, primed labels mark objects that belong to the other dimer. (B) Schematic representation of the changes of the helix pair angles as analysed in panels (D–E) during the closing motion. The tetramer is shown in a slanted view from the convex side with helix pairs depicted in identical colors. *Two arrows* indicate a decrease of the helix pair angle; a *double arrow* indicates an increase of the helix pair angle. (C) Minimal distance between side chain oxygens of Glu and the side chain nitrogen of Lys that can form salt bridges in the leg domain. Only four representative pairs out of the total of eight are shown (see panel (A)). A Bezier smoothing was applied to the curves representing interactions between K91 and E112′. (D) Angle between H3 and H4 of either leg domain as defined by the point triple (average coordinates of Cα atoms of H3; Cα atom of K91; average coordinates of Cα atoms of H4). Red and magenta curves refer to the two helix pairs on the convex side of the tetramer, blue and green curves to the two helix pairs on the concave side. A Bezier smoothing was applied to the curves. (E) Angle between H4 and H6 as defined by the point triple (average coordinates of Cα atoms of H4; average coordinate of Cα atoms of C-terminal residue of H4 and N-terminal residue of H6; average coordinates of Cα atoms of H6). Red and magenta curves refer to the two helix pairs on the convex side of the tetramer, blue and green curves to the two helix pairs on the concave side. A Bezier smoothing was applied to the curves.

of the hole. Notably, due to the closing, a *U*-shaped groove emerges, of which the base is formed by helices H3 and H4 and the sides by helices H1 and H2 of protomers A and A*. This groove

has a size that can accommodate a ds-DNA, and the groove-forming regions are strongly positively polarized (Fig. S7). Together this suggests that ds-DNA binds into this groove, with the helix

axis oriented *perpendicular* to an axis that runs through the ring (Fig. 9).

3.6. Identification of transient pockets in the interface of the head region

The presence of the specific DxD and RR-motif in the top of the head domain in combination with the proposed location of the ds-DNA binding site at the hole of the tetramer raises the question whether processes that occur at these two locations can communicate with each other. In order to analyze whether such interference could be mediated by the head domain structures, we performed a MD simulation of only the head region of Csn2 (see Section 2), which allowed for a more comprehensive sampling of 200 ns length compared to the tetrameric structure. Over the course of the trajectory, the structural changes in the head region are almost always <3.5 Å rmsd (Fig. S6) and, thus, are smaller than the deviations observed for the tetrameric structure. Surprisingly, however, conformations obtained after about 150 ns and almost 200 ns show pronounced indentations in the regions of the DxD motif, Y_{29}, and R_{199} ("indentation I") and the loop that connects β-strand 2 and H3 ("indentation II"), respectively (Fig. 8). Notably, these indentations overlap in the region between helices H7. When combined, the indentations thus form a tunnel that reaches from close to the DxD motif to the loop that connects β-strand 2 and H3 in the vicinity of the hole. Unfortunately, a continuous tunnel of that kind could not be observed in a single conformation generated by MD, which points to still a too limited sampling time. For the indentations to open up, mostly changes in the rotameric states of the side chains are sufficient. As such, R_{199} points to the solvent
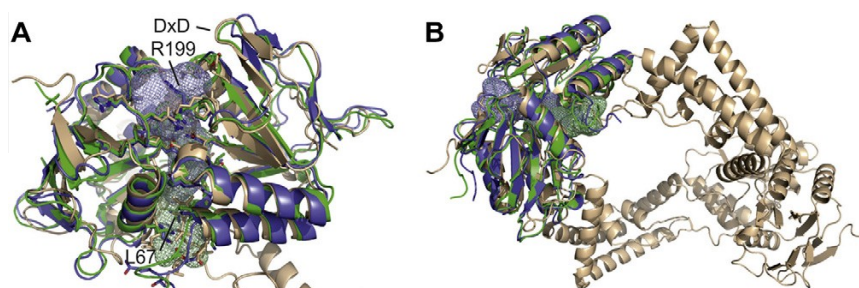
Fig. 8. Opening of transient pockets in the interface between the head domains. (A) Overlay of the head region of the crystal structure (wheat) and conformations extracted from the MD simulation of the head region after ∼150 ns (green) and ∼200 ns (blue) together with respectively colored mesh representations of pockets identified by PocketAnalyzer[PCA]. Sidechains of residues that are within 3 Å of the detected pockets are depicted as sticks. (B) Same as in (A), but now the full tetrameric structure is shown with a slanted view onto the convex side of the legs.
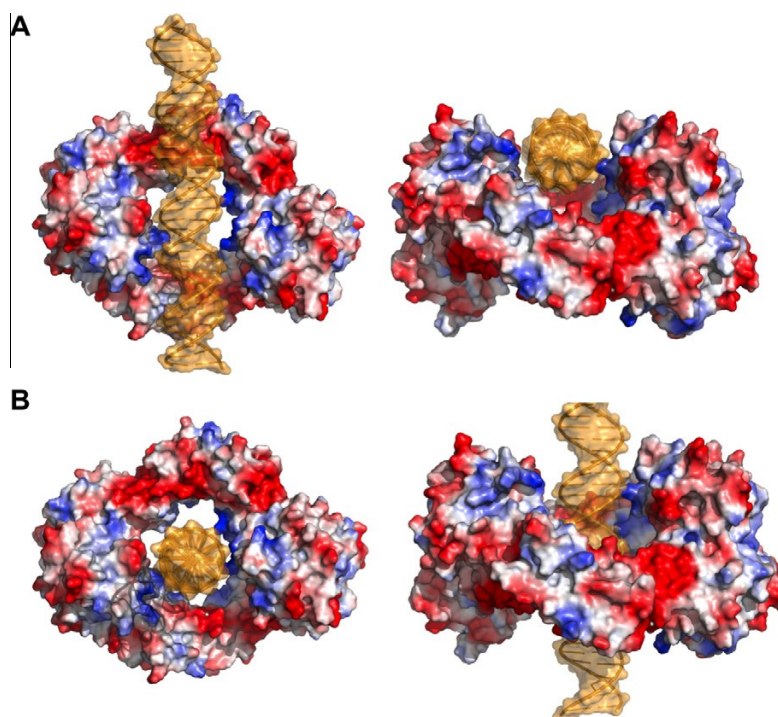
Fig. 9. Two possible binding sites of ds-DNA with Csn2. The structure of Csn2 is used as observed in the molecular simulation study (see Fig. 6). (A) Proposed binding site, where Csn2 "sits" on the ds-DNA. Used is a stretch of ideal ds-DNA (colored in light yellow). The surface of Csn2 is colored on the basis of the electrostatic potential contoured from −3 kT/e (red) to +3 kT/e (blue). White denotes 0 kT/e. The surface potential was calculated using APBS tools. (B) The ds-DNA binds through the hole within the tetramer.

in the conformations that lead to indentation I, whereas in particular L_{67} moves to open indentation II. Overall, these analyses suggest that a transient tunnel may open in the interface region of the head domain. The frequency of occurrence of conformations with such indentations during the MD simulation (~0.5%) suggests that these conformations are energetically accessible at 300 K.

4. Conclusion

Here, we report the crystal structure of Csn2 from *S. agalactiae* at 2.0 Å resolution. Csn2 is a stable tetramer in solution as determined by SEC-MALS and is forming a ring-like structure.

Recently, the structure of Csn2 of *E. faecalis* was solved at 2.7 Å resolution. Like Csn2 from *S. agalactiae*, the structure revealed a tetramer, forming a similar ring-like structure with a hole in the middle (Nam et al., 2011). In the structure of *E. faecalis* Csn2, Ca^{2+}-ions were observed at the same positions as in the structure reported here. This protein is only able to bind ds-DNA in the presence of Ca^{2+} (Nam et al., 2011). We also observed ds-DNA binding by Csn2, however, the binding is not depending on external addition of Ca^{2+} ions due to the co-purification of the divalent ion. These Ca^{2+} ions are important in that the addition of EGTA depleted Csn2 from ds-DNA binding although the tetramer is still stable as observed in SEC-MALS experiments. In contrast, a similar experiment with Csn2 of *E. faecalis* resulted in the formation of higher oligomers. This implies that for *S. agalactiae* Csn2 the Ca^{2+} ions are merely required to orientate the protein in a conformational state that allows the binding of ds-DNA.

Csn2 of *S. agalactiae* binds ds-DNA, and mutational analysis revealed that the RR-motif as well as the degenerated Walker 'A' motif both located in the head domain are not involved in ds-DNA binding. This demonstrates that the helicase fold present in the head is serving another activity. In contrast, the highly conserved K_{132}, part of a small helix in the leg region, is involved in ds-DNA binding because mutation to alanine resulted in a complete loss of DNA binding capacity.

Nam et al. suggested that DNA binds within the hole of the tetramer, which contains a substantial number of conserved positively charged residues (especially lysines) that are also present in Csn2 from *S. agalactiae* (Fig. 9A and Fig. S4). At variance with this, MD simulations performed in the present study suggest another ds-DNA binding site. For this, the head domain of one dimer within the tetramer undergoes a drastic movement and comes as close as 35 Å to the head domain of the other dimer, forming a *U*-shaped groove, which would allow binding of ds-DNA with a diameter of ~26 Å (Fig. 9B). That way, the binding site of ds-DNA would be on one side of Csn2, which implies that the hole in the middle of the tetramer can serve another purpose. Interestingly, the ds-DNA would be located exactly on the third twofold axis of the broken C_{222} symmetry, identical to what has been observed for other DNA binding proteins.

As such, the binding of ds-DNA could create a surface for the binding of yet another protein within the hole, which is required for the integration of the spacers, such as an integrase and/or a ligase protein. Both DNA binding modes ("sitting" or "going through") are still putative because until now no experimental evidence is available in favour of one or the other. We do not think that mutational studies located at the hole can reveal the exact mechanism because mutational effects will likely be not conclusive. For instance, the $K_{132}A$ mutation abolishes DNA binding, but visual inspection of the structure shows that this could happen in the case of either binding mode. Co-crystallization of Csn2 with DNA is likely the only method revealing the exact binding mode.

At present, the exact sequence of the DNA that binds to Csn2 is not known. The sequence of the co-purified DNA in the *E. faecalis*

Csn2 revealed no CRISPR-specific DNA sequence indicating an unspecific binding (Nam et al., 2011).

The head domain of the Csn2 structure adopts a helicase fold with DxD- and RR-motifs. However, the sequence motifs that are defining a helicase (Hall and Matson, 1999) are degenerated as revealed by a sequence alignment of Csn2 proteins: all proteins contain degenerated Walker 'A' and Walker 'B' motifs. This is likely the reason why the function of Csn2 could not be predicted *a priori* using bioinformatics programs. The DxD motif as well as the RR-motif are conserved in DNA and RNA polymerases (Steitz, 1998) as well as in toprim domains found in primases, (Frick and Richardson, 2001) and in nucleases of the OLD-family (Aravind et al., 1998). MD simulations of only the head domain of Csn2 suggest that processes occurring at the location of the DxD and RR-motifs as well as at the proposed location of the ds-DNA binding site could communicate with each other via a transient tunnel that can open in the interface region of the head domain.

Unfortunately, no polymerase activity of Csn2 could be identified in this study. Therefore, the exact roles of the DxD and RR-motifs remain unknown. Since the first experimental evidence of CRISPR as a novel bacterial defense system against foreign DNA was published (Barrangou et al., 2007), several studies have shed light on the mechanisms of this prokaryotic immune system. Common to all CRISPR/Cas systems are Cas1 and Cas2 functions during the immunization stage. These two proteins have been analyzed biochemically and structurally, demonstrating that both contain nuclease activities *in vitro* (Babu et al., 2011a; Beloglazova et al., 2008; Wiedenheft et al., 2009). Their precise role in the adaptation process remains unclear, however. In addition to these two proteins, the subtype-specific Cas proteins Csn2 and Cas4 are very likely involved in adaptation processes. Moreover, there is evidence that a unique CRISPR repeat sequence correlates with the presence of Csn2 (Kunin et al., 2007), suggesting a role in the adaptation process. However, binding affinity of Csn2 to repeat-spacer-repeat sequences are the same as compared to random DNA of similar length (data not shown).

Interestingly, the Csn2 protein family is only present in a unique subclass of the CRISPR family. Yet, the DxD as well as RR-motifs are present also in other operons. For example, Cas1 from *E. coli* contains a RR-motif ($_{131}RR_{132}$) as well as a DxD motif ($_{108}DxD_{110}$) structurally located next to each other (see Fig. S7). This raises the question whether in the CRISPR Nmeni-subtype the role of these motifs is mediated by Csn2 whereas in the other subtypes it is mediated by the other Cas and/or other proteins within the operon. Cas1 of *E. coli* has been characterized and exhibits a nuclease activity (Babu et al., 2011b). The described DxD- and RR-motifs are not located at the active site responsible for the nuclease activity. This indicates that Cas1 might have additional activities.

Taken together, the exact role *in vivo* of Csn2 within the CRISPR system is not yet known. However, recent studies describing the activity of the CRISPR operon of *S. thermophilus* located on a plasmid in *E. coli* (Sapranauskas et al., 2011) might be helpful to elucidate this role. Especially mutating single amino acids (i.e. in the DxD and RR-motifs) might shed light on the function of the head domain as well as the whole Csn2 protein and the other Cas proteins.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2012.04.006.

References

Al-Attar, S., Westra, E.R., van der Oost, J., Brouns, S.J., 2011. Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. Biol. Chem. 392, 277–289.

Aravind, L., Leipe, D.D., Koonin, E.V., 1998. Toprim–a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res. 26, 4205–4213.

Assenmacher, N., Wenig, K., Lammens, A., Hopfner, K.P., 2006. Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs. J. Mol. Biol. 355, 675–683.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E.V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., Yakunin, A.F., 2011a. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484–502.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E.V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A.M., Yakunin, A.F., 2011b. A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Mol. Microbiol. 79, 484–502.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709–1712.

Beloglazova, N., Brown, G., Zimmerman, M.D., Proudfoot, M., Makarova, K.S., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E.V., Edwards, A.M., Savchenko, A., Yakunin, A.F., 2008. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J. Biol. Chem. 283, 20361–20371.

Brissett, N.C., Martin, M.J., Pitcher, R.S., Bianchi, J., Juarez, R., Green, A.J., Fox, G.C., Blanco, L., Doherty, A.J., 2011. Structure of a preternary complex involving a prokaryotic NHEJ DNA polymerase. Mol. Cell 41, 221–231.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964.

Caruthers, J.M., McKay, D.B., 2002. Helicase structure and mechanism. Curr. Opin. Struct. Biol. 12, 123–133.

Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz Jr., K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., 2005. The Amber biomolecular simulation programs. J. Comput. Chem. 26, 1668–1688.

Chen, V.B., Arendall III, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21.

Cornell, W.D., Cieplak, C.I., Bayly, I.R., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., Kollman, P.A., 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. 117, 5179–5197.

Craig, I.R., Pfleger, C., Gohlke, H., Essex, J.W., Spiegel, K., 2011. Pocket-space maps to identify novel binding-site conformations in proteins. J. Chem. Inf. Model 51, 2666–2679.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald - A Nlog(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.

Deveau, H., Garneau, J.E., Moineau, S., 2010. CRISPR/Cas system and its role in phage–bacteria interactions. Annu. Rev. Microbiol. 64, 475–493.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132.

Frick, D.N., Richardson, C.C., 2001. DNA primases. Annu. Rev. Biochem. 70, 39–80.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., Moineau, S., 2010. The CRISPR/Cas

bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.

Grissa, I., Vergnaud, G., Pourcel, C., 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinform. 8, 172.

Haft, D.H., Selengut, J., Mongodin, E.F., Nelson, K.E., 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. 1, e60.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., Terns, M.P., 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956.

Hall, M.C., Matson, S.W., 1999. Helicase motifs: the engine that powers DNA unwinding. Mol. Microbiol. 34, 867–877.

Harding, M.M., 2001. Geometry of metal–ligand interactions in proteins. Acta Crystallogr. D Biol. Crystallogr. 57, 401–411.

Harding, M.M., 2006. Small revisions to predicted distances around metal sites in proteins. Acta Crystallogr. D Biol. Crystallogr. 62, 678–682.

Hayward, S., Berendsen, H.J., 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 30, 144–154.

Hendlich, M., Rippmann, F., Barnickel, G., 1997. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J. Mol. Graphics Modell. 15 (359–363), 389.

Hollenstein, K., Frei, D.C., Locher, K.P., 2007. Structure of an ABC transporter in complex with its binding protein. Nature 446, 213–216.

Holm, L., Rosenstrom, P., 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545–W549.

Jansen, R., Embden, J.D., Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43, 1565–1575.

Jorgensen, W.L., Chandrasekhar, J., Madura, J., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935.

Kabsch, W., 1993. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 26, 795–800.

Karginov, F.V., Hannon, G.J., 2010. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol. Cell 37, 7–19.

Kunin, V., Sorek, R., Hugenholtz, P., 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. 8, R61.

Lammens, K., Bemeleit, D.J., Mockel, C., Clausing, E., Schele, A., Hartung, S., Schiller, C.B., Lucas, M., Angermuller, C., Soding, J., Strasser, K., Hopfner, K.P., 2011. The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. Cell 145, 54–66.

Langer, G., Cohen, S.X., Lamzin, V.S., Perrakis, A., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171–1179.

Lowe, J., Cordell, S.C., van den Ent, F., 2001. Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. J. Mol. Biol. 306, 25–35.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., van der Oost, J., Koonin, E.V., 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. 9, 467–477.

Marraffini, L.A., Sontheimer, E.J., 2010a. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat. Rev. Genet. 11, 181–190.

Marraffini, L.A., Sontheimer, E.J., 2010b. Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463, 568–571.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733–740.

Murshudov, G., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D 53, 240–255.

Nam, K.H., Kurinov, I., Ke, A., 2011. Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca²⁺-dependent double-stranded DNA-binding activity. J. Biol. Chem.

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., Tucker, P.A., 2005. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Crystallogr. D Biol. Crystallogr. 61, 449–457.

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S., Tucker, P.A., 2009. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. Acta Crystallogr. D Biol. Crystallogr. 65, 1089–1097.

Pul, U., Wurm, R., Wagner, R., 2007. The role of LRP and H-NS in transcription regulation: involvement of synergism, allostery and macromolecular crowding. J. Mol. Biol. 366, 900–915.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., Wagner, R., 2010. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol. Microbiol. 75, 1495–1512.

Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C., 1977. Numerical-integration of Cartesian equations of motion of a system with constraints – molecular dynamics of n-alkanes. J. Comput. Phys. 23, 327–341.

Samai, P., Smith, P., Shuman, S., 2010. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. Acta Crystallogr. Sec. F: Struct. Biol. Crystall. Commun. 66, 1552–1556.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. Nucleic Acids Res.

Sawaya, M.R., Guo, S., Tabor, S., Richardson, C.C., Ellenberger, T., 1999. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. Cell 99, 167–177.

Simmerling, C., Strockbine, B., Roitberg, A.E., 2002. All-atom structure prediction and folding simulations of a stable protein. J. Am. Chem. Soc. 124, 11258–11259.

Slotboom, D.J., Duurkens, R.H., Olieman, K., Erkens, G.B., 2008. Static light scattering to characterize membrane proteins in detergent solution. Methods 46, 73–82.

Steitz, T.A., 1998. A mechanism for all polymerases. Nature 391, 231–232.

Steitz, T.A., 1999. DNA polymerases: structural diversity and common mechanisms. J. Biol. Chem. 274, 17395–17398.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., Brouns, S.J., 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. 34, 401–407.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S.M., Ma, W., Doudna, J.A., 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17, 904–912.

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res., http://dx.doi.org/10.1093/nar/gks216.

Supplementary information

The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*

Philipp Ellinger[a], Zihni Arslan[b], Reinhild Wurm[b], Britta Tschapek[a], Colin MacKenzie[c], Klaus Pfeffer[c], Santosh Panjikar[e, 1], Rolf Wagner[b], Lutz Schmitt[a], Holger Gohlke[d,*], Ümit Pul[b,*], Sander H.J. Smits[a,*]

Supplementary information comprises:

Supplementary Figures 1-8

Supplementary Table 1-2

Supplementary References

1 Present Address: Australian Synchrotron, 800 Blackburn Road, Clayton, VIC, Australia

Supplementary Figures

Figure S1:
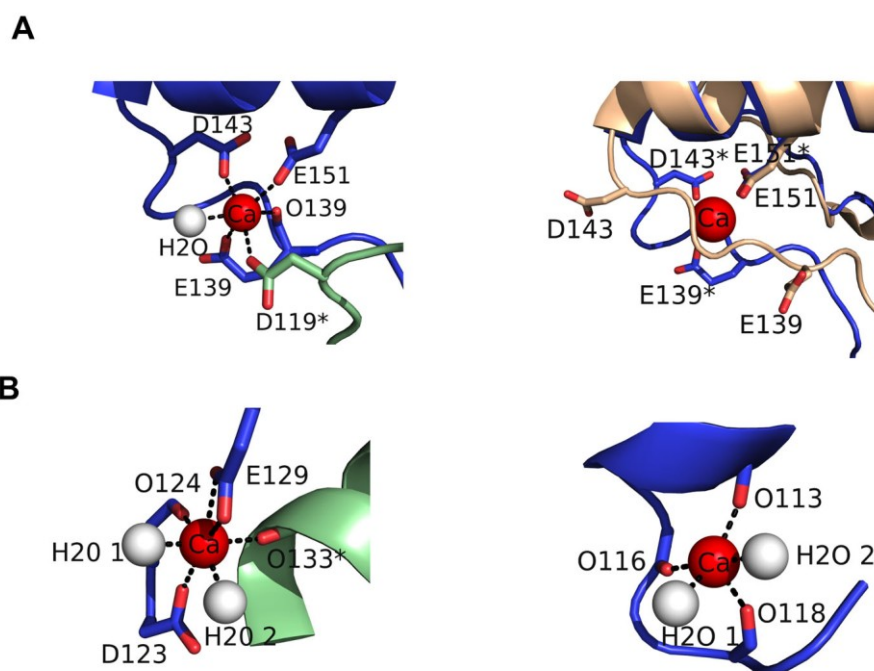
Ca^{2+} binding sites of Csn2

A Left panel: Calcium-ion 1 is shown as red sphere, which is only found in domain I of protomer A **Right panel:** Superimposition of monomer B with monomer A clearly shows that the loop leading to helix 6 (H6) in monomer A is not as wide opened as in monomer B and E_{139} and D_{143} are pointing away, preventing binding of calcium at this site in monomer A. The Ca^{2+}-ion is interacting with amino acids E_{139}, D_{143}, E_{151} (blue cartoon) and one water molecule in monomer B, as well as with D_{119} from monomer A of the other dimer (palegreen). **B** Binding sites for Ca^{2+}-ion 2 (left panel) and Ca^{2+}-ion 3 (right panel) are shown. (For interaction partners and distances see Table S1)

Figure S2
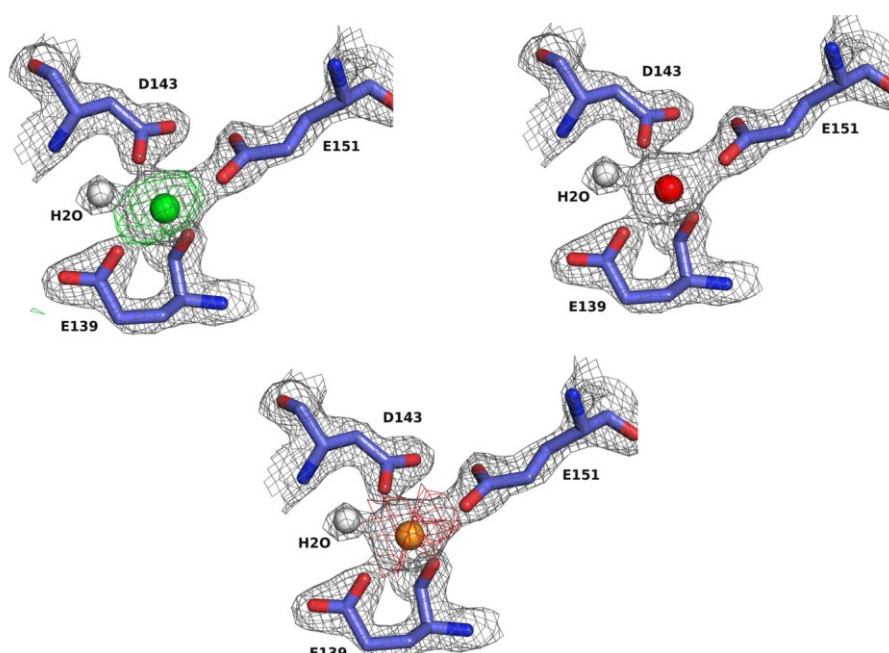
Experimental electron density around calcium binding site 1

The electron density is shown as a gray mesh ($2F_o$ - F_c map contoured at 1.5σ). To test if the observed electron density corresponds to calcium (C, red sphere), we also performed refinements with REFMAC5 with magnesium (A, green sphere) and manganese (C, orange sphere). As can be seen from the difference map (F_o - F_c map contoured at 1 σ), a magnesium-ion produces "positive" peaks (green mesh), indicating that the difference between the observed and calculated structure factors causes structure features which are not present in the model. In contrast, a manganese-ion produces "negative" peaks (red mesh), which are caused by features in the model that cannot be found in the experimental density. Only a calcium-ion does not show any difference in the F_o - F_c map, concluding that a magnesium-ion is too small (12 electrons) and a manganese-ion is too big (25 electrons) to be modeled into the density. Only the modeled calcium-ion (20 electrons) can interpret the density correctly.

Figure S3:
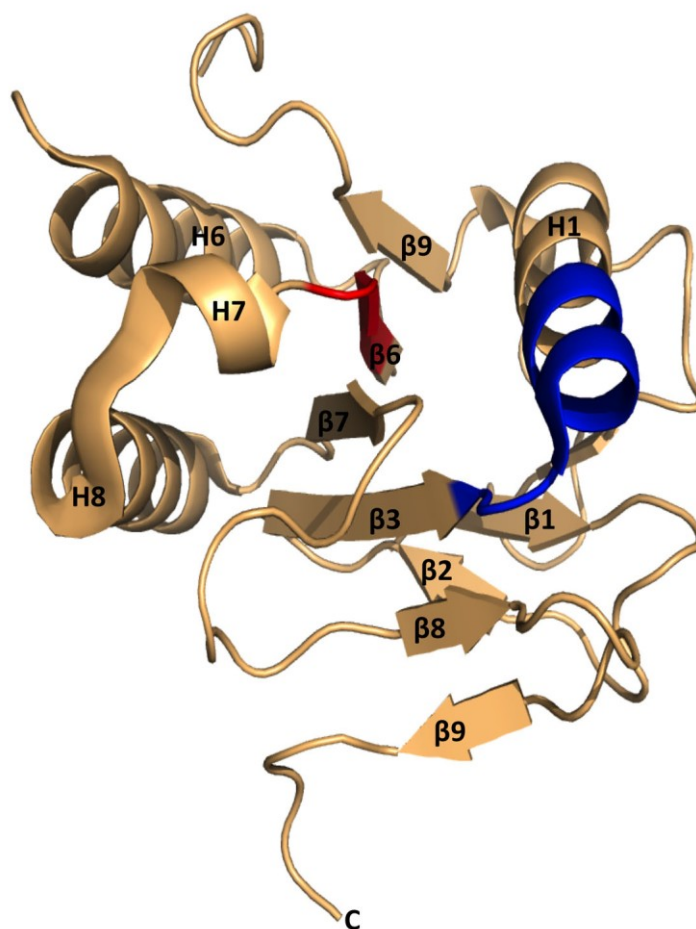
Head domain of Csn2

Shown is the head domain of a Csn2 monomer adopting a helicase fold. The structure is shown in cartoon and the degenerated Walker 'A' (blue) and Walker 'B' (red) are highlighted.

Figure S4:

```
                                            Walker A
S. agalactiae      MVKINFPILDEPLVLSNA--TILTIEDVSVYSSLVKHFYQYDVDEH--LKLFDDKQKSLK 56
S. pyogenes        -MNLNFSLLDEPIPLRGG--TILVLEDVCVFSKIVQYCYQYEEDSE--LKFFDHKMKTIK 55
S. mutans          -MKLNFPILDEPITLEKS--TILVLEDVQVFAQMVRNLYQYDEDSE--LKFFNRKFKSLK 55
E. faecium         MLNLNFPLLEQPLSLEKA--TFLIVEDVTVFAKIVKWFYQYSESHE--LKLFNDKYKELK 56
L. monocytogenes   -MNFNFKLLDDPIQIEDS--TIFVIEDVRVFANVTRFFYQYEEIEE--LTIFDAKHQPLK 55
E. faecalis        -MRVNFSLLEEPIEIEKA--TFLTIKDVQTFAHLVKLIYQYDGENE--LKLFDAQQKGLK 55
S. thermophilus    -MKINFSLLDEPMEVNLG--TVLVIEDVSVFAQLVKEFYQYDEQSN--LTIFDSKIRSIR 55
R. lactaris        -MKLVHINLNEGILVDAINCTEWVIESPEYFSEYVMELAGQVEGKEGRFVLSDNEKEVDI 59
                                            GxxGxGKS/T

S. agalactiae      ATELMLVTDILGYDVNSAPILKLIHGDLENQFNEKPEVKSMVEKLAATITELIAFECLEN 116
S. pyogenes        ESEIMLVTDILGFDVNSSTILKLIHADLESQFNEKPEVKSMIDKLVATITELIVFECLEN 115
S. mutans          PSELMLVTDILGYDVNAPSLLKLVHADLENQFNEKPEVKSMVEKLANTITELIAYECLEN 115
E. faecium         ESDLLIITDVLGFDINSASTLKLIYADLERQLNEKVEVKSMIDKLTATISELIGYELLDH 116
L. monocytogenes   SSELMLITDVLGHDINSAATLKLIYADLEQQLNEKPEVKSMIDKLTATISELIGYELLEH 115
E. faecalis        PTELFVVTDILGYDVNSAATLKLIYGDLEAQLNDKPEVKSMIEKLIGTISQLIGYELLEH 115
S. thermophilus    SSELLLITDILGYDINTSQVLKLLHTDIVSQLNDKPEVRSEIDSLVSLITDIIMAECIEN 115
R. lactaris        SKNVELIFNIFALDINERKLISKLYMELEKLTADERFYVKTQEMKQYLQEYLLQLEQETD 119

                                                            Walker B
S. agalactiae      ELDLEYDEITILELIKALGVKIETQSDTIFEKCFEIIQVYHYLTKKNLLVFVNSGAYLTK 176
S. pyogenes        ELDLEYDEITILELIKSLGVKVETQSDTIFEKCLEILQIFKYLTKKKLLIFVNSGAFLTK 175
S. mutans          ELDLEYDEITILELIKALGVKIETQSDTIFEKMFEVLQVYKYLNKKKLLVFINTLSYFKR 175
E. faecium         ELDLEGDEITVLELFKALGIKIETRSDTIFEKLLELLQVYKYLSKKKLLVLINVCSYLTE 176
L. monocytogenes   ELDLEEDEITVIELFKALGIKIETKSDTVFEKLIEIVQVYKYLSKKKLLVLINVCSYLTE 175
E. faecalis        EMDLEEDGITVQELFKALGIKIETTSDTIFEKVMEITQVHRYLSKKKLLIFINACTYLTE 175
S. thermophilus    ELDIEYDEITLLELIKALGVRIETKSCTVFEKIFEILQIFKYLVKKRILVFVNSLSYFSK 175
R. lactaris        YILDLADEIDFASLFKAFGIKYEVLEENFLERLVRYMRIVERLLKKKLFVFVNLRSYLSN 179
                                                            hhhhD

S. agalactiae      DEVIKLCEYINLMQKSVLFLEPRR---LYDLPQYVIDKDYFLIGENMV 221
S. pyogenes        DEVASLQEYISLTNLTVLFLEPRE---LYDFPQYILDEDYFLITKNMV 220
S. mutans          EEIAQILEYIHLSDMVVLFIEPRK---IDDFAQYILDEDYFLITESNN 220
E. faecium         EEISEVQDYISLYHMEVLFVEPRS---IKGIKQFVLDKDYFLNLENMV 221
L. monocytogenes   EELLELRRYISLYQVKVLFIEPRK---IKGSPQVTLDSDYFLHVENSV 220
E. faecalis        DEVQQVVEYISLNNVDVLFLEQRV---VQNRFQYILDENFYLSYEKA- 219
S. thermophilus    DEIYQILEYTKLSQADVLFLEPRQ---IEGIQQFILDKDYILMPYNN- 219
R. lactaris        QQIDELIKEATYQEIQLLLIETCARDCIEGVTRYIIDKDGCEI----- 222
                                                      DxD
```

Sequence alignment of different Csn2 proteins

The following proteins are aligned: Csn2 from *S. agalactiae* ATCC13813 (sequenced from genomic DNA), *S. pyogenes* (NCBI Accession code: NP_269218.1), *S. mutans* (YP_003484615), *E. faecium* (ZP_03981024), *L. monocytogenes* (ZP_05235633), *E. faecalis* (EFT90893), *S. thermophilus* (ADQ63487) and *R. lactaris* (ZP_03166524). All colored amino acids are highly conserved among the Csn2 protein family. Colored in red are amino acids, which were mutated to alanine. Blue amino acids represent the putative acidic triad are colored in green or orange are the degenerated Walker 'A' and 'B' motifs with their consensus sequence are given below.

Figure S5:
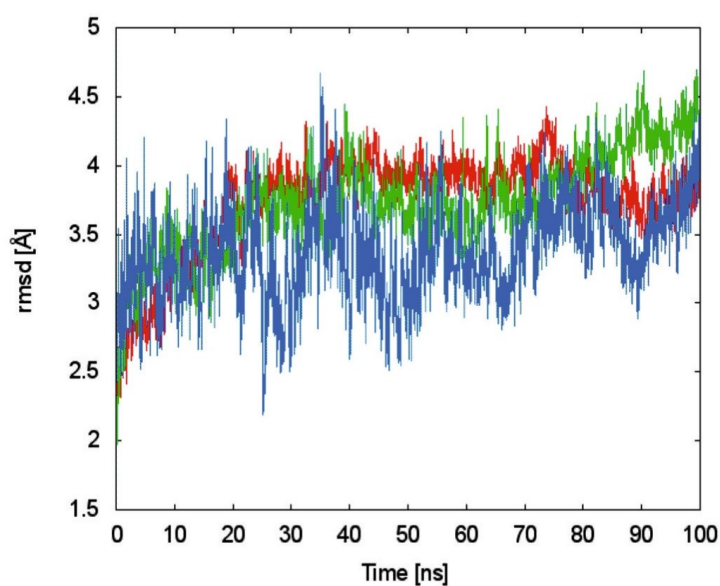
Rmsd of the tetrameric Csn2 structure during the MD simulation

Rmsd of Cα atoms of either head region (red, green) with respect to the crystal structure; each head region was aligned to itself. In addition, the rmsd of Cα atoms of the legs with respect to the crystal structure is shown (blue); the legs were aligned to themselves.

Figure S6:
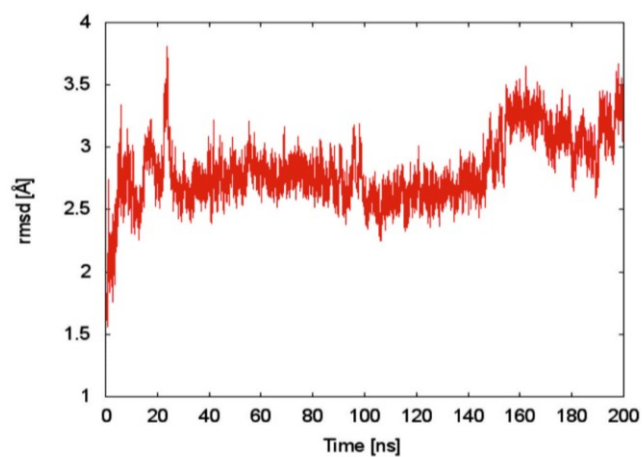
Rmsd of the head region of Csn2 during the MD simulation

Rmsd of Cα atoms of the head region with respect to the crystal structure.

Figure S7:
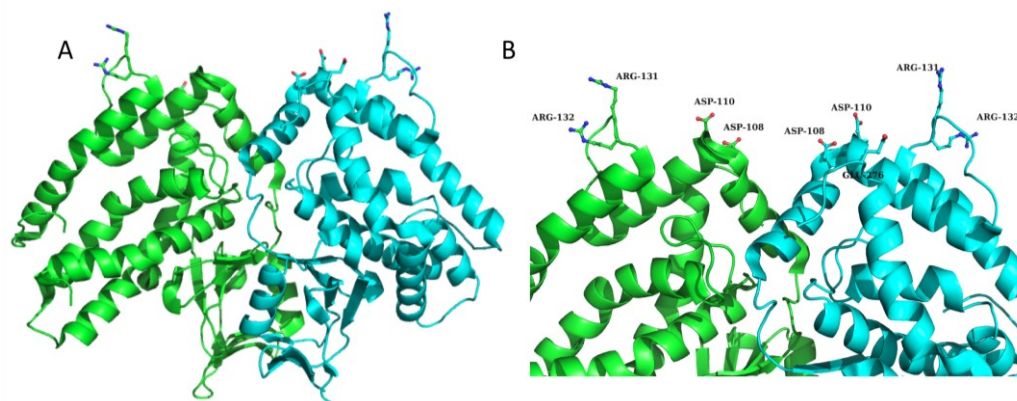
Structure of Cas 1 from *E.coli*

A Structure of Cas 1 from *E.coli* (PDB code: 3NKE). Two monomers are shown as cartoon with the DxD and RR-motifs highlighted as sticks. **B** Zoom into the top of the helical domain visualizing the presence of the DxD and RR-motifs ($_{108}$DxD$_{110}$ and $_{131}$RR$_{132}$).

Figure S8:
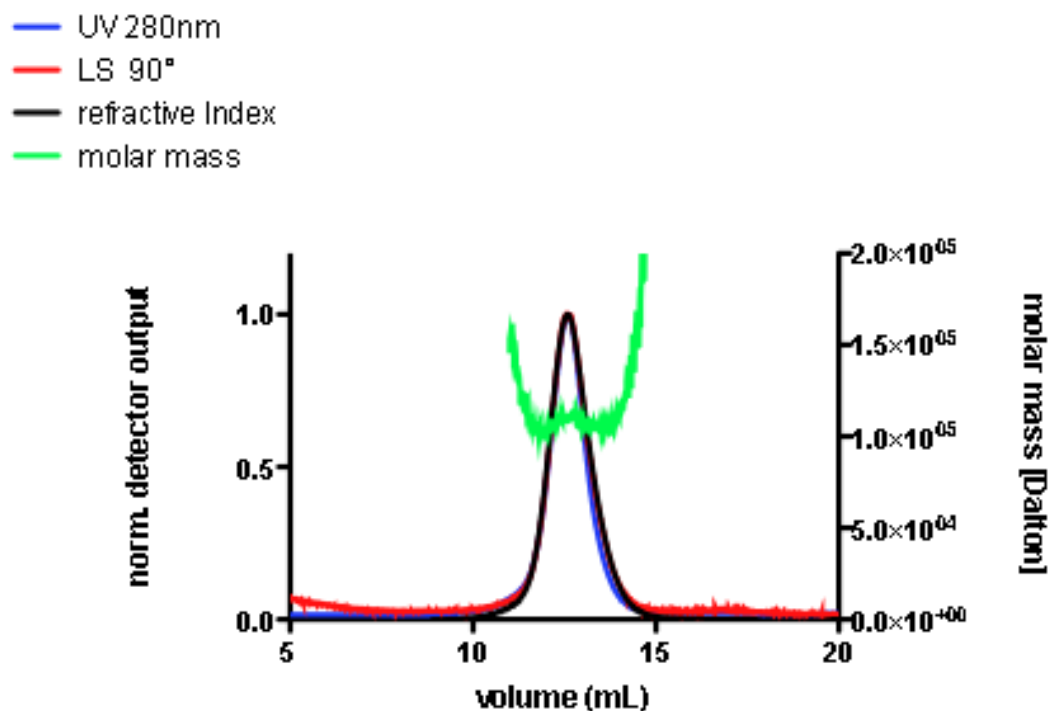
Oligomeric status of Csn2 after addition of 20 mM EGTA

The elution profile from a Superdex 200 10/300 column of Csn2 supplemented with EGTA (final concentration 20 mM) blue line: normalized UV-Signal; red line: normalized LS-(90°)-Signal; black line: refractive index signal. The green line indicates the corresponding molar mass. Csn2 has a mass of 104.1 ± 0.3 kDa in the peak fraction indicating a homogenous, tetrameric species in solution.

Supplementary Tables

Table S1. Coordination partners and distances for the observed calcium-ions and their coordination geometry

Ca^{2+} site 1		Ca^{2+} site 2		Ca^{2+} site 3	
Interaction	Distance in Å	Interaction	Distance in Å	Interaction	Distance in Å
Oϵ E139	2.46	Oδ D123	2.40	Main-chain O C133	2.45
Main-chain O E139	2.48	Main-chain O E124	2.25	Main-chain O N116	2.31
Oδ D143	2.33	Oϵ E129	2.51 (average)	Main-chain O L118	2.26
Oϵ E151	2.38	Main-chain O A133*	2.33	H$_2$O 1	2.40
Oδ D119*	2.48	H$_2$O 1	2.29	H$_2$O 2	2.29
H$_2$O	2.49	H$_2$O 2	2.37		
Coordination type	octahedral	octahedral		tetragonal-pyramidal	

Table S2. Result of the Dali server [1] search. Shown are the ten first hits with PDB code, Z-score, r.m.s.d. value and sequence identity.

Protein	*PDB code*	*Z-score*	*rmsd*	*Sequence Identity (%)*
SMC Head Domain from *Thermotoga maritima*	1E69	9.1	3.1	7
ABC Trasnporter ModBC	2ONK	8.9	3.5	9
Structure of PH0203 Protein from *Pyrococcus*	2IT1	8.8	3.3	12
Cobalt Import ATP-Binding Protein	3QFO	8.8	3.3	13
replicative Helicase-Primase of Bacteriophage T7	1CR1	8.8	3.3	6
GlcV, ABC-ATPase	1OXU	8.5	3.7	7
MJ1267 ATP binding cassette	1G6H	8.4	3.5	10
DNA Double-Strand Break Repair RAD50 ATPase	3QF7	8.4	3.3	6
Maltose/Maltodextrin Import ATP-Binding Protein	2AWN	8.4	3.4	6
Bifunctional Primase-Helicase of Bacteriophage T7	1Q57	8.4	3.2	7

Supplementary References

[1] Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic acids research 38: W545-549.

Chapter 2

Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2

Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2

Zihni Arslan[1], Reinhild Wurm[1], Oleksandr Brener[1,2], Philipp Ellinger[3], Luitgard Nagel-Steger[1,2], Filipp Oesterhelt[1], Lutz Schmitt[3], Dieter Willbold[1,2], Rolf Wagner[1], Holger Gohlke[4], Sander H. J. Smits[3] and Ümit Pul[1,*]

[1]Institut für Physikalische Biologie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany, [2]ICS-6, Forschungszentrum Jülich, 52425 Jülich, Germany, [3]Institut für Biochemie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany and [4]Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany

ABSTRACT

The adaptive immunity of bacteria against foreign nucleic acids, mediated by CRISPR (clustered regularly interspaced short palindromic repeats), relies on the specific incorporation of short pieces of the invading foreign DNA into a special genomic locus, termed CRISPR array. The stored sequences (spacers) are subsequently used in the form of small RNAs (crRNAs) to interfere with the target nucleic acid. We explored the DNA-binding mechanism of the immunization protein Csn2 from the human pathogen *Streptococcus agalactiae* using different biochemical techniques, atomic force microscopic imaging and molecular dynamics simulations. The results demonstrate that the ring-shaped Csn2 tetramer binds DNA ends through its central hole and slides inward, likely by a screw motion along the helical path of the enclosed DNA. The presented data indicate an accessory function of Csn2 during integration of exogenous DNA by end-joining.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins constitute a prokaryotic adaptive immunity system to prevent viral infections or the invasion by mobile DNA elements (1–3). The immunization is achieved by the specific storage of foreign nucleic acid sequences as 'portraits' of the targets, which are later used in the form of small RNAs for the specific recognition and inactivation of the invading DNA. Ten CRISPR-Cas systems have been described, which differ in the organization of the CRISPR operon, constituted by a set of diverse CRISPR-associated genes (*cas*) and the CRISPR arrays (4). The *cas* genes encode the protein components of the system, which exhibit various biochemical activities (5).

The mechanism of the CRISPR-mediated immunity is divided into three stages (2,6). The first stage describes the immunization of the host cells by the integration of foreign DNA-derived spacer sequences into the CRISPR array (adaptation stage). The second stage comprises the transcription of the CRISPR array to the precursor CRISPR RNA (pre-crRNA), the expression of Cas proteins and the processing of the pre-crRNA to small crRNAs (expression/processing stage). In the third stage, the crRNA-loaded Cas protein complexes screen the foreign DNA for spacer-matching sequences and initiate the nucleolytic hydrolysis of the target DNA (interference stage). According to the recent classification of the different CRISPR-Cas systems, three major types (type I, II and III) have been defined (4), which exhibit mechanistic variability in the maturation of the crRNAs and the inactivation of the target nucleic acids (7–15).

The type II CRISPR-Cas systems, mainly found in pathogenic bacteria (16), consist of the CRISPR array, the *cas* genes encoding for the proteins Cas9, Cas1, Cas2 and Csn2 (in type II-A) or Cas4 (in type II-B) and a gene for the trans-encoded crRNA (tracrRNA). The tracrRNA contains sequence elements complementary to the repeat of the pre-crRNA (11). The RNA duplex formation of the tracrRNA and the pre-crRNA mediates the maturation of the crRNAs by RNaseIII and the Cas9 protein (11,17). Loaded with the tracrRNA:crRNA duplex, the Cas9 protein introduces double-strand DNA breaks at the spacer-matching protospacer region (12). Several studies have shown that the Cas9 protein along with designed crRNA:tracrRNA is active in eukaryotic cells and can be used for multiplex genome editing in bacteria and eukaryotes (18–23). Moreover, engineered Cas9 protein with inactivated nuclease domains has been used for sequence-specific control of transcription in *Escherichia coli* and eukaryotic cells (24).

*To whom correspondence should be addressed. Tel: +49 211 81 14945; Fax: +49 211 81 15167; Email: pul@hhu.de

The activities of the proteins Cas1, Cas2 and Csn2 are dispensable for maturation of the crRNAs and the cleavage of the target DNA, and thus are likely required in the immunization stage (10,11). The nucleases Cas1 and Cas2 are involved in incorporation of new spacer sequences into the CRISPR array (25–27). Little is known about the function of the Csn2 protein; although early studies showed that a knockout of the *csn2* gene prevents the development of phage-insensitive mutants by an uptake of new spacer sequences (7,28), the role of the Csn2 protein during immunization is unknown. The structures and DNA-binding activities of Csn2 proteins from *Enterococcus faecalis*, *S. agalactiae* and *Streptococcus pyogenes* have been analyzed independently (29–31). These Csn2 proteins are ring-shaped homotetramers, and the structure of a protomer consists of a 'head' and a 'leg' domain. Whereas the leg domain consists of a pair of long stretched helices that form the interface for the tetrameric assembly, the function of the head domain is still unclear. More recently, based on the crystal structure of the Cas protein Stu0660 from *Streptococcus thermophilus*, a second Csn2 protein family has been proposed, which shares no sequence similarity to Csn2 proteins; however, its members have a similar tetrameric ring-shaped structure (32). All Csn2 proteins exhibit Ca^{2+}-dependent double-stranded DNA (dsDNA)-binding activities without any notable sequence dependence. The tetrameric structure contains a large positively charged central hole of ~4 nm in diameter, which is stabilized by Ca^{2+}-ions bound to the helices in the leg domain. This hole is large enough to encircle the dsDNA. An alternative DNA binding mode has been suggested by molecular dynamics simulations, which revealed that the Csn2 tetramers can adopt a conformation with a groove outside the ring that could interact with the dsDNA (30).

To elucidate the role of the Csn2 protein in the CRISPR immunization stage, we have analyzed the DNA-binding mechanism of Csn2 from the human pathogen *S. agalactiae* in detail, using different biochemical techniques, atomic force microscopy (AFM) and molecular dynamics (MD) simulations. We provide evidence that Csn2 binds to the ends of linear dsDNA and moves inward through rotation-coupled translocation. The encircling of the dsDNA by Csn2 tetramers was probed by capping the DNA ends of Csn2–DNA complexes with streptavidin, keeping the Csn2 rings arrested on the DNA. The MD simulations give mechanistic insights at the molecular level how a Csn2 ring may rotate along the helical path of the DNA in a Ca^{2+}-dependent manner. The data suggest that Csn2 is a DNA end-loading toroidal protein, whose function is directly related to DNA end metabolism. Csn2 thus resembles proteins involved in the repair of dsDNA breaks and recombination.

MATERIALS AND METHODS

Purification of Csn2

Streptococcus agalactiae Csn2 protein was expressed and purified as previously described (30).

Preparation of relaxed plasmid DNA for competition experiments

Relaxation of supercoiled plasmid DNA was performed with *E. coli* Topoisomerase I (New England Biolabs, NEB). Supercoiled pUC18-1 plasmid DNA (33) was extracted and purified from *E. coli* cells grown to stationary phase in YT medium. Ten micrograms of supercoiled plasmid DNA were incubated in 1 × NEBuffer 4 (50 mM potassium acetate, 20 mM Tris acetate, pH 7.9, 10 mM magnesium acetate, 1 mM DTT) and 1 µg ml^{-1} bovine serum albumin in a total volume of 30 µl. Five units *E. coli* Topoisomerase I were added and the reaction mixture was incubated for 1 h at 37°C to allow complete relaxation. After extraction with phenol/chloroform and ethanol precipitation, the DNA was resuspended in deionized water (Millipore). The relaxation was verified on a 1% agarose gel. DNA bands were stained with ethidium bromide and visualized under UV light.

Analyses of protein–DNA interaction by electrophoretic mobility shift assays

The 155 bp DNA fragment was obtained by *Eco*RI-*Dra*I digestion of the plasmid pUC18-1. DNA fragments were purified by agarose gel electrophoresis and end labeled by Klenow (Promega) incorporation of [α-^{32}P]-dATP. Binding reaction mixtures contained 2 nM DNA, 10 mM $CaCl_2$, 20 ng µl^{-1} heparin and indicated concentrations of Csn2 or unlabeled plasmid DNA (supercoiled, relaxed or linearized) as given in the individual experiments. After incubation for 15 min at room temperature, the complexes were separated from free DNA on native 5% (w/v) Tris/glycine polyacrylamide gels. DNA bands were visualized by autoradiography.

Preparation of DNA for AFM

The pCR001 plasmid (5125 bp) used in the AFM analyses was constructed by ligation of *cas1*-*cas2* genes into the *Nde*I-*Pac*I cleaved pACYCDuet-1 vector (Novagen). The insert with *cas1*-*cas2* genes were obtained by PCR from genomic DNA of *E. coli* MG1655 using the following primers: forward 5′-GAATGC<u>CATATG</u>ACCTGGC TTCCCCTTAAT-3′ and reverse 5′-CCGAC<u>CTTAATTA</u> <u>A</u>CATTTCCTTATTATTAAAGATCAGCT-3′.

The plasmid pCR001 was extracted and purified from *E. coli* cells grown to stationary phase in YT medium with the Qiagen Plasmid Midi Kit. To obtain relaxed DNA, 10 µg of the supercoiled plasmid was treated with *E. coli* Topoisomerase I as described above. Linear plasmid DNA used in the AFM analyses was produced by cleavage with the restriction enzyme *Ecl136*II. After relaxation or linearization reactions, the DNA was extracted with phenol/chloroform, purified with PCR purification Kit (Qiagen) and precipitated with ethanol. The relaxed and linearized DNA probes were resuspended in deionized water (Millipore).

Atomic force microscopic (AFM) analyses of Csn2–DNA complexes

Linearized or relaxed pCR001 was incubated with the Csn2 protein in the presence of 2 mM $CaCl_2$ in a total

volume of 5 μl. The DNA and protein concentrations are given in the individual experiments. The samples were incubated 10 min at room temperature to allow complex formation and diluted with 95 μl of adsorption buffer (5 mM $CaCl_2$, 2 mM Tris–HCl, pH 8.0, 10 mM NaCl). Forty-microliter aliquots of the samples were immediately transferred to freshly cleaved mica. After 1 min incubation the mica surface was washed three times with 100 μl deionized water (Millipore). Excess water was dried with compressed N_2. Imaging of Csn2–DNA complexes was performed with Nanowizard II (JPK Instruments AG, Berlin) in intermittent contact in air, using standard silicon cantilevers (OLYMPUS OMCL-AC160).

Biotinylation of dsDNA fragment and Csn2 binding studies

The 256 bp DNA fragment was obtained by *Eco*RI-*Bam*HI digestion of the plasmid pUC18-1. After purification of the DNA fragments by agarose gel electrophoresis, the biotinylation of the ends was achieved by Klenow incorporation of Biotin-11-dUTP. Ten microgram of the DNA fragment was incubated with 5 U Klenow (Promega), each 0.2 mM dCTP, dGTP, dATP and Biotin-11-dUTP (GeneOn) in 1 × Klenow buffer (Promega, 50 mM Tris–HCl, pH 7.2, 10 mM $MgSO_4$ and 100 μM DTT) in a total volume of 50 μl for 1 h at room temperature. After extraction with phenol/chloroform and ethanol precipitation, the DNA was dissolved in 10 μl TE buffer.

The binding studies with the biotinylated DNA were performed by sequential incubation of 168 ng of the biotinylated DNA in the presence of 7.2 mM $CaCl_2$ with 4.7 μg Csn2, 14 mM ethylene glycol tetraacetic acid (EGTA) and/or 2 μg streptavidin (Sigma-Aldrich) in a total volume of 14.4 μl. The volumes of the binding reaction without EGTA or streptavidin were adjusted by addition of deionized water (Millipore). The incubation of the samples occurred in three steps, each for 15 min at room temperature. The complexes were separated from free DNA on a native 2% Tris/Acetate agarose gel. The bands were visualized by ethidium bromide staining.

MD simulations

MD simulations were performed with the AMBER 11 suite of programs (34) together with the force field as described by Cornell *et al.* (35) using modifications suggested by Simmerling *et al.* (36). To generate a starting structure of Csn2 binding to dsDNA, canonical B-DNA of 36 bp length (sequence: 5′-GTTTTAGAGCTGTGCTG TTTCGAATGGTTCCAAAAC-3′) was inserted into the central hole of tetrameric Csn2 such that the DNA is perpendicular to a least-squares fit plane through the protein atoms and that Csn2 is displaced by ~6 Å from the center of the DNA along the double strand. The tetrameric Csn2 was taken from Ellinger *et al.* (30) (PDB code: 3QHQ), including the Ca^{2+}-ions. In addition, a starting structure of Csn2 binding to dsDNA without Ca^{2+}-ions was generated. The starting structures were placed into octahedral periodic boxes of TIP3P water molecules (37). The distance between the edges of the water box and the

closest atom of the solutes was at least 11 Å, resulting in system sizes of ~186 000 atoms. The systems were minimized by 50 steps of steepest descent minimization followed by 450 steps of conjugate gradient minimization. The particle mesh Ewald method (38) was used to treat long-range electrostatic interactions, and bond lengths involving bonds to hydrogen atoms were constrained using SHAKE (39). The time-step for all MD simulations was 2 fs, with a direct-space, non-bonded cutoff of 8 Å. Applying harmonic restraints with force constants of 5 kcal $mol^{-1}Å^{-2}$ to all solute atoms, canonical ensemble (NVT)-MD was carried out for 50 ps, during which the system was heated from 100 to 300 K. Subsequent isothermal isobaric ensemble (NPT)-MD was used for 150 ps to adjust the solvent density. Finally, the force constants of the harmonic restraints on solute atom positions were gradually reduced to zero during 100 ps of NVT-MD. The following NVT-MD at 300 K with a time constant of 10 ps for heat-bath coupling was used for analysis, with conformations extracted every 20 ps. The simulation lengths amount to 100 ns of which the first 30 ns were discarded. Each simulation was repeated once with varied starting velocities, resulting in four independent simulations in total.

For analyzing the trajectories, conformations were superimposed with respect to the phosphorous atoms of the dsDNA or the C_α atoms of Csn2. The conformational variability of Csn2 binding to dsDNA agrees with that observed in MD simulations of Csn2 alone (30). For determining the translocation along the dsDNA, the distance between the center of the dsDNA and the average coordinates of the C_α atoms of all four helices H5 was computed. For determining the rotation of Csn2 relative to the dsDNA, the angle defined by the point triple (average coordinates of the phosphorous atoms of the five most central nucleotides on one strand; average coordinates of the phosphorous atoms of the five most central nucleotides on the other strand; average coordinates of the C_α atoms of one pair of helices H5 located in one of the legs of the tetrameric structure) was computed. For determining the kink angle of the dsDNA, the angle defined by the point triple (average coordinates of the phosphorous atoms of the five outermost base pairs at one end of the DNA; average coordinates of the phosphorous atoms of the five most central base pairs of the DNA; average coordinates of the phosphorous atoms of the five outermost base pairs at the other end of the DNA) was computed.

RESULTS

Binding of Csn2 requires free DNA ends

The Csn2 tetramers from *S. agalactiae* are stable in the range of pH 7.0–9.0, a wide range of ionic strengths as well as in the presence of Ca^{2+}-chelating EGTA (30). Owing to the high stability of the ring-shaped Csn2 tetramers, it is unlikely that the quaternary structure can be opened wide enough to bind and enclose dsDNA as it is known to occur in several DNA-enclosing proteins (40,41). Therefore, if the Csn2 tetramers encircle the dsDNA the

tetramers should bind DNA by threading onto the DNA ends. In contrast, a binding of the DNA outside the central hole will be independent on the presence of free DNA ends. To distinguish between these two alternative DNA-binding modes, we performed electrophoretic mobility shift assays (EMSAs) and tested whether the binding of Csn2 requires free DNA ends.

First, we incubated a 155 bp ^{32}P-radiolabeled DNA fragment with increasing amounts of Csn2 and separated the protein–DNA complexes on a 5% native Tris/glycine polyacrylamide gel. As can be seen in Figure 1A, Csn2 formed one complex band with an estimated apparent $K_{0.5}$ value between 30 and 60 nM (Figure 1A, lanes 3 and 4). The declining mobility of the complex band with increasing protein concentration indicates the binding of multiple Csn2 molecules to the DNA fragment. To test a putative requirement for a free DNA end for the interaction, we performed competition experiments with unlabeled plasmid DNA, either in closed circular or linearized form. As shown in Figure 1B (lanes 3–6), the complex formation of Csn2 with the linear radiolabeled DNA was challenged with increasing concentrations of the linearized pUC18-1 plasmid. However, the same plasmid in supercoiled or relaxed form almost did not compete for binding of Csn2 (Figure 1B, lanes 7–14; Figure 1C, lanes 19–22). This indicates that free DNA ends are necessary for Csn2-binding and supports the threading of dsDNA into the central channel of the Csn2 tetramers.

To further evaluate the dependence of the Csn2–DNA interaction on free dsDNA ends, we repeated the competition experiments with the plasmid DNA cleaved by restriction enzymes at two sites, with the aim to double the concentration of the DNA ends. The competition efficiency was higher with DNA cleaved with double cutters *Pvu*II or *Ava*II compared with the *Sca*I-linearized DNA at the same final competitor concentration (Figure 1C, lanes 7–14). Moreover, plasmid DNA cleaved with *Hae*III at 12 sites reduced the binding to the radiolabeled DNA already at the lowest competitor concentration tested (Figure 1C, lanes 15–18). Thus, the challenging efficiency depended on the concentration of free DNA ends at equal total competitor DNA concentration. The appearance of three intermediate complexes during the decay of the Csn2–DNA complexes is consistent with a binding of multiple Csn2 tetramers to the 155 bp dsDNA, and suggests a sliding of the bound proteins along the DNA after loading the DNA from its ends.

Cleavage of the pUC18-1 plasmid with the restriction enzymes *Sca*I, *Pvu*II and *Hae*III produces blunt ends, while cleavage with *Ava*II gives 5′-overhangs of three bases. The comparable challenging extent of the *Pvu*II- and *Ava*II-cleaved DNA demonstrated that the DNA end-binding activity of Csn2 is not restricted to blunt ends but short 5′-protruding DNA ends are also bound by Csn2 (Figure 1C, lanes 7–14). The ability of Csn2 to bind dsDNA with 3′- and/or 5′-overhangs was further examined with different synthetic DNA substrates
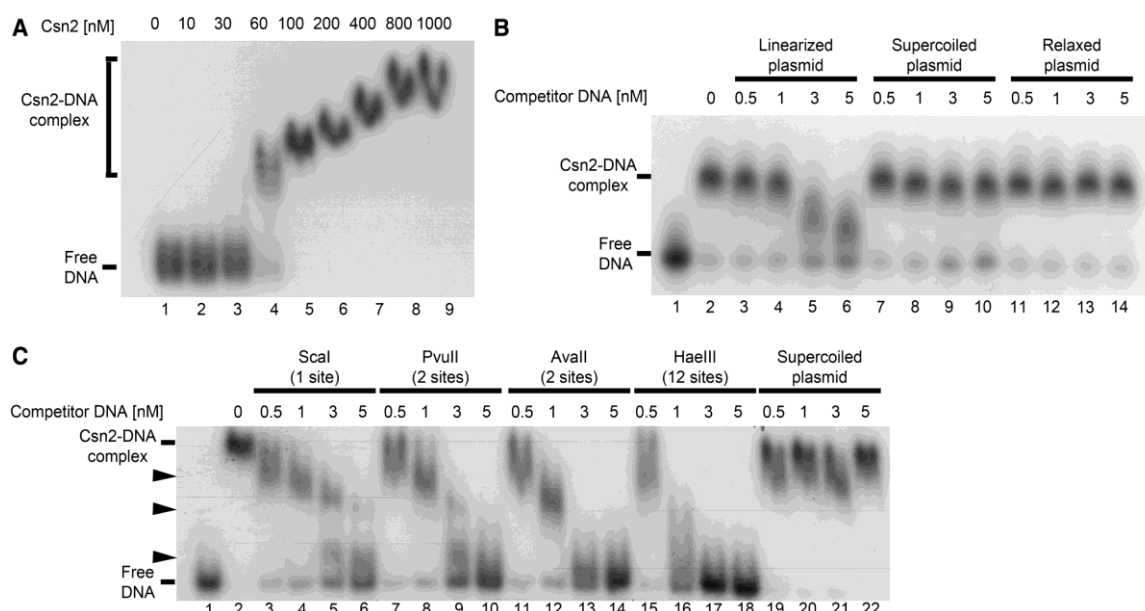
Figure 1. Electrophoretic mobility shift assays of a radiolabeled 155 bp DNA fragment with Csn2 either in the absence (**A**) or in the presence of competitor DNA (**B** and **C**) are presented. In each reaction 2 nM ^{32}P-labeled DNA, 20 ng/μl heparin, and 10 mM CaCl$_2$ were employed. (A) Titration of Csn2 in the range of 0 to 1 μM is shown. (B) Csn2 binding to the radiolabeled DNA fragment was competed with indicated amounts of 2915 bp unlabeled plasmid DNA either in *Sca*I-linearized (lanes 3–6), supercoiled (lanes 7–10) or in relaxed (lanes 11–14) form. The concentration of Csn2 was constant at 60 nM in lanes 2–14. Lanes 1 and 2 show the control reactions, performed either in the absence of Csn2 (lane 1) or in the absence of competitor DNA (lane 2). (C) The same competition experiment as in (**B**) but with *Pvu*II-, *Ava*II-, or *Hae*III-cleaved competitor plasmid. The numbers of cleavage sites of the different endonucleases are given in the brackets. The black arrowheads indicate intermediate Csn2-DNA complexes, resulting from decomposition of the fully occupied complexes.
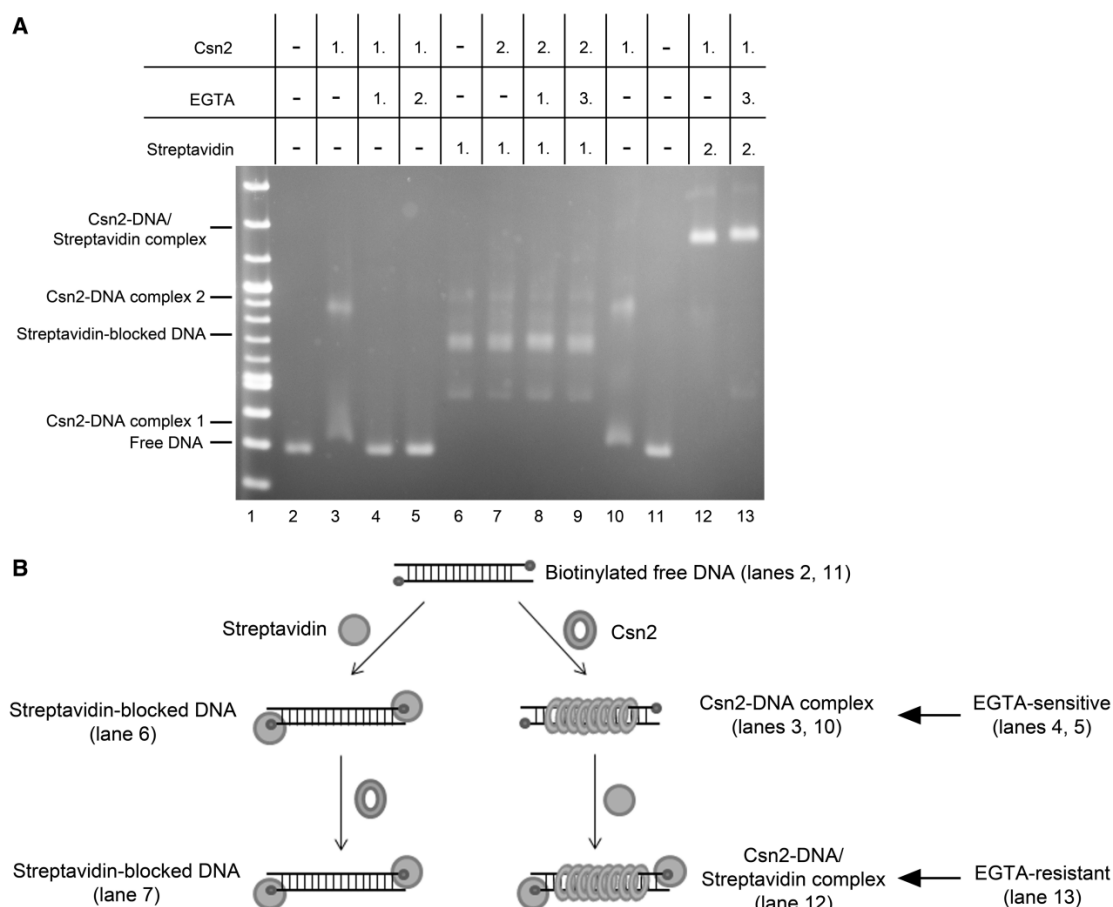
Figure 2. (**A**) Binding analyses of Csn2 in the presence and absence of EGTA and free DNA ends on 2% Tris-acetate agarose gel. In each lane 168 ng linear DNA and 7.2 mM $CaCl_2$ were employed. The numbers above the lanes indicate the order of addition of streptavidin (2 μg), Csn2 (4.7 μg), or EGTA (14 mM) in a total volume of 14.4 μl. Lanes 2–5: Influence of EGTA on Csn2-DNA interaction is shown. Lanes 6–9: 168 ng of the end-biotinylated DNA fragment were incubated first with streptavidin to block the DNA ends. Lanes 10 and 11: Streptavidin was added after binding of Csn2. After separation of the complexes the agarose gel was stained with ethidium bromide. (**B**) Schematic presentation of the binding analysis, shown in (A).

(Supplementary Figure S1). Csn2 did not form complexes with single-stranded DNA (ssDNA) (Supplementary Figure S1A and B, lanes 1–3). However, the hybridization of complementary oligonucleotides to ssDNA restored the Csn2–DNA interaction (Supplementary Figure S1A and B, lanes 4–6). Moreover, Csn2 formed faint complex bands with a DNA substrate that contains a 36-bp dsDNA region flanked by 29 or 34 nt ssDNA tails (Supplementary Figure S1B, lanes 7–9). These results indicate that Csn2 is in principle able to enter dsDNA regions by threading ssDNA tails in its central hole.

Dissociation of Csn2 occurs at the DNA ends: evidence for encircling the DNA by Csn2 rings

The crystal structure of the Csn2 from *S. agalactiae* revealed three Ca^{2+}-ions bound to the leg domain of each protomer (30). These ions have also been found in the crystal structures of Csn2 proteins from *E. faecalis* and *S. pyogenes* (29,31). The loss of one of the Ca^{2+}-ions within one protomer of Csn2 resulted in a reorientation of α-helices in the leg domain, which form the inner rim of the central hole (30) (Supplementary Figure S2). Thus, the Ca^{2+}-ions are considered to be important for the stabilization of the tetramer conformation that is able to bind the dsDNA (29–31).

Because capturing the Ca^{2+}-ions by EGTA did not disassemble the tetramers but abolished their DNA-binding activity (30,31), we suggested that in the presence of EGTA the conformation of the central hole becomes disordered, resulting in weakened interactions of the tetramers with the DNA backbone and leading to the dissociation of the sliding rings from the DNA ends. If this were the case, the obstruction of the DNA ends of

Csn2–DNA complexes should result in arrested Csn2 rings on the DNA, and the complexes should not dissociate in the presence of EGTA. To test this, we attached biotin to both ends of a dsDNA fragment, allowing the blocking of both DNA-termini with streptavidin.

Incubation of Csn2 with biotinylated dsDNA led to the formation of two complex bands, demonstrating that biotin itself did not interfere with Csn2 binding (Figure 2A, lanes 2, 10). As shown previously (29–31), in the presence of EGTA the complex formation was completely inhibited (Figure 2, lane 4). To test whether EGTA causes the dissociation of DNA-bound Csn2 molecules, we incubated the DNA with Csn2 in the presence of Ca^{2+} for 15 min and then added EGTA to the binding reaction followed by incubation for additional 15 min. Indeed, the lack of Csn2–DNA complexes indicated the decomposition of the complexes in the presence of EGTA (Figure 2A, lane 5).

As expected, the tethering of streptavidin to the DNA ends inhibited the binding of Csn2 (Figure 2A, lanes 6 and 7), consistent with the requirement of free DNA ends for Csn2 binding. In contrast, adding streptavidin after the binding of Csn2 resulted in supershifted Csn2-DNA/streptavidin complexes (Figure 2A, lane 12). Moreover, the Csn2 proteins in these complexes were captured on the DNA because adding EGTA did not lead to a decay of the complexes (Figure 2A, lane 13). Thus, both the association as well as the dissociation of Csn2 occurs at the DNA ends, indicating a sliding activity of the Csn2 rings (Figure 2B). In the absence of Ca^{2+}-ions, the tetramers slip down from the DNA ends, likely due to weakened electrostatic interactions between the distorted α-helices of the ring channel and the phosphate backbone of the DNA helix.

AFM of Csn2–DNA complexes

To confirm the DNA end requirement and to probe a binding of multiple Csn2 molecules, which would confirm a sliding activity of the Csn2 rings, we imaged the Csn2–DNA complexes by AFM in intermittent contact mode in air.

The AFM analyses were performed with the 5125 bp plasmid pCR001 as a substrate, either in linearized or relaxed circular form. AFM images of equal amounts of both DNA forms (each 1.3 nM) in the absence of Csn2 are shown Figure 3A and B. Some of the molecules of both DNA forms contained nodes, likely due to a crossover of DNA helices or a distortion/kinking of the DNA during the drying process of the probe on the mica surface (Figure 3B). AFM images of the same DNA mixture in the presence of 176 nM Csn2 showed the formation of protein–DNA complexes (Figure 3C–F). In average the ratio of the heights and widths of bound Csn2 and free DNA (Csn2-DNA/DNA) were 4.2 and 1.3, respectively. In contrast, the ratio of heights and widths of the DNA nodes relative to straight DNA were in average 1.9 and 1.5, respectively. Thus, these DNA nodes, which were also observed in the protein–free sample (Figure 3B), were not considered as protein–DNA complexes.
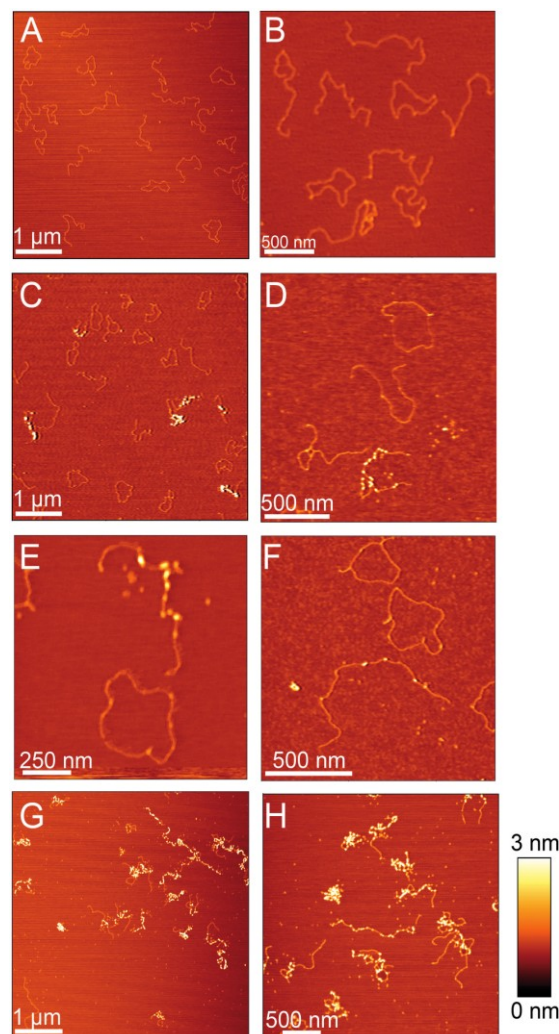
Figure 3. Representative AFM images of 5125 bp plasmid DNA in the absence or presence of Csn2 are shown. (**A**) and (**B**) show images of 1.3 nM relaxed and 1.3 nM linear plasmid DNA in the absence of Csn2; (**C–F**) images of equal amounts of relaxed and linear plasmid DNA (each 1.3 nM) incubated with 176 nM Csn2; (**G**) and (**H**) images of 2.6 nM linear plasmid DNA incubated with 800 nM Csn2. The relative color scale range is 0–3 nm in all images.

Csn2 was exclusively associated with linear DNA (Figure 3C–F), consistent with its DNA end-dependent–binding activity as observed in the mobility shift assays. The binding was not restricted to the DNA ends but internal sites along the entire DNA fragment were also occupied by several Csn2 proteins (Figure 3D–F), indicating a sliding activity of the tetramers after initial binding at the DNA ends. Moreover, the binding of Csn2 seems to be highly cooperative, as some linear DNA molecules were free of any protein, whereas others were bound by several Csn2 proteins in the same sample (Figure 3D–F). Such an apparent cooperative binding

activity is also known for the 'recombination associated protein' RdgC from *E. coli*, a ring-shaped DNA-binding protein with preference for DNA ends (42). As control, incubation of Csn2 with the relaxed plasmid in the absence of linear DNA confirmed that Csn2 does not bind to circular DNA (Supplementary Figure S3).

At higher Csn2 concentration, highly condensed nucleo-protein complexes were observed, likely caused by clustering of multiple Csn2 proteins and DNA-dependent self-association of DNA-bound Csn2 (Figure 3G and H). The formation of these higher-order nucleoprotein complexes should not be caused by an unspecific aggregation of the protein, as previous study has shown that Csn2 exists as stable tetramer in solution (30). To further evaluate the DNA end-dependence and to exclude an unspecific aggregation of Csn2 in the presence of DNA, we analyzed the elution profile of Csn2 in the presence of linear or closed circular plasmid DNA by size exclusion chromatography (Figure 4). Incubation of Csn2 with the linear DNA caused a co-elution of the entire Csn2 tetramers in the void fraction (Figure 4C). In contrast, the same DNA in supercoiled form did not change the elution profile of the tetrameric Csn2 (Figure 4D). The results are consistent with a selective binding of the Csn2 tetramers to the linear dsDNA and exclude an unspecific protein aggregation in the presence of DNA.

MD simulations reveal Ca^{2+}-dependent coupled translocation-rotation motions of Csn2 and kinking of the DNA

To further study Csn2 binding to DNA and the sliding activity of the tetramers, we performed two independent all-atom MD simulations of 100 ns length each of the tetrameric Csn2 with dsDNA of 36 bp length bound within the central hole. The sequence of the dsDNA corresponds to the CRISPR repeat sequence of the type II-A CRISPR system of *S. agalactiae*. Although we do not have an indication for a sequence-specific binding of Csn2, a bioinformatics analysis revealed a correlation between the presence of the *csn2* gene and a particular CRISPR repeat sequence (43).

The simulations reveal sliding motions of Csn2 along the DNA of a magnitude of $>10\,\text{Å}$ (Figure 5A), which is equivalent to about one-third of the B-DNA pitch. The motion pattern shows both inward and outward movements accompanied by repeated changes in the direction of motion, as expected for a process driven by thermal energy. Yet, during the last 70 ns of the simulations, the Csn2 molecules prefer to be more closely located to the center of the DNA: although distances up to 12 Å away from the center are observed during the simulations (Figure 5A), the combined likelihood from both trajectories of Csn2 being within 6 Å of the center is 66%. Thus, with the present length of the simulations, no preference for Csn2 to move toward one of the DNA ends could be detected.

The sliding motions of Csn2 are weakly ($r^2 = 0.21$) but significantly ($P < 0.001$) correlated to a rotation of the protein around the DNA (Figure 5B, D), as found for both independent MD simulations (data for the second

simulation is shown in Supplementary Figure S4). The slope of the correlation line is $\sim13°/\text{Å}$, which is only slightly larger than the turn angle per 1 Å rise of $\sim11°$ for B-DNA (44). The N-termini of helices H3 are likely responsible for this screw motion as the two helices within one dimer of Csn2, first, are almost perfectly collinearly oriented, second, are tilted by $\sim30°$ with respect to the axis of the dsDNA binding within the central hole and, third, possess a Lys (K78) at the N-terminus (Figure 5D). Taken together, this allows the pair of helices H3 of one dimer of Csn2 to fit with their N-termini in between the phosphate backbones of the major groove, while the helix pair of the other dimer is closer to the phosphate backbone of the minor groove. Therefore, on sliding, Csn2 must rotate around the dsDNA for helices H3 need to follow the course of the phosphate backbones, ultimately leading to a screw motion.

To probe the effect of removing Ca^{2+}-ions on Csn2 binding to the DNA, the above MD simulations were repeated without Ca^{2+}. No gross structural distortions of the tetrameric state of Csn2 were observed (Supplementary Figure S5), in agreement with the observed high stability of Csn2 even in the presence of EGTA (30). No release of Csn2 from the DNA was observed either, which is not surprising given the limited size of the water box enclosing the Csn2–DNA complex in the simulation that prevents a slipping down of Csn2. However, in contrast to the simulations with Ca^{2+}, no correlation between a sliding motion of Csn2 and its rotation could be detected in either one of the simulations ($r^2 < 0.07$). Thus, the absence of Ca^{2+} apparently disturbs the screw motion of Csn2, pointing to weakened interactions between Csn2 and DNA.

Both MD simulations furthermore reveal weak ($r^2 = 0.31$) or very weak ($r^2 = 0.06$) but significant ($P < 0.001$) correlations between the translocation of Csn2 from the DNA center and DNA bending in that a large kink angle in the DNA is preferentially observed if Csn2 moves toward a DNA end (Figure 5C and E). The DNA bending is fostered by the insertion of the N-terminus of at least one helix H3 of one of the Csn2 dimers into the major groove, and interactions between the phosphate backbones of the minor groove at two positions with Lys-rich loops between β4 and H2 of the other dimer, respectively (Figure 5E). As no correlation could be detected when plotting the kink angle against the distance of Csn2 from one end of the DNA, no indication was found either that the bending occurs preferentially at one end (and thus at a specific sequence) of the DNA.

DISCUSSION

In this study, we have characterized the DNA end-binding activity of the Cas protein Csn2 and investigated the structure of the Csn2–DNA complexes by atomic force microscopic imaging and MD simulations. Our results assign Csn2 as a new member of the class of multimeric toroidal proteins involved in DNA end-metabolism and indicate a pivotal role of Csn2 as an accessory protein
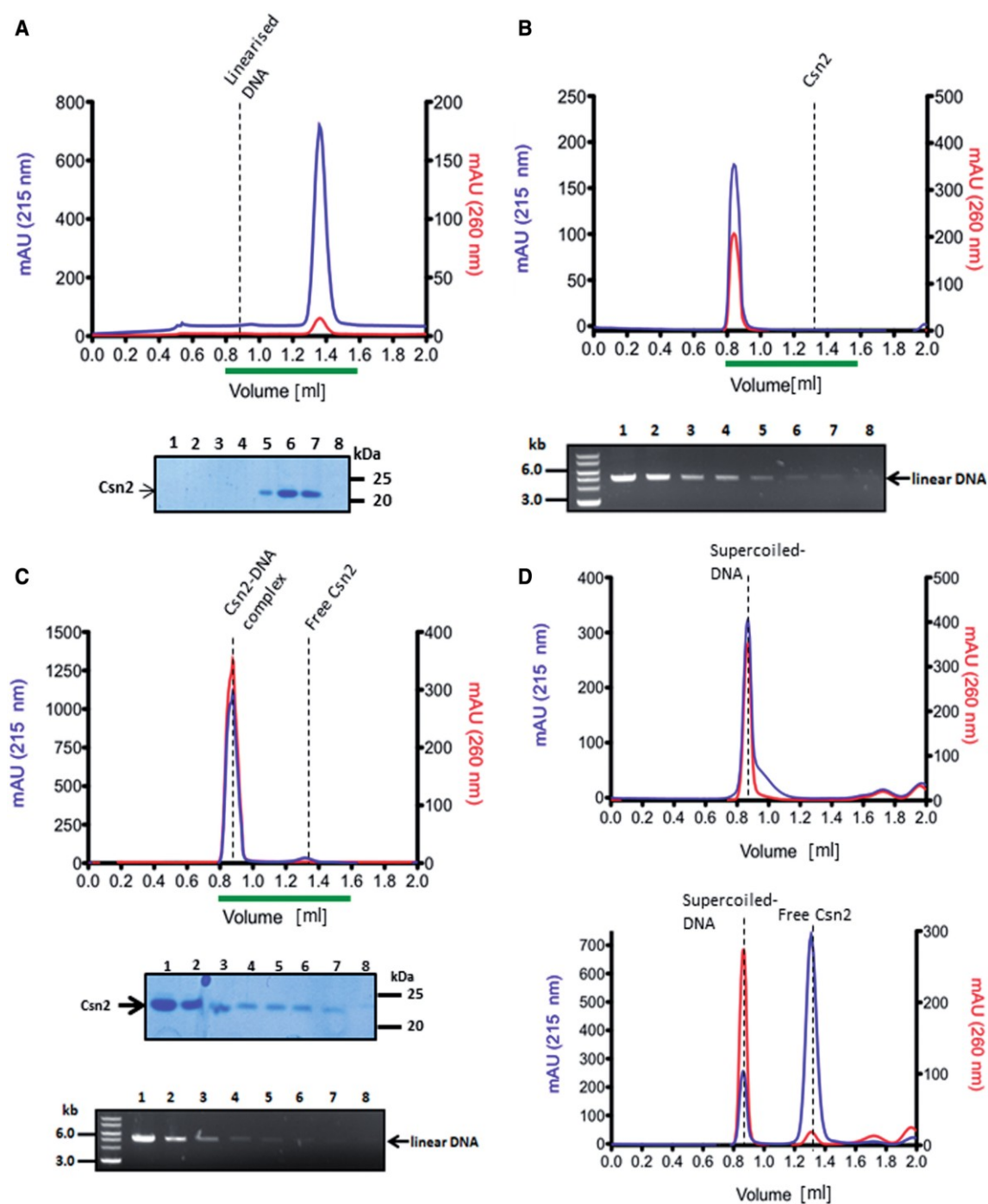
Figure 4. Analytical gel filtration analysis of Csn2 and Csn2-DNA complexes performed with a Superdex 200 PC 3.2/30 column is shown. Elution profiles of 20 μM Csn2 (**A**), 30 nM linearized pCR001 plasmid (**B**) and Csn2-DNA complexes (**C**) are shown. In (**D**) the elution profiles of 30 nM supercoiled pCR001 alone (upper part) or in the presence of 20 μM Csn2 (lower part) are shown. 100 μl fractions were collected starting at an elution volume of 0.8 ml. Aliquots of the fractions 1 to 8, indicated by the green lines below the elution profiles, were analyzed on 10% SDS gels (A, C) and on 1% agarose gels (B, C).
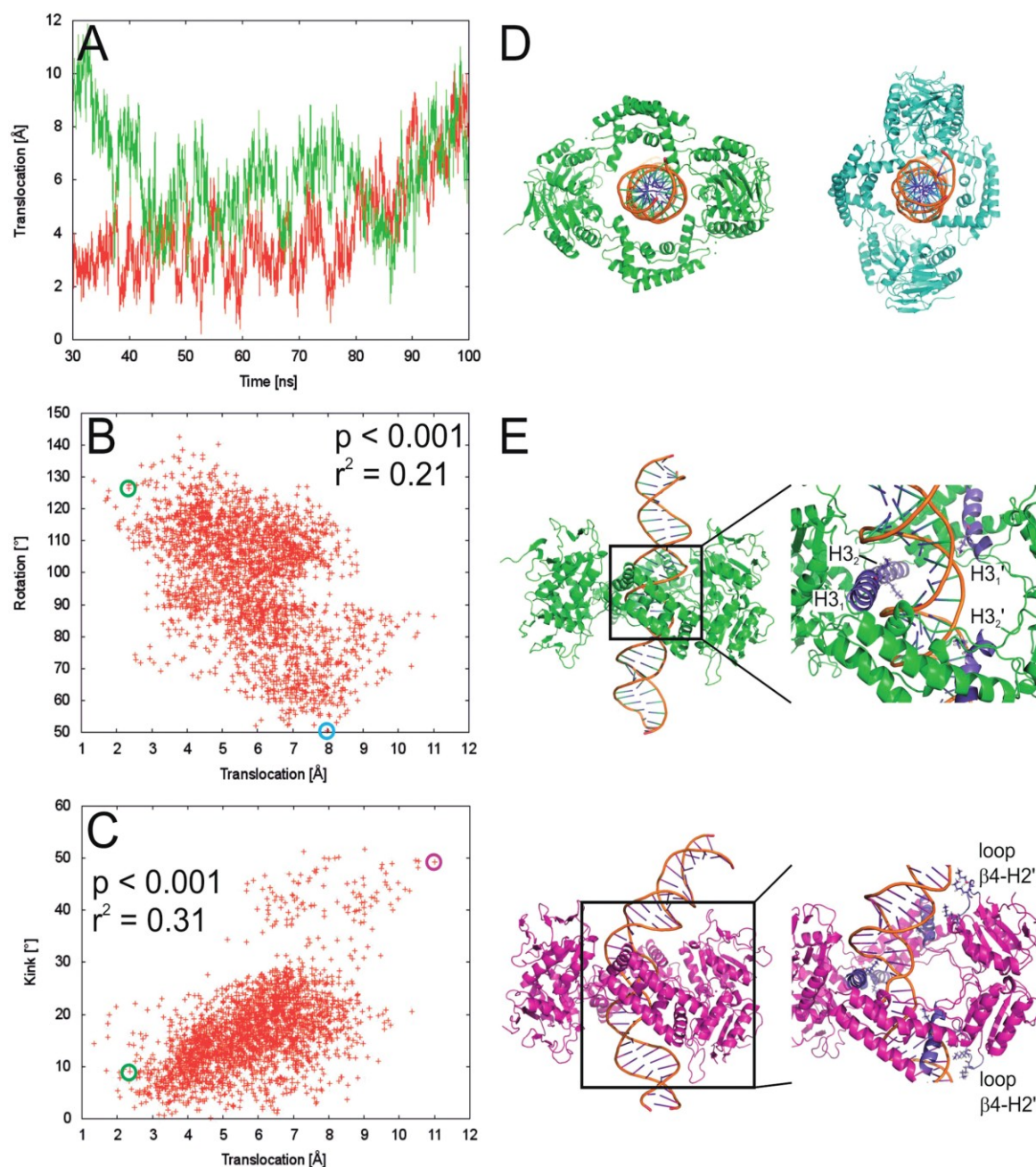
Figure 5. (**A**) Movement of Csn2 along the DNA determined from two independent MD simulations (red and green lines, respectively) over a time course of 70 ns. A translocation value of zero refers to Csn2 being centered on the DNA; translocation values >0 indicate a shift of Csn2 toward the termini of the DNA. (**B**) Coupled motions of translocation and rotation of Csn2 when moving along the DNA over the last 50 ns of the MD simulation. To determine the rotational motion, all conformations of the trajectories were aligned with respect to the phosphorous atoms of the DNA. (**C**) Coupling between DNA bending as determined by a kink angle and the translocation of Csn2 along the DNA over the last 50 ns of the MD simulation. (**D**) View along the DNA axis on Csn2-DNA conformations extracted from the MD trajectory at 50.3 ns (green) and 97.3 ns (cyan). The complexes were aligned with respect to the phosphorous atoms of the DNA and display a rotation of ∼75° of Csn2 during a translocation of ∼5.7 Å. In panel B, the respective data points are marked by circles. (**E**) Side view on Csn2-DNA complexes extracted from the MD trajectory at 50.3 ns (green) and 98.5 ns (magenta). The complexes display a kinking of the DNA by ∼40° during a translocation of Csn2 of ∼8.7 Å. In the close-up figures, helices H3 and loops β4-H2 are colored in blue, and Lys residues in these structural elements are depicted as sticks; unprimed labels mark helices and loops that belong to one dimer, primed labels mark objects that belong to the other dimer. In panel C, the respective data points are marked by circles.

during the integration of spacer sequences into the CRISPR array.

Csn2 tetramers bind at dsDNA ends and slide inward along the DNA

The toroidal structure of the DNA-binding protein Csn2 in different studies led to the suggestion that it binds to DNA through its positively charged central hole and encloses the DNA, although alternative binding mechanisms were also considered (29–31). Furthermore, the crystal structure of the Cas protein Stu0660 of the *S. thermophilus* has revealed the presence of a second group of the Csn2 protein family (Stu0660-like Csn2), the members of which share no sequence similarity but have a highly similar tetrameric structure (32). In contrast to the canonical Csn2 proteins, the Csn2-like Stu0660 protein has an extended C-terminal domain and binds to dsDNA without the need of Ca^{2+}-ions. Moreover, due to the lack of a mobility shift with a circular DNA, the authors suggested that Stu0660 'does not or barely' binds to circular DNA but selectively to linear dsDNA (32). Our study is consistent with this suggestion and clearly demonstrates that the Csn2 tetramer binds DNA through its positively charged central hole from the DNA ends and then slides inward along the DNA.

The sliding activity of Csn2 tetramers is not energy-dependent and thus likely not a directed translocation but rather a thermal energy-driven random walk fostered by electrostatic interactions between positively charged amino acids of Csn2 and the negatively charged phosphate backbone of the DNA. MD simulations revealed a rotation of Csn2 when moving along the helical axis of the DNA, similar to a rotation-coupled sliding of the sliding clamp PCNA (45), endonuclease EcoRV (46) or human oxoguanine DNA glycosylase 1 (47). One-dimensional diffusion of DNA-binding proteins is considered to facilitate finding of the specific binding sites (48). Although we have no evidence for a sequence-specific interaction *in vitro*, the reported correlation of the presence of the *csn2* gene and a particular CRISPR repeat sequence (43) could point to a specific role of Csn2 at the repeat sequence. The MD simulations were performed with the CRISPR repeat sequence, and one of the simulations indicated a bending of the DNA preferentially when Csn2 was located at an end. AFM is in principle suitable to study protein-induced DNA bending (49); however, the predicted Csn2-induced kinking is located at the ends of the DNA, and thus difficult to detect with AFM. Therefore, the structural deformation of the DNA ends on Csn2-binding, as proposed by the MD simulations, needs further experimental validation.

The cooperative binding of Csn2, observed in the EMSA and AFM analyses, indicates that multiple Csn2 tetramers are needed at the DNA ends to form stable Csn2–DNA complexes. We suggest that the binding of the first Csn2 tetramer at one DNA end is stabilized by loading of additional tetramers onto the same DNA end, which could impede a slipping down of the initially bound tetramers from the DNA end. This is consistent with the observation that the dissociation of Csn2–DNA complexes occurs through slipping down of the sliding tetramers at the DNA termini (Figure 2). In addition, at high concentrations, Csn2 tends to cluster along the DNA, leading to highly condensed nucleoprotein complexes. This apparent DNA-dependent self-association of the tetramers could point to a physical contact of the tetramers, contributing to the cooperative DNA end-loading of Csn2, as known for the Ku protein (50). The heterodimeric Ku protein forms a ring-like structure, binds preferentially at DNA ends, encircles the DNA (51), can slide along the DNA and promotes looped DNA structures (52). Moreover, it tends to cluster along the DNA as visualized by AFM (53). The DNA-binding properties of Ku are thus similar to the features of Csn2 observed in this study. The function of Ku *in vivo* is to mediate the formation of DNA end-synapsis and recruitment of recombination proteins to allow the repair of double-strand DNA breaks (54).

DNA end-metabolism and spacer integration

A series of multimeric ring-shaped proteins is known, which meet different functions in DNA metabolism, such as replication, transcription, recombination or DNA repair (40,41). In principle, there are three general ways how the DNA can get encircled by a ring-shaped protein: the quaternary structure can be opened wide enough to bind and enclose the DNA [e.g. RdgC (55)], the ring-shaped structure gets directly assembled around the DNA [e.g. PCNA (56)] or the DNA end has to pass through the central hole of the protein [e.g. Ku protein (51)]. In the latter case, the function of the protein is directly linked to DNA ends. Thus, the results presented here strongly suggest that the biological function of the CRISPR adaptation protein Csn2 is related to free DNA ends, which are likely formed during spacer acquisition.

The study of Yosef *et al.* (27) characterized a minimal requirement for spacer integration in type I-E CRISPR systems, and demonstrated the involvement of leader DNA and the nucleases Cas1 and Cas2 in the immunization process. The uptake of new spacers at the leader proximal end is also observed for other CRISPR subtypes, and together with the universality of Cas1 and Cas2, it seems that the main principle mechanism for spacer acquisition is similar in the different CRISPR subtypes. However, some of the CRISPR-Cas subtypes require in addition to Cas1 and Cas2 the proteins Csn2 or Cas4 for the acquisition of new spacer sequences. The Cas4 protein is proposed to be a subunit of the multiprotein complex Cascis (CRISPR-associated complex for the integration of spacers), mediating the CRISPR-mediated immunization (57). A recent study of the Cas4 protein from *Sulfolobus solfataricus* demonstrated that it contains a 5′–3′ DNA exonuclease activity, implying the requirement of DNA end modification/resection during integration of new spacer sequences into the CRISPR array (58). Moreover, the crystal structure of the Cas4 protein from *S. solfataricus* has been deposited in PDB database, showing that the Cas4 protein forms, like Csn2, a ring-shaped structure (PDB ID: 4IC1). Thus, the structural
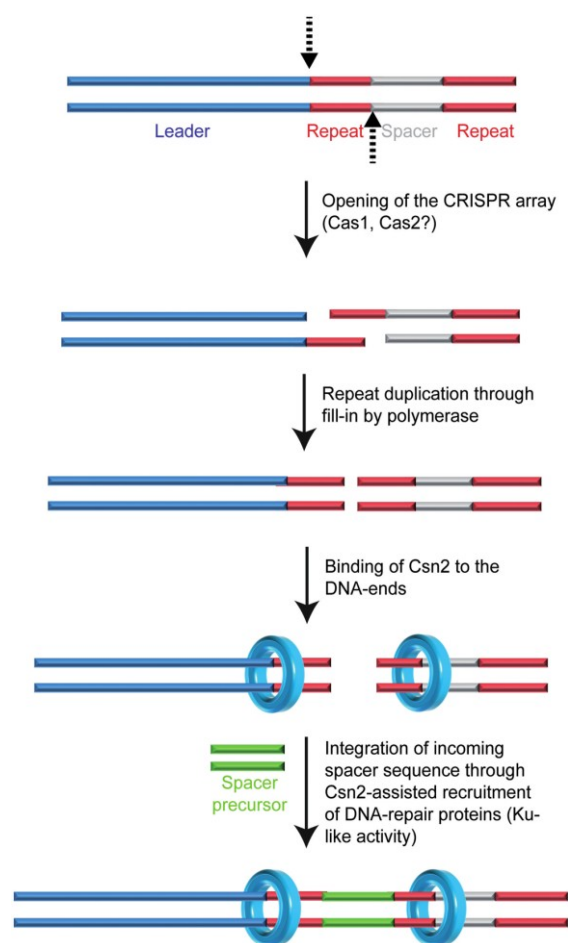
Figure 6. A model for the proposed function of Csn2 DNA end-binding in spacer integration is shown. The leader proximal repeat sequence serves as template for duplication of the repeat sequence (25,26,64), suggesting a cleavage at the leader-repeat and repeat-spacer borders (indicated by the dashed arrows). Complementary strand synthesis and binding of Csn2 to the DNA ends could assist in the integration of new spacer DNA fragments through holding the DNA ends together, while simultaneously recruiting DNA-repair proteins and protecting the double-stranded DNA breaks from exonucleolytic degradation.

similarity of Cas4 to AddB, its RecB-like exonuclease activity (58) and the DNA end-binding of Csn2 presented in this study are in agreement with DNA end-processing steps during the spacer integration. Another indication for the participation of DNA end-processing proteins in the adaptation stage is the genetic association of *nurA* and *herA* genes with type III-A CRISPR operons in Thermoproteales (59), encoding for an exonuclease–helicase complex involved in DNA end-resection (60).

The lack of enzymatic activity of Csn2 suggests that Cas4 and Csn2 are likely not functional homologs. However, the conserved co-localization within the type II systems and the DNA end-related activities of both proteins indicate that they are involved in DNA end-

resection processes during integration of spacers. The reported protein–protein interaction of Cas1 with RecB in *E. coli* (61) and the AddB-like structure and activity of the Cas4 protein (58) could support an involvement of DNA repair systems during the uptake of CRISPR spacer sequences. The study of Chayot *et al.* (62) has demonstrated that the insertion of non-related exogenous sequences in *E. coli* depends on the RecBCD complex. The role of RecBCD is to process the DNA ends exonucleo-lytically on the 3′-end to produce microhomology with the incoming DNA, a mechanism termed alternative end-joining (A-EJ) (62). A similar mechanism could be catalyzed by Cas4 in type II-B or AddAB proteins in type II-A systems lacking Cas4.

The DNA end-binding activity of Csn2 suggests that it has a function as accessory protein on DNA ends. Csn2 could fulfill a task analogous to Ku during acquisition of exogenous DNA by non-homologous end-joining, by binding to the double-strand breaks to hold the ends to-gether and/or by recruiting recombination proteins (63) (Figure 6). Alternatively, the DNA end binding activity of Csn2 could be related to the nuclease activity of Cas9, introducing specific double-strand DNA breaks into the target DNA (12,17). It is possible that after rec-ognition and cleavage of the foreign DNA by Cas9-tracrRNA:crRNA ribonuleoprotein complex, Csn2 proteins are loaded at the DNA ends, e.g. to mark the foreign DNA as a substrate for new spacer sequences and/or to recruit the nucleases for specific generation of new spacer precursors. Indeed, such a coupling of the CRISPR interference with the immunization stage has been described to occur in the type I-E systems, known as 'primed acquisition' of spacer sequences (25). Although our analyses do not directly address the mechanisms of CRISPR spacer integration, the DNA end-binding of the adaptation protein Csn2 and the structure of the Csn2–DNA complexes resemble known recombination proteins acting on DNA end-metabolism. Therefore, we propose that the integration of new spacer DNA may occur by a recombination mechanism similar to non-homologous end-joining.
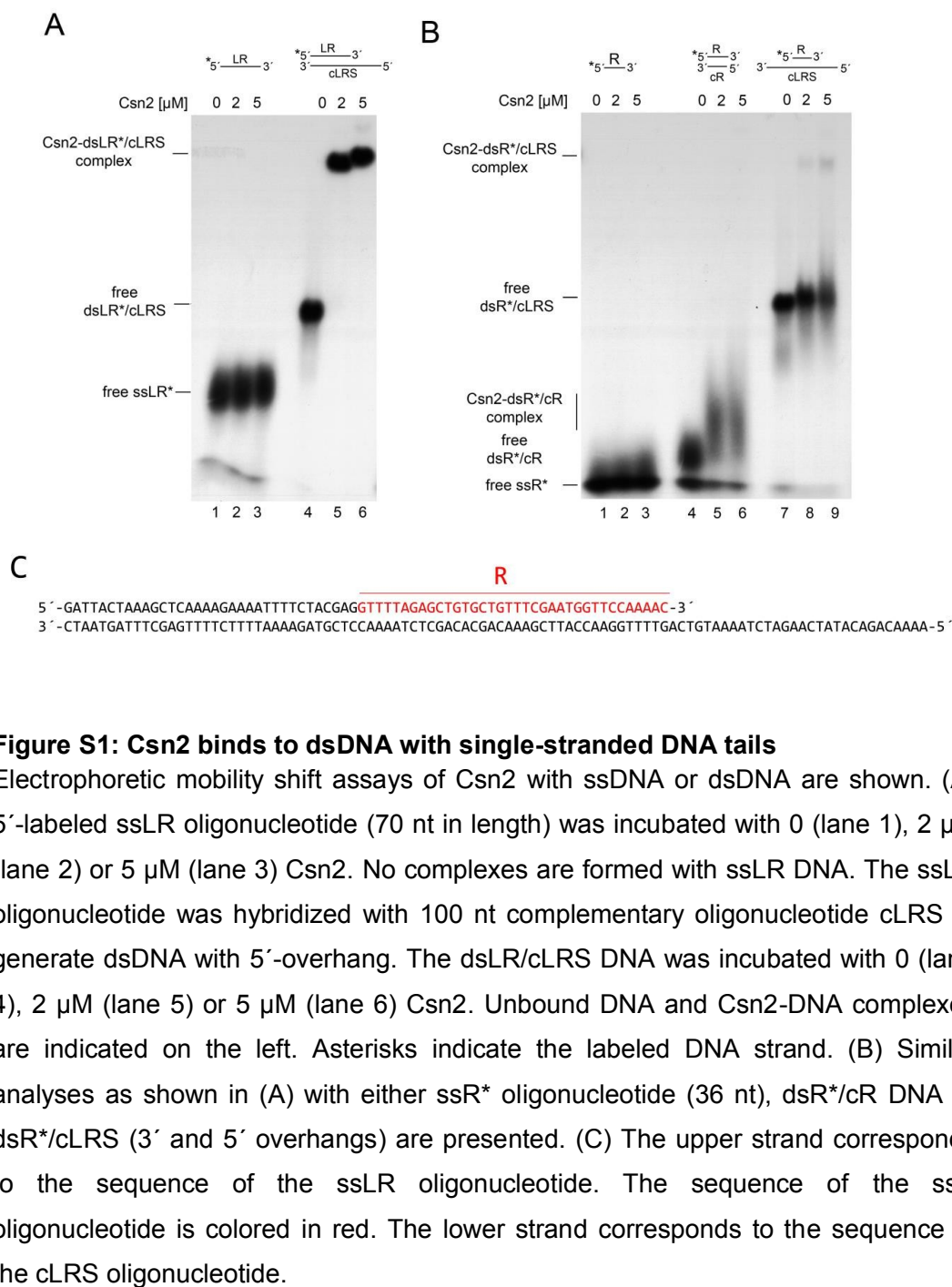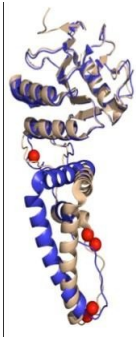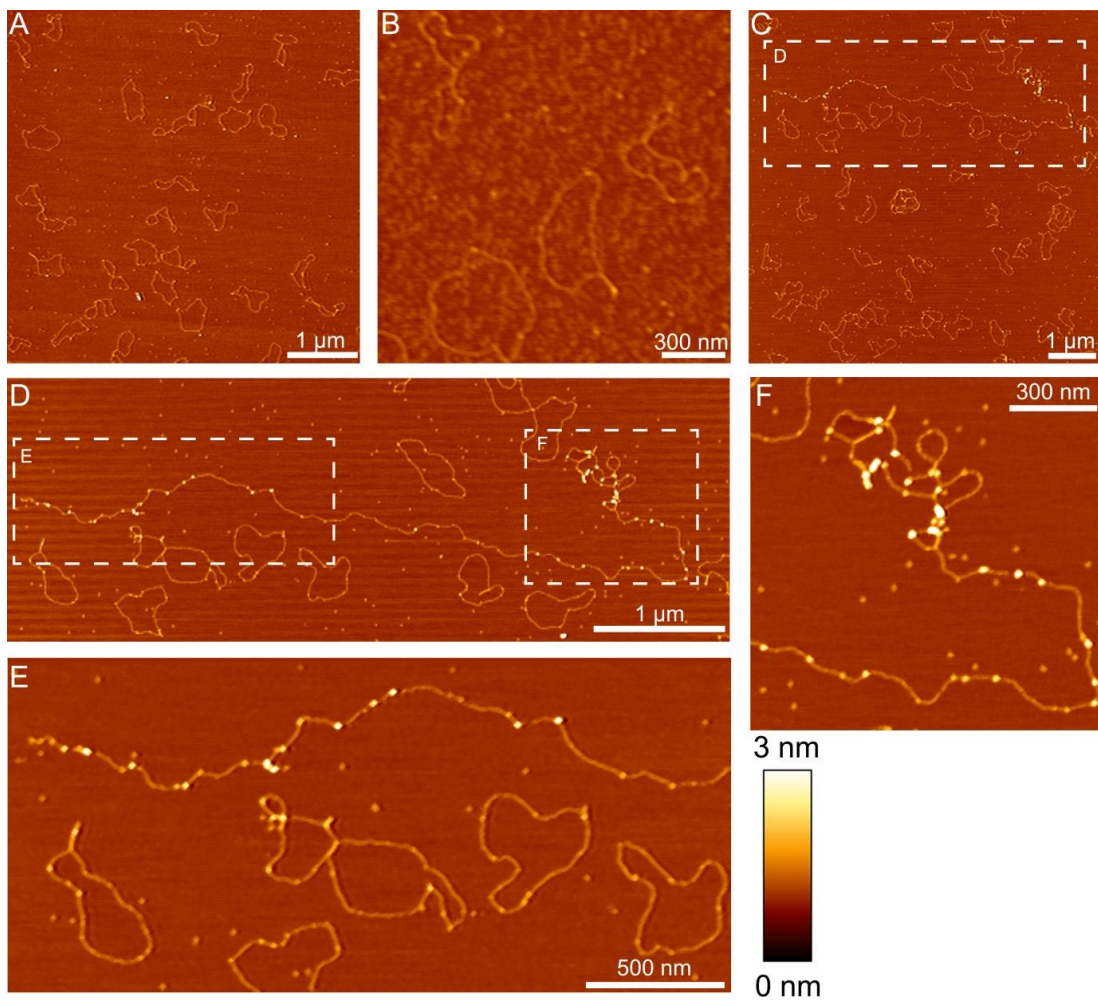
REFERENCES

1. Bhaya,D., Davison,M. and Barrangou,R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.*, **45**, 273–297.
2. Al-Attar,S., Westra,E.R., van der Oost,J. and Brouns,S.J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.*, **392**, 277–289.
3. Wiedenheft,B., Sternberg,S.H. and Doudna,J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331–338.
4. Makarova,K.S., Haft,D.H., Barrangou,R., Brouns,S.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J., Wolf,Y.I., Yakunin,A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
5. Makarova,K.S., Grishin,N.V., Shabalina,S.A., Wolf,Y.I. and Koonin,E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.*, **1**, 7.
6. van der Oost,J., Jore,M.M., Westra,E.R., Lundgren,M. and Brouns,S.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.*, **34**, 401–407.
7. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
8. Brouns,S.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J., Snijders,A.P., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
9. Westra,E.R., van Erp,P.B., Kunne,T., Wong,S.P., Staals,R.H., Seegers,C.L., Bollen,S., Jore,M.M., Semenova,E., Severinov,K. *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.
10. Sapranauskas,R., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.*, **39**, 9275–9282.
11. Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
12. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
13. Hale,C.R., Majumdar,S., Elmore,J., Pfister,N., Compton,M., Olson,S., Resch,A.M., Glover,C.V. III, Graveley,B.R., Terns,R.M. *et al.* (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell*, **45**, 292–302.
14. Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945–956.
15. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R. *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
16. Haft,D.H., Selengut,J., Mongodin,E.F. and Nelson,K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.*, **1**, e60.
17. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA*, **109**, E2579–E2586.
18. Jinek,M., East,A., Cheng,A., Lin,S., Ma,E. and Doudna,J. (2013) RNA-programmed genome editing in human cells. *Elife*, **2**, e00471.
19. Cho,S.W., Kim,S., Kim,J.M. and Kim,J.S. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, **31**, 230–232.
20. Jiang,W., Bikard,D., Cox,D., Zhang,F. and Marraffini,L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.*, **31**, 233–239.
21. Hwang,W.Y., Fu,Y., Reyon,D., Maeder,M.L., Tsai,S.Q., Sander,J.D., Peterson,R.T., Yeh,J.R. and Joung,J.K. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.*, **31**, 227–229.
22. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
23. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
24. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
25. Datsenko,K.A., Pougach,K., Tikhonov,A., Wanner,B.L., Severinov,K. and Semenova,E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, **3**, 945.
26. Swarts,D.C., Mosterd,C., van Passel,M.W. and Brouns,S.J. (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One*, **7**, e35888.
27. Yosef,I., Goren,M.G. and Qimron,U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.*, **40**, 5569–5576.
28. Deveau,H., Barrangou,R., Garneau,J.E., Labonte,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
29. Nam,K.H., Kurinov,I. and Ke,A. (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *J. Biol. Chem.*, **286**, 30759–30768.
30. Ellinger,P., Arslan,Z., Wurm,R., Tschapek,B., MacKenzie,C., Pfeffer,K., Panjikar,S., Wagner,R., Schmitt,L., Gohlke,H. *et al.* (2012) The crystal structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. *J. Struct. Biol.*, **178**, 350–362.
31. Koo,Y., Jung,D.K. and Bae,E. (2012) Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. *PLoS One*, **7**, e33401.
32. Lee,K.H., Lee,S.G., Eun Lee,K., Jeon,H., Robinson,H. and Oh,B.H. (2012) Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins*, **80**, 2573–2582.
33. Pul,U., Wurm,R. and Wagner,R. (2007) The role of LRP and H-NS in transcription regulation: involvement of synergism, allostery and macromolecular crowding. *J. Mol. Biol.*, **366**, 900–915.
34. Case,D.A., Cheatham,T.E. III, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
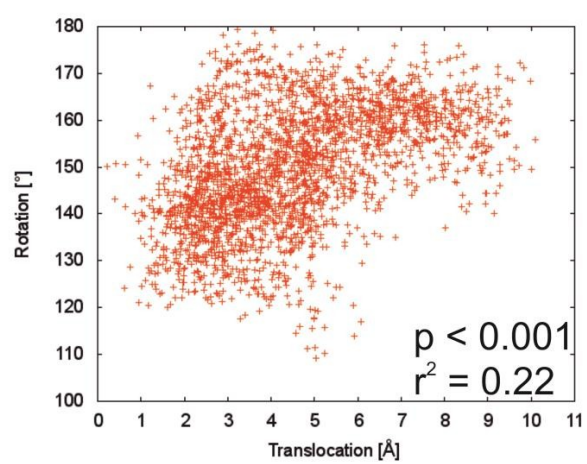
35. Cornell,W.D., Cieplak,C.I., Bayly,I.R., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

36. Simmerling,C., Strockbine,B. and Roitberg,A.E. (2002) All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, **124**, 11258–11259.

37. Jorgensen,W.L., Chandrasekhar,J., Madura,J. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

38. Darden,T., York,D. and Pederson,L. (1993) Particle Mesh Ewald—a Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

39. Ryckaert,J.P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, **23**, 327–341.

40. Hingorani,M.M. and O'Donnell,M. (1998) Toroidal proteins: running rings around DNA. *Curr. Biol.*, **8**, R83–R86.

41. Hingorani,M.M. and O'Donnell,M. (2000) A tale of toroids in DNA metabolism. *Nat. Rev. Mol. Cell. Biol.*, **1**, 22–30.

42. Tessmer,I., Moore,T., Lloyd,R.G., Wilson,A., Erie,D.A., Allen,S. and Tendler,S.J. (2005) AFM studies on the role of the protein RdgC in bacterial DNA recombination. *J. Mol. Biol.*, **350**, 254–262.

43. Kunin,V., Sorek,R. and Hugenholtz,P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.*, **8**, R61.

44. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.

45. Kochaniak,A.B., Habuchi,S., Loparo,J.J., Chang,D.J., Cimprich,K.A., Walter,J.C. and van Oijen,A.M. (2009) Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.*, **284**, 17700–17710.

46. Dikic,J., Menges,C., Clarke,S., Kokkinidis,M., Pingoud,A., Wende,W. and Desbiolles,P. (2012) The rotation-coupled sliding of EcoRV. *Nucleic Acids Res.*, **40**, 4064–4070.

47. Blainey,P.C., Luo,G., Kou,S.C., Mangel,W.F., Verdine,G.L., Bagchi,B. and Xie,X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.*, **16**, 1224–1229.

48. von Hippel,P.H. and Berg,O.G. (1989) Facilitated target location in biological systems. *J. Biol. Chem.*, **264**, 675–678.

49. Dame,R.T., van Mameren,J., Luijsterburg,M.S., Mysiak,M.E., Janicijevic,A., Pazdzior,G., van der Vliet,P.C., Wyman,C. and Wuite,G.J. (2005) Analysis of scanning force microscopy images of protein-induced DNA bending using simulations. *Nucleic Acids Res.*, **33**, e68.

50. Ma,Y. and Lieber,M.R. (2001) DNA length-dependent cooperative interactions in the binding of Ku to DNA. *Biochemistry*, **40**, 9638–9646.

51. Walker,J.R., Corpina,R.A. and Goldberg,J. (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, **412**, 607–614.

52. Dynan,W.S. and Yoo,S. (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res.*, **26**, 1551–1559.

53. Cary,R.B., Peterson,S.R., Wang,J., Bear,D.G., Bradbury,E.M. and Chen,D.J. (1997) DNA looping by Ku and the DNA-dependent protein kinase. *Proc. Natl Acad. Sci. USA*, **94**, 4267–4272.

54. Feldmann,E., Schmiemann,V., Goedecke,W., Reichenberger,S. and Pfeiffer,P. (2000) DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining. *Nucleic Acids Res.*, **28**, 2585–2596.

55. Briggs,G.S., Yu,J., Mahdi,A.A. and Lloyd,R.G. (2010) The RdgC protein employs a novel mechanism involving a finger domain to bind to circular DNA. *Nucleic Acids Res.*, **38**, 6433–6446.

56. Kong,X.P., Onrust,R., O'Donnell,M. and Kuriyan,J. (1992) Three-dimensional structure of the beta subunit of E. *coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell*, **69**, 425–437.

57. Plagens,A., Tjaden,B., Hagemann,A., Randau,L. and Hensel,R. (2012) Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon T*hermoproteus tenax.* *J. Bacteriol.*, **194**, 2491–2500.

58. Zhang,J., Kasciukovic,T. and White,M.F. (2012) The CRISPR associated protein cas4 Is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS One*, **7**, e47232.

59. Bernick,D.L., Cox,C.L., Dennis,P.P. and Lowe,T.M. (2012) Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum. *Front. Microbiol.*, **3**, 251.

60. Blackwood,J.K., Rzechorzek,N.J., Abrams,A.S., Maman,J.D., Pellegrini,L. and Robinson,N.P. (2012) Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex. *Nucleic Acids Res.*, **40**, 3183–3196.

61. Babu,M., Beloglazova,N., Flick,R., Graham,C., Skarina,T., Nocek,B., Gagarinova,A., Pogoutse,O., Brown,G., Binkowski,A. et al. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.*, **79**, 484–502.

62. Chayot,R., Montagne,B., Mazel,D. and Ricchetti,M. (2010) An end-joining repair mechanism in *Escherichia coli. Proc. Natl Acad. Sci. USA*, **107**, 2141–2146.

63. Doherty,A.J. and Jackson,S.P. (2001) DNA repair: how Ku makes ends meet. *Curr. Biol.*, **11**, R920–R924.

64. Goren,M.G., Yosef,I., Auster,O. and Qimron,U. (2012) Experimental definition of a clustered regularly interspaced short palindromic duplicon in *Escherichia coli. J. Mol. Biol.*, **423**, 14–16.

SUPPLEMENTARY DATA

Double-strand DNA end binding and sliding of the toroidal CRISPR-associated protein Csn2

Zihni Arslan[1], Reinhild Wurm[1], Oleksandr Brener[1,2], Philipp Ellinger[3], Luitgard Nagel-Steger[1,2], Filipp Oesterhelt[1], Lutz Schmitt[3], Dieter Willbold[1,2], Rolf Wagner[1], Holger Gohlke[4], Sander H.J. Smits[3] and Ümit Pul[1,*]

1

Figure S1: Csn2 binds to dsDNA with single-stranded DNA tails

Electrophoretic mobility shift assays of Csn2 with ssDNA or dsDNA are shown. (A) 5´-labeled ssLR oligonucleotide (70 nt in length) was incubated with 0 (lane 1), 2 µM (lane 2) or 5 µM (lane 3) Csn2. No complexes are formed with ssLR DNA. The ssLR oligonucleotide was hybridized with 100 nt complementary oligonucleotide cLRS to generate dsDNA with 5´-overhang. The dsLR/cLRS DNA was incubated with 0 (lane 4), 2 µM (lane 5) or 5 µM (lane 6) Csn2. Unbound DNA and Csn2-DNA complexes are indicated on the left. Asterisks indicate the labeled DNA strand. (B) Similar analyses as shown in (A) with either ssR* oligonucleotide (36 nt), dsR*/cR DNA or dsR*/cLRS (3´ and 5´ overhangs) are presented. (C) The upper strand corresponds to the sequence of the ssLR oligonucleotide. The sequence of the ssR oligonucleotide is colored in red. The lower strand corresponds to the sequence of the cLRS oligonucleotide.

α

~

<div style="background:black;color:white">

Chapter 3

</div>

Structure and function of hepatic ABC transporters

Published in: *Hepatobiliary Transport in Health and Disease, Editors,*
                  D.Häussinger, R. Kubitz and V. Keitel, de Gruyter,
                  Berlin

Impact factor: not available

Own Proportion

 to this work: Writing of the manuscript

2 Structure and function of hepatic ABC transporters

Philipp Ellinger, Marianne Kluth, Susanne Przybylla, Sander H. J. Smits, and Lutz Schmitt

2.1 Introduction to human ABC transporters expressed in the liver

Several membrane transporters that belong to a group of ATP-dependent primary transporters, the so-called ABC (ATP binding cassette) transporters, are found in the human genome. In general, ABC transporters contain two transmembrane-spanning domains (TMDs) and two characteristic nucleotide-binding domains (NBDs) localized in the cytosol. In the membrane the two TMDs form a pore-like structure, which facilitates substrate transport against a chemical gradient. One TMD is predicted to have six α-helices, whereas the soluble NBDs are essential for the supply of energy by hydrolysis of ATP. Compared with the TMD, the NBD harbors highly conserved sequence motifs: the Walker A (GXXGXGKS/T, where X can be for any amino acid), Walker B ($\Phi\Phi\Phi\Phi D$, where Φ can be any hydrophobic residue) motifs, and the C-loop (ABC–signature motif, LSGGQ) (1). The C loop, which is located roughly 90 amino acids downstream of the Walker A motif and roughly 30 amino acids upstream of the Walker B motif, is actually the characteristic sequence motif of this family; together with the Walker A and B motifs, it serves as a diagnostic clue to the identification of new family members. Additional sequence motifs present in ABC transporters are the Q loop, the D loop (SALD), and a highly conserved histidine residue essential for ATP hydrolysis, which is positioned 30 amino acid downstream of the D loop (2).

To achieve a thermodynamic uphill transport of the substrate, transport has to be coupled to the cycle of ATP hydrolysis. Several high-resolution structures of full-length ABC transporters and isolated NBDs, in combination with biochemical analysis, have provided important contributions to a molecular understanding of substrate binding, ATP hydrolysis, and substrate transport. For example, the highly conserved NBD has an L-shaped structure consisting of a catalytic domain and a helical domain. The catalytic domain contains the Walker A and B motifs while the helical domain harbors the C loop. These two domains are connected by the Q and Pro loops (3). Further analysis of, for example, the isolated haemolysin B-NBD demonstrated that in the presence of ATP, the two NBDs form a homodimer (2). The Walker A and B motifs of one NBD and the C loop of the opposing NBD bind one ATP, so that the two NBDs are set in a head-to-tail arrangement. This ATP-induced dimerization generates mechanical work, which in principle can be transmitted to the TMDs and might serve as another source of energy (see section 2.2.4). The dimeric NBDs cooperate in hydrolysing ATP and provide the free energy to drive the directional transport of the substrate against a concentration gradient. After ATP hydrolysis, ADP and P_i dissociate from the NBD, the dimer falls apart, and the ground state of the NBDs is restored.

Furthermore, different models for the transport mechanism have been proposed. The simplest model, the alternating access model, describes two basic conformations. One conformation is open to the cytosolic side (inward-facing), with a substrate-binding

site accessible for the substrate only from the cytosol, and the second conformation is open to the extracellular side, containing a binding site, which has a low affinity to the substrate and is accessible only from the extracellular space (4). A continuous model is the ATP-switch model. In the ground state the transporter is in the inward-facing conformation with a high-affinity substrate-binding site and the NBDs exist as monomers, with low affinity to ATP. The ATP-induced dimerization of the NBDs leads to a conformational change in the TMDs such that the substrate-binding site is exposed to the extracellular space, the substrate affinity is reduced, and the bound substrate is finally released (5). However, the exact molecular coupling of the ATP–hydrolysis cycle and substrate transport is still not entirely clear.

In the human hepatocyte, several ABC transporter are expressed: for example, the bile salt export pump (BSEP, ABCB11), responsible for bile salt transport; ABCG5/ABCG8, involved in sterol transport; multidrug resistance protein 3 (MDR3, ABCB4), flopping phosphatidylcholine from the inner to the outer membrane leaflet; and ABCG2, transporting a variety of hydrophobic substances (▶Fig. 2.1). Mutations in one of these transporters are associated with different kinds of liver diseases of varying severity. For example, Dubin-Johnson disease is related to mutations in MRP2 (ABCC2), and progressive familial intrahepatic cholestasis type 2 (PFIC2) is associated with a mutations with the bile salt export pump BSEP.

This chapter summarizes experimental insights and focuses on the canalicular ABC transporters BSEP, MDR3, and ABCG2, highlighting their discovery and evolution and the *in vitro* assays from which a mechanistic understanding may be derived.
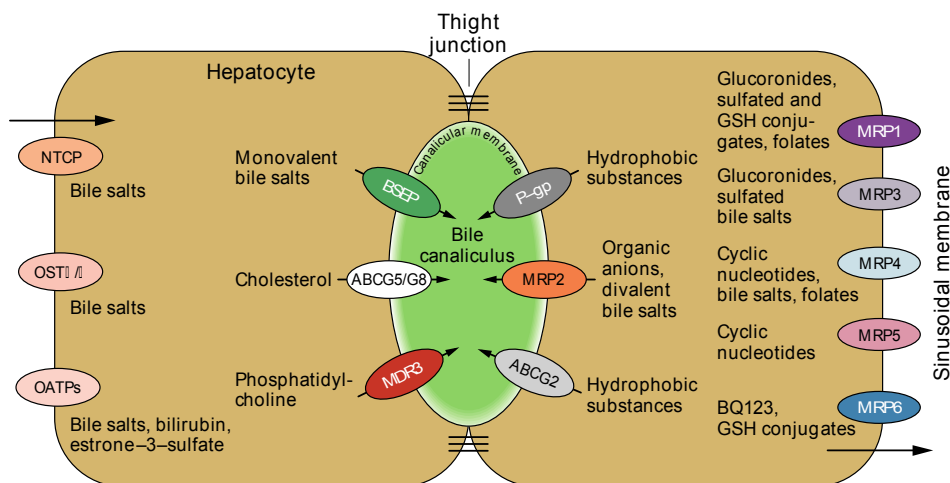
Fig. 2.1: Localization of transporters in the hepatocytes. Bile salts are taken up at the sinusoidal (basolateral) membrane through the sodium–taurocholate cotransporting peptide (NTCP) in a sodium–dependent manner and to a lesser extent through a sodium–independent transport by organic anion transporting proteins (OATPs). They are then further shuttled to the canalicular membrane and transported via the bile salt export pump (BSEP) into the canaliculus. Multidrug resistance protein 3 (MDR3) and Sterolin 1

Fig. 2.1: (*Continued*)

(ABCG5/G8) complete the bile formation by flopping phosphatidylcholine (PC) from the inner to the outer leaflet as well as transporting cholsterol. Bile salts, PC and cholesterol form mixed micelles which constitute the basis of bile. P–glycoprotein (P–gp) and the breast cancer resistance protein (ABCG2) transport a variety of hydrophobic substances into the bile and confer multidrug resistance (MDR). Furthermore, a substanial number of multidrug–related proteins (MRPs) are localized in the sinusoidal membrane except MRP2. They transport a broad range of organic anions and conjugated substances. They also participate in MDR and some of them, e.g. MRP4 as well as the organic solute transporter (OST) act as salvage system for too high bile salt concentrations within the cell to prevent toxicity.

2.2 Structure and function of the bile salt export pump (ABCB11; BSEP)

2.2.1 Liver transport of bile salts

Bile salts are essential for the absorption of lipids and fat-soluble vitamins, originated from food intake, by the enterocytes of the small intestine and also for the excretion of endo- and xenobiotics with the bile. They are synthesized by multiple enzymatic reactions in the liver, more precisely in the hepatocytes from cholesterol as educt; this constitutes one of the key function of the liver (6). From there bile salts enter the biliary tree and are stored in the gallbladder upon food intake (7). After they have fulfilled "their mode of action", bile salts pass through the enterohepatic circulation, meaning that they are reabsorbed to ~90% in the small intestine and then transported back to the liver via the portal blood. There, they are transported again into the hepatocyte and the cycle starts anew with their secretion into the canaliculi (8,9). A single bile salt molecule traverses the cycle approximately up to 10 times a day until it is excreted via the intestine, which makes this circulation an extremely efficient recycling system (10).

Because bile salts are amphipathic molecules, they display a detergent character. Hence a high concentration within the cell is deleterious, leading to damaged mitochondria and apoptosis or necrosis of the hepatocytes owing to the salts' ability to solubilize or create defects within biological membranes. To prevent this and keep bile salts circulating, a specialized set of bile salt transporters in the hepatocyte is required (9,11,12). In the basolateral membrane (also called the sinusoidal membrane), bile salts are taken up from the portal blood. This is accomplished by the sodium taurocholate co-transporting peptide (NTCP, SLC10A1) in a sodium-dependent transport process (13). In addition, there is the less frequently used sodium-independent transport by the organic anion-transporting polypeptides (OATPs) (14).

After entering the cell, bile salts reach the apical membrane (also called the canalicular membrane); the exact mechanism of this is not yet completely understood. For example, one mechanism involves bile salt−binding proteins (15).

At the canalicular membrane, bile salts are transported into the canalicular lumen by the ATP-binding cassette transporter (ABC transporter) bile salt export pump (ABCB11;

BSEP) (16,17). BSEP is the main driving force for the bile salt−dependent part of bile flow and a bottleneck in the enterohepatic circulation. It must transport bile salts against a steep concentration gradient to maintain circulation, since the concentration of bile salts in the canaliculus is 1000 fold higher than in the cell, 1 mM and 1 µM, respectively (10).

2.2.2 Discovery of the bile salt export pump

The electrochemical gradient across the canalicular membrane is ∼ −35 mV and its discovery marked the first explanation for bile salt transport across this membrane (18,19). However, this electrochemical gradient alone could not be the entire explanation. Finally, in 1991, an ATP-dependent system for the transport of taurocholate in isolated canalicular membranes of rat liver was described. Other laboratories subsequently confirmed this finding (20,21). Evidence that an ABC transporter was responsible for bile salt secretion into the canaliculus appeared in 1995. It involved an increased level of mRNAs, detected by Northern blotting, in combination with the overexpression of an ABC transporter found via the Western blot technique with a P-glycoprotein antibody (ABCB1, MDR1, P-gp); these were demonstrated in a bile salt−resistant rat hepatoma−derived cell line (22). This suggested that an ABC transporter closely related to P-gp became upregulated in this system. In the same year, Childs et al. screened a pig cDNA library with a probe consisting of a P-gp sequence and identified a gene exclusively expressed in the liver that had a sequence identity of 61% to human P-gp on the amino acid level (23). This gene was named "sister of P-gp" (sP-gp), but its function remained unknown. Gerloff et al. were the first to demonstrate that oocytes exhibited a stimulated taurocholate efflux when liver sP-gp cRNA was injected into *Xenopus laevis* oocytes and the first to express sP-gp in *Sf*9 (*Spodoptera frugiperda*) cells (24). Furthermore, membrane vesicles derived from these *Sf*9 cells demonstrated an ATP-dependent taurocholate uptake, much as in previous studies with isolated canalicular membranes. Because of these findings the "sister of P-gp" was renamed "bile salt export pump (BSEP)" and was considered to be the predominant bile salt transporter in the apical membranes of hepatocytes (24). Further strong support for this consideration was obtained by positional cloning of the human *BSEP* gene and mapping it to chromosome 2q24, a locus linked to progressive familial intrahepatic cholestasis type 2 (PFIC2), a severe liver disease (25).

2.2.3 Evolution of the bile salt export pump

The production of bile salts and their subsequent transport into the canaliculi is highly conserved among the livers of vertebrates. Over the years, BSEP has been detected and studied in the pig (23), rat (26), mouse (27,28), rabbit (29), dog (30) and human (31,32). Interestingly, full-length BSEP cDNAs has been identified in a variant of the small skate (*Raja erinacea*), a 200-million-year-old marine vertebrate with an amino acid sequence identity of 68.5% to the human orthologue (33). Here, bile salts are transported in large amounts by BSEP. Furthermore, it was demonstrated that mutations leading to PFIC2 in humans had the same effect on substrate transport in skate BSEP (33,34). The bile of this elasmobranch normally consists of bile alcohols (scymnol sulfate) rather than bile salts, which cannot be found in its bile. Probably bile alcohols were the original

substrates for BSEP, and mammalian evolution led to different substrates owing to a selective pressure – for example, more fat in the diet. Interestingly, skate bile does not contain any phospholipids and no MDR3 protein (a phospholipid floppase) is found in the hepatocyte. The function of BSEP and its tight correlation with MDR3 is described in section 2.3. Importantly, however, this finding suggests that BSEP evolved much earlier than the highly identical MDR3 protein (sequence identity between P-gp and MDR3 of ~80%) and probably also by gene duplication (35). The occurrence of lipids in bile was potentially the result of the more deleterious bile acids than of the bile alcohols that arose during evolution. All of these indications demonstrate that BSEP diverged very early from P-gp and that it is highly conserved in vertebrate evolution.

2.2.4 The bile salt export pump – a member of the ABC transporter family

BSEP belongs to the group of ABC transporters. They can be found in all the taxonomic kingdoms (from bacteria to humans), and all possess the same modular architecture and act either as importers or exporters (36). In humans, 48 ABC transporter genes have been identified in addition to a small number of pseudogenes, which are not expressed (37). All known eukaryotic ABC transporters are exporters, whereas ABC importers can be found only in Archaea and Bacteria. In humans, ABC transporters are expressed throughout the body, but some highly tissue-specific and ABC transporters are restricted to the liver (37). Phylogenetic analysis of the entire human ABC transporter sequences has led to the classification of seven subfamilies (A to G) (38). Because of their important roles in human physiology, dysfunction is the cause of very severe diseases, such as cystic fibrosis (39). In terms of mutations of liver ABC transporters BSEP and MDR3, for example, PFIC2 (25) and PFIC3 (40) may develop (see chapter 9). BSEP belongs to the group B (MDR/TAP) subfamily of human ABC transporters because of its high sequence identity to P-gp. The gene is located on chromosomes 2q24 (25) and the 28 exons code for a 1321 amino acid glycosylated ABC transporter with a molecular mass of ~160 kDa (31,32).

ABC transporters have a core architecture consisting of two NBDs and two TMDs. In eukaryotes these modules are encoded on a single gene, but one must distinguish between the full-size transporters (two TMDs and two NBDs) and half-size transporters (only one of each domain). The latter homo- or heterodimerize to form a functional transporter. BSEP is a full-size ABC transporter with a core molecular weight of 146 kDa. Interestingly, the N-terminal NBD (NBD1) of BSEP contains a methionine instead of a glutamate within the Walker B motif. The glutamate normally interacts with ATP through a catalytic water molecule that catalyzes the nucleophilic attack onto the γ–phosphate. ATP binding sites in ABC transporters are composed of the Walker A and B motifs of one NBD and the C-loop of the other NBD. Therefore the ATP-binding site that contains NBD1 (site 1) is a degenerated site. Degeneration of this conserved residue is also seen in other human ABC transporters like TAP1/2 or CFTR (41). Functional studies of other degenerated ABC transporters have demonstrated that this mutation leads to an ATP-deficient site within the NBD dimer. This, of course, implies an asymmetric function of the two NBDs and further suggests that ATP binding site 1 of BSEP is catalytically inactive or active only at drastically reduced levels compared with the other ATP binding site in the composite dimer. This phenomenon has not been investigated for BSEP so far but is of high concern for a molecular understanding of ATP hydrolysis coupled to bile salt transport. The TMDs are located within the membrane and provide the translocation
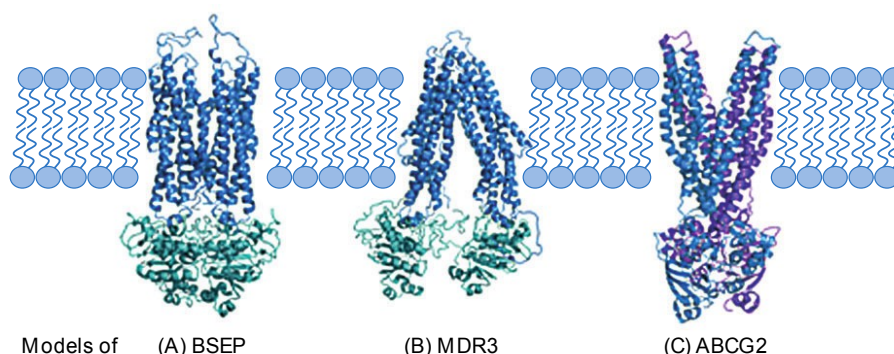
Models of   (A) BSEP   (B) MDR3   (C) ABCG2

Fig. 2.2: Models of BSEP, MDR3 and ABCG2. (**A**) The model of BSEP based on the known structure of Sav1866 from *Staphylococcus aureus*. The transmembrane domain is highlighted in blue and the nucleotide binding domain in cyan. The used template is deposited under protein data bank (PDB) code 2HYD (**B**) The model of MDR3 based on the known structure of P-gp from *Mus musculus*. The transmembrane domain is highlighted in blue and the nucleotide binding domain in cyan. The used template is deposited under PDB code 3G61. (**C**) The model of ABCG2 based on the known Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP. The used template is deposited under PDB code 2ONJ. Since ABCG2 is a halfsize transporter the two monomers are colour coded differently. Monomer I is highlighted in blue and monomer II in purple. Due to the bound AMP-PNP in the template structure the conformation of ABCG2 represents the potential nucleotide bound state. It is important to clarify that these models are based on the known X-ray structure and the structures obtained from the actual protein might look differently.

pathway for the substrate. In contrast to the NBDs, the TMDs are highly variable in their sequence and thus determine the substrate specificity. It is assumed that many human ABC transporters show the 6×6 topology, meaning that they contain six TM helices (TMH) traversing the membrane followed by a cytoplasmically located NBD and again six TMHs and an NBD. This assumption was originally proposed based on cysteine scanning mutagenesis of P-gp and the recently developed x-ray structures of mouse P-gp and bacterial homologues (42); it may be true for BSEP as well. BSEP is also a full-size ABC transporter containing 12 TMHs and two NBDs (▶Fig. 2.2). This number of helices is derived from hydrophobicity calculations because structural information on BSEP is lacking. To date only one eukaryotic ABC transporter structure (of mouse P-gp) has been published (42). The second available structure (ABCB10) is deposited only in the Protein Data Bank (PDB database). The P-gp structure shows the typical bundle of six helices crossing the membrane. However, as first observed for Sav1866, a domain swap is present in P-gp, suggesting that such a swapping is a conserved feature of ABC drug pumps. Here, four helices of one bundle and two helices of the other bundle build up one TMD. The TMDs provide a large cavity for substrate binding for mouse P-gp, which may also be true for BSEP. According to the "cholesterol fill-in mechanism," cholesterol also participates in substrate recognition and fills the volume of the cavity that is not occupied by the substrate, as postulated for P-gp. Besides the amino acids, which constitute the substrate-binding site, this could also e an additional explanation for the fact that those two closely homologous transporters have different substrate spectra and

BSEP is restricted to bile salts. It has been demonstrated that the activity of BSEP critically depends on cholesterol. This might be because BSEP is targeted to detergent-resistant microdomains (DRMs) in the canalicular membrane, which exhibit a high amount of cholesterol and sphingomyelin, or because of the "cholesterol fill-in model." Whatever the molecular reason(s) for the differences between P-gp and BSEP may be, all the models proposed must be verified experimentally for BSEP in the future. So far most if not all functional information on BSEP results from disease-linked mutations found in patients with, for example, PFIC2 (see chapter 9) and offers insights into the way single amino acids influence the trafficking, stability, and transport capabilities of BSEP.

2.2.5 Cloning and expression systems for BSEP

To study a protein biochemically *in vitro,* it is often necessary to obtain sufficient amounts of pure, homogeneous protein. Therefore an expression system must be chosen and recombinant expression constructs must be cloned. All this is true for BSEP, but it has one big drawback. The human cDNA that codes for BSEP has been found to be unstable in *Escherichia coli* (31,32,43). This phenomenon has been observed for several other mammalian membrane proteins as well (44). Because *E coli* is the most widely used cloning and expression host for standard molecular biology techniques, other strategies must be applied in utilizing the favored expression system. In the case of human BSEP, after several years of struggle, efforts to clone the cDNA into an expression vector were eventually successful. However this led, even after a bacterial promoter in the cDNA was silenced, to a construct with several point mutations within the coding sequence (six missense mutations) and the loss of specific parts of the coding sequence during expression construct propagation (32). All these findings led to the notion that the cDNA of BSEP is "toxic" or "unstable" for cloning and/or the expression host since colonies would no longer grow. One way of circumventing this laborious work is to use homologous recombination (HR) in the yeast *Saccharomyces cerevisiae* (45,46). We therefore established a workflow for human BSEP that can also be applied to any other target (47). Here the expression vector of interest was modified by the introduction of an origin of replication (ori) and a selection marker for *S cerevisiae* into the backbone of the plasmid. The linearized expression vector and the PCR-amplified BSEP cDNA, which has overlapping ends to the expression vector, are then transformed into yeast. *S cerevisiae* is capable of recombining those overlapping ends to a circular vector, and only clones that do this correctly are able to grow under selection. The expression vector can be recovered from yeast and transformed in *E coli* for amplification. We have found that *E coli* is capable of handling the BSEP cDNA if it is in a closed, circular plasmid form (no nicks, etc.) and grown strictly at or below 30°C. If required, our expression construct can be designed to remove the origin of replication or selection marker to prevent a potential influence of these additional sequences on balanced expression systems. Furthermore, we developed a mutagenesis strategy relying only on yeast. A changed primer design (primers carrying the mutation are not completely complementary to each other) results in a PCR product with overlapping ends (5' and 3' ends) that can be recombined by yeast, resulting in a plasmid containing the desired mutation. With the directed recombination–assisted mutagenesis (DREAM) method, mutations can be introduced more easily and quickly than with commercially available strategies. Therefore this method is seen as a DREAM (47).

30 | 2 Structure and function of hepatic ABC transporters

The major bottleneck in studying membrane proteins in vitro (e.g. structural and functional studies), is their homo– or heterologous overexpression, making it difficult to purify the protein in adequate amounts. Therefore one must choose between prokaryotic and eukaryotic expression systems. As a prokaryotic expression system, *E coli* is the most widely used host (48), although there are others like *Lactococcus lactis*, which is also successfully used for the overexpression of membrane proteins (49). Mammalian transporter can be expressed in *E coli*, but sometimes in an inactive manner. Therefore we also tried to overexpress human BSEP in *E coli* using the T7-RNA polymerase/promotor system in combination with a synthetic gene, which sequence was optimized for use in *E coli*. Unfortunately cells stopped growing upon induction of BSEP expression and we were not able to detect BSEP in cell lysates via Western blotting. BSEP is a plasma membrane protein and *E coli* does not possess the eukaryotic posttranslational modification system, and no cholesterol is present in the inner membrane. Therefore, eukaryotic expression systems are likely the methods of choice for BSEP. Three different expression systems, which are also commercially available, are used the most: yeast, insect, and mammalian cell lines. Mammalian cell lines have the great advantage that they present the native environment of BSEP; these cells contain the native lipid environment, the native secretory/posttranslational pathways, and a known functional expression. Human BSEP could be expressed in different mammalian cell lines (e.g. HEK293 cells (50), HepG2 cells (51), MDCK cells (52) and LLC PK1 (53) cells) and characterized functionally without purification. The most widely used system for the heterologous expression of BSEP is the insect cell system. Human BSEP was expressed in *Sf*9 (32,54) as well as HighFive cells (31). Insect cells exhibit a nonnative lipid environment with low levels of cholesterol as well as nonnative glycosylation (generally of the high-mannose type), but they resemble the native conditions more than yeast does. Expression in this system is used to investigate the transport properties of BSEP in vesicular-based transport assays. Cell culture−based systems may generally be suitable for addressing questions of a cellular phenotype, protein trafficking, and the modification of protein interactions. But mammalian and insect cell systems are also costly and maybe not be producible in large the amounts required for purification and structural studies of BSEP. From this point of view, a better choice might be yeast. Two yeast-based systems are used to overexpress mammalian membrane proteins: the previously mentioned *S. cerevisiae* and *Pichia pastoris*. The advantages are obvious: yeast is inexpensive, requires simple culture media, and exhibits well-studied genetics. Furthermore, yeasts have the eukaryotic modification machinery. Of course there are disadvantages, since, for example, the lipid composition of the membrane is different from that of mammalian cells (they contain ergosterol instead of cholesterol) and *S. cerevisiae* often hyperglycosylates proteins (highly branched and extended high-mannose structures), which is not observed in *P. pastoris*. It has been shown that BSEP requires glycosylation for transport activity in MDCK cells (55), although the type of glycosylation seems not to be important, as shown by functional expression in insect cells. *S. cerevisiae* was also used to express human P-gp (56) and MRP1 (57), and we were also able to express BSEP in this host (unpublished data), yielding only low amounts of fully translated protein. Therefore we switched to *P. pastoris*. This yeast was shown to overexpress 25 human ABC transporters (BSEP was not among them) and was also the expression host for mouse P-gp, which in the end and after a long endeavor resulted in the three dimensional x-ray structure (42,58). Recently we demonstrated the heterologous overexpression of human BSEP in

this yeast (47). Another advantage of *P. pastoris* is that this methylotrophic yeast strain can be fermented to high cell densities generating large amounts of biomass, which can be used for subsequent purification.

2.2.6 *In vitro* assays to study BSEP

Although BSEP has not been purified to homogeneity yet, assays have been described to study the function of BSEP *in vivo*. A vesicular transport assay is the most important one. It consists of three steps: (a) preparation of membrane vesicles, (b) addition of substrate and an energy source, and (c) readout of substrate uptake into the vesicles. In general there are two ways to prepare membrane vesicles from cells, right-side-out (RSO) and inside-out (IO) vesicles. In RSO vesicles, the cytosolic side of the transporter is localized in the lumen of the vesicles, whereas in IO vesicles it is vice versa. The latter ones are commonly used for primary transporters such as BSEP. With the addition of ATP and substrate, transport is initiated and the substrate begins to accumulate in the lumen of the vesicles. After a defined amount of time, the reaction is stopped by, for example, a rapid-filtration method; then the transported amount of substrate, which is retained in the vesicle on the filter, is quantified (via radioactivity, fluorescence, or LC/MS). With this assay, the substrate spectrum of BSEP was elucidated (see ▶ Tab. 2.1 for human BSEP). These assays were mainly performed with BSEP derived from insect cell vesicles (31,32) but also with vesicles originated from HEK293 cells or isolated canalicular membranes (52). Human BSEP transports monovalent conjugated bile salts in the order of taurochenodeoxycholate > taurocholate > tauroursodeoxycholate > glycocholate (it has to be

Tab. 2.1: Substrate spectra and Michaelis-Menten constant for human BSEP from different expression systems. For an excellent overview, see reference 60.

Substrate	K_M / µM	Source
Taurocholate	8 (32), 20 (30, 61), 15(54)	Sf9
	4 (31)	HighFive
	6 (50)	
Taurochenodeoxycholate	4 (54), 5 (32), 13 (61)	Sf9
	7 (50)	HEK293
Tauroursodeoxycholate	12 (32)	Sf9
Taurodeoxycholate	34 (61)	Sf9
Taurolithocholate	4 (61)	Sf9
Taurolithocholate 3 sulfate	10 (50)	HEK293
Glycocholate	11 (32), 36 (54)	Sf9
	22 (50)	HEK293
Glycochenodeoxycholate	2 (54)	Sf9
	8 (50)	HEK293
Pravastatin	124 (62)	HEK293

mentioned, that the K_M values vary slightly between different expression systems but not unconjugated ones) (see ▶ Tab. 2.1). In most of these studies bile acids are tritium-labeled for readout, but fluorescent bile acid derivatives, such as cholylglycylamidofluorescein and chenodeoxycholylglycylamidofluorescein, have also been investigated (59). In addition, inhibitors were analyzed for their potential impact on BSEP. Inhibition of BSEP by different drugs causes drug-induced cholestasis, leading to severe liver injury (63). Examples of inhibitors that were determined for human BSEP by a vesicular uptake assay in competition experiments with bile salts include cyclosporine, rifampicin, and bosentan. These assays and the recommendation of the European Medicines Agency (64) emphasize the importance of BSEP for drug development. With a vesicular uptake assay for BSEP commercially available, the screening of drug libraries is in principle straightforward. One disadvantage of this kind of assay, however, is that besides the target transporter, the vesicles contain many irrelevant membrane proteins that probably affect the uptake assay. This can be excluded if proper controls are performed, but it complicates the assay.

Mutations in the *BSEP* gene can lead to an impairment of bile salt transport due to the protein's dysfunction. This can lead to PFIC2 or BRIC2, a severe liver disease, which at present can be cured only by liver transplantation (65). Currently, according to the Human Gene Mutation Database (http://www.hgmd.org/), 179 disease-related BSEP mutations are known. Thus an understanding of the effect of such mutations could, in the future, lead to therapeutic innovations that might cure this disease without transplantation. If, for example, mutated BSEP is still able to transport and the disease is caused by a trafficking defect, it could also be investigated by the vesicular transport assay, with mutation and localization studies in cell culture systems involving immunostaining or with a fluorescent tag like eGFP or YFP.

Trafficking and the regulation of BSEP in the apical membrane of hepatocytes also requires adaptor proteins. HCLS1-associated protein X-1 (Hax1), for example, was identified using yeast two-hybrid screens as well as pull-down assays with glutathione-S-transferase (GST) tag fusion proteins (soluble parts of BSEP with GST tag) and co-immunoprecipitation (66). Other adaptor proteins are still not known and would be of high interest, especially for the short-term regulation of BSEP or for their potential involvement in trafficking mutants.

2.3 Structure and function of the multidrug resistance protein 3 (ABCB4; MDR3)

As described in section 2.2.1, BSEP is essential for the circulation of bile salts. However, bile salts are harsh detergents and possess the power to solubilize any biological membrane. The outer leaflet of the canalicular membrane is destabilized by bile salts, which are translocated in the canaliculus by BSEP (ABCB11). To dampen this effect, bile salts and phosphatidylcholine (PC) form mixed micelles with cholesterol translocated by ABCG5/G8. These mixed micelles have a lower capacity to extract lipids from the membrane. A second function of PC is the solubilization of cholesterol, which prevents the crystallization of cholesterol in the biliary duct and the formation of cholesterol gallstones. The bulk of PC is reabsorbed in the intestine and returns to the hepatocyte within the enterohepatic cycle. However, the half-time of PC to flip spontaneously from

the inner to the outer leaflet of a lipid bilayer is very low; therefore PC must be translocated across the membrane of the hepatocyte by an active transporter. The multidrug resistance protein 3 (MDR3), also called ABCB4, is localized only in the canalicular membrane of the hepatocyte (▶Fig. 2.1) and is indispensable for the primary active transport of PC from the inner to the outer leaflet of the canalicular membrane against a concentration gradient. The mouse homologue is called Mdr2 and fulfills the same function as MDR3 to flop PC across the apical membrane of hepatocytes. Mutations in the MDR3 gene caused different types of liver diseases, such as progressive familial intrahepatic cholestasis type 3 (PFIC3), intrahepatic cholestasis of pregnancy (ICP), and low phospholipid−associated cholestasis (LPAC).

2.3.1 A brief history of MDR3

During an analysis of cDNAs from human liver in 1987, van der Bliek et al. identified a gene that is highly homologous to the human P–gp and designated it *MDR3*. One year later the complete cDNA sequence was published (67). This sequence is composed of two similar halves. One half consists like BSEP of six putative TMHs and one NBD. The NBDs are identical to those of the human MDR1. Furthermore, the TMDs showed up to 80% identity. Divergence between MDR1 and MDR3 is greatest at the N-terminus and in the 60−amino acid linker connecting the two halves (67).

While MDR1 transports a wide variety of structural unrelated substances and is involved in multidrug resistance (MDR), no drug-pumping activity has been demonstrated for MDR3 (68). Smit and coworkers characterized mice with a disruption of *mdr2* in 1993. They ascertained that the homozygous disruption of the murine homologous *mdr2* gene leads to a complete absence of PC and cholesterol from bile (69). Furthermore, mice heterozygous for *Mdr2* (Mdr2$^{−/+}$) have normal amounts of cholesterol and only 40% of PC in bile. Human MDR3 can functionally replace *mdr2* in knockout mice (70). This demonstrates that the closely related Mdr2 and MDR3 carry out the same function. Direct evidence that MDR3 can translocate endogenous PC has been obtained in enhanced transport of newly synthesized [^{3}H]choline-labeled PC to the surface of transgenic fibroblast (71). This suggested that MDR3 translocates specifically PC from the inner to the outer leaflet of the canicular membrane.

Van Helvoort and coworkers (72) were the first to demonstrated specific transport of a short-chain PC in polarized pig kidney epithelial cells transfected with MDR3. In this study they measured lipid translocation across the plasma membrane by extracting fluorescently labeled short-chain lipids from the cell surface into the basolateral and apical media. MDR3 translocated fluorescently labeled PC but not the other lipid analogues (72). However van Helvoort et al. showed that radiolabeled short-chain PC lacking the fluorescence moiety was not translocated into the apical medium by MDR3.

2.3.2 MDR3 – an ATP-binding cassette (ABC) transporter

The 141-kDa lipid translocase is postranslationally modified by glycosylation at two predicted asparagine residues (N91 and N97) and is allocated to the group of P glycoproteins based on amino acid sequence homology. Like BSEP, MDR3 is a so called full-size transporter and is encoded on one structural gene (NBD-TMD)$_2$ (see ▶Fig. 2.2).

2.3.2.1 Transport machinery – the flippase model

Two models, the "vacuum cleaner model" and the "flippase model", are postulated for the transport of hydrophobic substrates by ABC transporters. The vacuum cleaner model proposes that the molecule in the cytosol interacts with the transporter, enters a hydrophobic cavity of the ABC-transporter, and is pumped into the extracellular space.

In contrast, Higgins and Gottesman proposed a flippase model for mammalian P-glycoproteins (73). P-gp binds an amphipathic molecule located in the inner leaflet of the plasma membrane and flips the molecule to the exoplasmic leaflet. Therefore a substrate-binding site must be accessible from the lipid phase. The substrate accumulates in the outer leaflet, forming a concentration gradient between the cytosolic and exoplasmic leaflet of the plasma membrane. From the leaflet the substrate can freely diffuse into the extracellular medium. On the basis of the flippase model, it is feasible to explain the observation that PC secretion depends on the expression of Mdr2, the mouse homolog of MDR3, and the bile salt concentration (74). Elferink and coworkers showed if either PC or bile salts were lacking, PC would not be detectable in bile, concluding that bile salts translocation is the main driving force for the secretion of phospholipids (75).

It is assumed that P-gp, which is over 76% identical to MDR3, can bind substrates within the inner leaflet of the membrane as well as from the cytosol. How ABC transporters recognize and translocate substrates is still unclear and the subject of intensive investigation.

2.3.2.2 MDR3 – a drug ABC transporter?

MDR3 shares 78% amino acid sequence identity with the well-characterized drug-pumping ABC transporter P-gp. Because of the high amino acid sequence homology between MDR3 and MDR1 (over 85%) it was assumed that MDR3 also translocates drugs. However, initial experiments with MDR3 cDNA or its mouse homolog Mdr2 transfected cells showed no drug resistance (67,76–78) and MDR3 was not detected in MDR cell lines (67,79). The first indication that MDR3 translocates drugs was obtained by Kino et al. (80). They observed that MDR3 transfected yeast cells showed low-level resistance against the antifungal agent aureobasidin A. Another study of MDR3 was performed by Smith et al. (81), who investigated vectorial substrate transport by polarized pig kidney monolayers transfected with MDR3 cDNA of several MDR1 substrates. They observed that the transport of digoxin, paclitaxel, vinblastine, and ivermectine into the apical medium was significantly increased in the MDR3-transfected cells compared with the control cells. Digoxin transport by MDR3 was efficiently inhibited by the MDR1-specific inhibitor verapamil, cyclosporine, and PSC833, which also inhibited the transport of short-chain PC. Verapamil had also previously been shown to inhibit the translocation of short-chain C_6-NBD-PC (72,82). No significant transport of some other MDR1 substrates, such as cyclosporine or dexamethasone, was determined.

These results suggest that MDR3 is not specific for PC and is able to translocate various typical MDR1 substrates as well. But why is drug transport observed only in polarized monolayers transfected with MDR3 cDNA? Currently there is no satisfactory explanation. Further studies on the translocation of long-chain PC and drugs by MDR3 are required.

2.3.3 Analysis of the substrate specificity of the PC translocator

The analysis of lipid transporters is very complex by reason of the difficulty of developing a reliable assay for the molecular mechanism of lipid transporters. Following are described two different ways of analyzing the function of P-glycoproteins and especially lipid translocases. On the one hand, MDR3 translocates PC across the membrane; three different approaches to this have been reported. On the other hand, MDR3 hydrolyzes ATP. The resultant ATPase activity correlates indirectly with the substrate transport.

2.3.3.1 Transport of lipids and lipid analogues by the ABC transporter

Currently no sensitive assay for measuring naturally occurring long-chain protein-mediated lipid translocation from one leaflet to the other leaflet of the membrane exists. Nevertheless Sleight and Pagano used the lower hydrophobicity of short-chain lipids (C_5-C_6 acyl chain) to determine lipid transport of lipid translocases, which allows their free exchange as monomers via the aqueous phase (83). At first short-chain lipids are easily integrated into the surface of the membrane of interest and can be detected by a spin-, fluorescent-, or radiolabel on the short-chain. The transport can be measured by chemically quenching of the spin-labeled or fluorescent analogue in the outer leaflet (84) or by "back-exchange." To date two different systems to determine short-chain PC transport by MDR3 or the mouse-homologous Mdr2 have been described.

Ruetz and Gros expressed Mdr2 in the membrane of secretory vesicles obtained from a yeast secretion mutant (82). These vesicles can be easily isolated and consist of a pure population of inside-out vesicles, meaning that the cytoplasmic NBDs of the ABC transporter are located on the outside of the vesicle. To determine Mdr2-driven transport of PC from the outer leaflet into the inner leaflet of the vesicular membrane they used fluorescent-labeled short-chain PC – C_6-NBD-PC: (N-6[7-nitro-2,1,3-benzoxadiazol-4-yl]-amino-hexanoyl-phosphatidylcholine) – which is chemically reduced to the non-fluorescent compound by a membrane-impermeable reducing agent such as sodium dithionite. Dithionite reduces only the C_6-NBD-PC located in the outer leaflet, whereas the translocated C_6-NBD-PC in the inner leaflet remains unaffected. Detergent disruption of the vesicles lead to a decrease of fluorescence emission because of the release of translocated C_6-NBD-PC. With this system Ruetz and Gros proved indeed a very small but specific transport of short-chain PC analogue by Mdr2. Second, they showed that transport was ATP-dependent and inhibited by verapamil, a specific inhibitor for MDR1.

In the "back-exchange" method, short-chain lipids are extracted from the outer leaflet by bovine serum albumin (BSA). BSA has the ability to selectively bind short-chain lipids from the outer leaflet. The lipids are analyzed by two-dimensional thin layer chromatography (TLC) and the transport activity is calculated by the ratio between translocated and total amount of short-chain PC (85).

The floppase activity of MDR3 was confirmed by van Helvoort et al. using LLC-PK1 pig cells transfected with an MDR3 cDNA construct (72). LLC-PK1 cells are able to

2.3.3 Analysis of the substrate specificity of the PC translocator

The analysis of lipid transporters is very complex by reason of the difficulty of developing a reliable assay for the molecular mechanism of lipid transporters. Following are described two different ways of analyzing the function of P-glycoproteins and especially lipid translocases. On the one hand, MDR3 translocates PC across the membrane; three different approaches to this have been reported. On the other hand, MDR3 hydrolyzes ATP. The resultant ATPase activity correlates indirectly with the substrate transport.

2.3.3.1 Transport of lipids and lipid analogues by the ABC transporter

Currently no sensitive assay for measuring naturally occurring long-chain protein-mediated lipid translocation from one leaflet to the other leaflet of the membrane exists. Nevertheless Sleight and Pagano used the lower hydrophobicity of short-chain lipids (C_5-C_6 acyl chain) to determine lipid transport of lipid translocases, which allows their free exchange as monomers via the aqueous phase (83). At first short-chain lipids are easily integrated into the surface of the membrane of interest and can be detected by a spin-, fluorescent-, or radiolabel on the short-chain. The transport can be measured by chemically quenching of the spin-labeled or fluorescent analogue in the outer leaflet (84) or by "back-exchange." To date two different systems to determine short-chain PC transport by MDR3 or the mouse-homologous Mdr2 have been described.

Ruetz and Gros expressed Mdr2 in the membrane of secretory vesicles obtained from a yeast secretion mutant (82). These vesicles can be easily isolated and consist of a pure population of inside-out vesicles, meaning that the cytoplasmic NBDs of the ABC transporter are located on the outside of the vesicle. To determine Mdr2-driven transport of PC from the outer leaflet into the inner leaflet of the vesicular membrane they used fluorescent-labeled short-chain PC – C_6-NBD-PC: (N-6[7-nitro-2,1,3-benzoxadiazol-4-yl]-amino-hexanoyl-phosphatidylcholine) – which is chemically reduced to the non-fluorescent compound by a membrane-impermeable reducing agent such as sodium dithionite. Dithionite reduces only the C_6-NBD-PC located in the outer leaflet, whereas the translocated C_6-NBD-PC in the inner leaflet remains unaffected. Detergent disruption of the vesicles lead to a decrease of fluorescence emission because of the release of translocated C_6-NBD-PC. With this system Ruetz and Gros proved indeed a very small but specific transport of short-chain PC analogue by Mdr2. Second, they showed that transport was ATP-dependent and inhibited by verapamil, a specific inhibitor for MDR1.

In the "back-exchange" method, short-chain lipids are extracted from the outer leaflet by bovine serum albumin (BSA). BSA has the ability to selectively bind short-chain lipids from the outer leaflet. The lipids are analyzed by two-dimensional thin layer chromatography (TLC) and the transport activity is calculated by the ratio between translocated and total amount of short-chain PC (85).

The floppase activity of MDR3 was confirmed by van Helvoort et al. using LLC-PK1 pig cells transfected with an MDR3 cDNA construct (72). LLC-PK1 cells are able to grow as monolayers on filters and MDR3 is found only in the apical membrane. Cells are cultured in the presence of a short-chain lipid precursor, which is taken up and converted into the corresponding short-chain lipid analogue. The intracellularly synthesized C_6-NBD-PC was specifically transported by MDR3 but not C_6-NBD-phosphatidylethanolamine, C_6-NBD-sphingomyelin, or C_6-NBD-glucosylceramide. Remarkably, radiolabeled short-chain PC with two C_8 fatty acids lacking the fluorescence moiety

(C_8C_8-[^{3}H]PC) were slightly translocated into the apical medium. Van Helvoort and colleagues confirmed that the high specificity of MDR3 is determined by the choline head group.

To date, only Smith and coworkers have been able to generate a system for the translocation of long-chain PC through the membrane of fibroblasts from transgenic mice by MDR3 (71). Intracellular synthesized radioactively labeled PC is inserted into the inner leaflet and translocated to the outer leaflet in the presence of MDR3. PC-TP, a PC-specific transfer protein, carries out the exchange of labeled PC from the outer leaflet to acceptor liposomes in the medium. In this study Smith et al. determined an increased translocation of long-chain PC in the presence of MDR3. One main drawback of this system is the high background in the absence of MDR3 by vesicular transport. This makes usage of this assay extremely complicated.

Thus far no in vitro system for the translocation of PC by MDR3 is established because of the challenge of cloning, expressing, and purifying functional MDR3 in sufficient amounts and the technical difficulty of measuring the translocation of natural PC.

2.3.3.2 Substrate-stimulated ATPase activity

ABC transporters hydrolyze ATP to energize the transport across the membrane. Since ATP hydrolysis is linked by substrate translocation, the transport activity can be visualized indirectly. Most ABC transporters offer a basal ATPase activity. This ATPase activity is stimulated or inhibited by adding the substrate or inhibitor. There are two assays for measuring the ATPase activity by the determination of released inorganic phosphate: the malachite green assay (86) and the NADH-coupled assay (87,88). Both assays measure the release of free orthophosphate.

The highly sensitive malachite green assay is based on the complex formation of free phosphate with molybdate. The reaction of phosphomolybdate and the dye malachite green results in a green complex, whose absorbance can be easily determined at a wavelength of 620 to 650 nm. Nevertheless, a disadvantage of this method is its inability to observe the hydrolytic reaction continuously. The NADH-coupled assay enables one to follow the rate of ATP hydrolysis in real time by coupling the release of P_i and the oxidation of NADH to NAD$^+$. The ATPase hydrolyzes ATP to ADP and P_i. ADP is converted to ATP and phosphoenolpyruvate (PEP) to pyruvate by pyruvate kinase. The lactate dehydrogenase reduces pyruvate to lactate, while NADH is oxidized to NAD$^+$. The decrease of NADH is then determined at a wavelength of 340 nm.

The precondition to measuring ATPase activity is simple: sufficient expression of MDR3. To date it has not been possible to clone and express functional MDR3 in bacterial systems such as *E coli* or *L lactis* because of the "toxic" or "unstable" DNA sequence (as described in section 2.2.5). The expression of MDR3 in mammalian cell lines such as LLC PK1 and insect cells has been demonstrated by different groups (72,82). However, the obtained protein amounts are not sufficient to purify MDR3. Thus, up to now, it has not been possible to measure the PC-stimulated ATPase activity of membrane vesicles containing MDR3 and/or of isolated MDR3 in detergent solution or reconstituted into liposomes. To overcome this major obstacle it is crucial to study MDR3 in vitro and obtain a more detailed knowledge of this interesting ABC transporter as expressed inside the liver.

2.4 Structure and function of the breast cancer resistance protein (ABCG2; BCRP)

2.4.1 History of ABCG2

ABCG2 was first identified in human (BCRP, ABCG2) carcinoma cells. Despite the absence of overexpression of known multidrug transporters, like P-gp or MRP1, these cells displayed a remarkable resistance to multiple chemotherapeutic drugs such as doxorubicin and mitoxantrone. The gene conferring this resistance was isolated and subsequently used to transfect carcinoma cells, which then displayed a diminished accumulation of daunorubicin in flow cytometry assays. Additionally, this transport function appeared to depend on the presence of ATP, and this transport protein was termed breast cancer resistance protein (BCRP; ABCG2) (89). Independently, ABCG2 was discovered as the determinant responsible for the resistance of human colon carcinoma cells selected in mitoxantrone. Isolated cDNA clones displayed high levels of resistance to mitoxantrone. The gene showed relation to the *Drosphila melanogaster* white gene and homology to ABC transporters; it was named MXR for "mitoxantrone resistance" (90). Furthermore, ABCG2 was identified among a group of new human ABC transporters that were found to be highly expressed in the placenta. The isolated cDNA contained an open reading frame of 655 amino acids consisting an ABC half-size transporter with an N-terminal NBD and a C-terminal TMD (91). Although it was discovered three times in different contexts, the gene involved always encoded ABCG2.

2.4.2 Structure and function of ABCG2

ABCG2 is a 72-kDa 655−amino acid glycoprotein. Among the members of the ABC transporter family, ABCG2 has, like other members of the ABCG subfamily, a reverse topology, meaning that the NBD is located N-terminal to the TMD. With only one NBD and one TMD encoded on a single gene, ABCG2 is considered to be a half-size transporter and thought to dimerize to become a functional ABC transporter (see ▶Fig. 2.2). Several studies have focused on this oligomerization behavior. Interestingly, intermolecular disulfide bonds are required to obtain a dimeric protein. Cysteine scanning mutagenesis revealed that residue C603 of ABCG2 is involved in intermolecular crosslinking via disulfide bonds (92). Additionally, no mutation of any other cysteine residue had an effect on the dimerization of ABCG2 or its activity. In agreement with these results, Henriksen et al. showed that the oligomeric species of ABCG2 was observed with the use of a nonreducing SDS-PAGE can be gradually disrupted by the addition of a reducing agent (93). Here, mutational analysis of the three cysteine residues located in the third extracellular loop showed that only the C603A mutant impaired dimerization. However, a cell survival assay with mitoxantrone showed that this mutation was still as resistant as the wild type, indicating that the disulfide bond is not essential for the transport function. A biotinylation assay supported the idea that the other two cysteine residues in this loop, C592 and C608, form an intramolecular disulfide bond. However, this disulfide bond is important for protein degradation (94).

An important feature of ABCG2 is the GXXXG motif, which has been identified as a recurring transmembrane sequence and is proposed to be an interaction site between the transmembrane α-helices of different monomers. Polgar et al. investigated the only putative GXXXG motif in transmembrane helix 1 of ABCG2. Mutation of one or both of

the glycine residues resulted in lower ATP hydrolysis and a reduced substrate transport rate, although the protein was still expressed at similar levels on the cell surface. These findings support the hypothesis that the GXXXG motif plays a role for the correct orientation of the transmembrane segments toward each other in the functional transporter. Mutational studies of G553 indicate an involvement of this residue in the dimerization of ABCG2 (95).

Another important characteristic of ABCG2 is its hyperglycosylation, deduced from the apparent molecular weight of the protein in SDS PAGE gels and susceptibility to PNGaseF treatment. The glycosylation, however, appears to have no influence on the trafficking of ABCG2 to the plasma membrane. Surface expression was investigated by immunostaining of human ovarian carcinoma cells and hamster ovary cells. Although three glycosylation sites are predicted to be potentially located in the third extracellular loop, only the N569Q mutant showed impaired glycosylation. As mentioned in section 2.2.5, this impaired glycosylation does not result in misstrafficking, in contrast to, for example, the N557 alanine mutation, which results in a ER localization of ABCG2 (96).

Nonglycosylated ABCG2 still showed reduced accumulation of the substrate rhodamine 123 in flow-cytometric assays and normal ATPase activity, which can be stimulated by prazosin. The results were comparable to levels found for glycosylated ABCG2 in crude membrane preparations, indicating that glycosylation is not essential for the function of ABCG2. Many studies investigating the function of ABCG2 have employed mutagenesis to clarify the role of different residues in the protein. Residues C592, C603, and C608 are involved in intra- or intermolecular disulfide bonds and N596 is glycosylated. Furthermore, residue R482 has been extensively characterized. Early isolates of ABCG2 from carcinoma cell lines showed a mutation at this position. By testing the accumulation of rhodamine 123 in cells expressing the variants R482G and R482T, broader substrate specificity was observed (97). Whereas the wild-type protein conferred no resistance to compounds like rhodamine 123, doxorubicin, or daunorubicin, expression of ABCG2 and the mutants R482G and R482T reduced the accumulation of the drugs and prolonged cell survival in cytotoxicity assays. Other compounds – like mitoxantrone, prazosin, and Hoechst 33342 – are substrates for both mutant and wild-type transporters (98,99) . A later study confirmed previous results and additionally observed binding of substrates, which are not transported to the wild-type transporter (100).

A common single-nucleotide polymorphism encoding the mutation Q141K is linked with the occurrence of gout. ABCG2 was shown to be located in the brush-border membrane of kidney proximal tubule cells. Functional assays with *X. laevis* oocytes expressing wild-type ABCG2 or Q141K mutant showed that the latter exhibited urate efflux, thereby linking ABCG2 to this genetic disease (101). A recent study revealed that this mutant is exhibiting increased susceptibility for lysosomal and proteasomal degradation (102).

2.4.3 Analysis of the substrate specificity of ABCG2

Owing to the discovery of ABCG2 in drug-resistant cells, the first reported substrates for it were predominantly chemotherapeutic drugs. These included mitoxantrone, flavopiridol, metothrexate, irinotecan and its active metabolite SN-38, porphyrines, and tyrosin-kinase inhibitors such as imatinib and gefitinib (103). Other substrates are antibiotics (104,105), flavonoids, antivirals (106,107), folic acid (108), and fluorescent dyes such

as Hoechst 33342. Mutation of the arginine residue at position 482 conveys a broader substrate spectrum including rhodamine 123 and anthracyclines such as doxorubicin. Because of its broad substrate spectrum and its expression in several tissues apart from the liver – such as the small intestine, colon, central nervous system, testis, ovary, and placental syncytiotrophoblasts – the transporter is thought to have a protective role (103,109,110). The number of ABCG2 inhibitors identified is equally large. Fumitremorgin C was the first inhibitor described (111). Its analog, Ko143, was found to be one of the most effective ABCG2 inhibitors (112). Some inhibitors were also inhibitors of P-gp or MRP – among them cyclosporine (113) and elacridar (GF120918)(114). Many compounds are both inhibitors and transported substrates, such as dihydropyridines (115). Despite the great number of substrates and inhibitors described to date, no clear structural requirements for a binding compound could be identified.

2.4.4 Expression, purification, and biochemical studies of ABCG2

To date, ABCG2 has been successfully expressed in a number of different vector systems and host organisms. Early studies have been done with drug-selected mammalian cell lines. Finally, the isolation of the cDNA offered the opportunity to move the expression to some heterologous hosts, such as *Xenopus* oocytes, insect cells, yeast, or bacteria.

Baculovirus-infected insect ovary cells (*Sf*9) and High Five cells offer an alternative to mammalian cell lines and have been successfully used to overexpress ABCG2, although in both cases hypoglycosylation, transport, and ATPase activity were observed (116,117). Other expression systems include yeasts like *P. pastoris* and *S. cerevisiae*. Mao et al. expressed ABCG2 in *P. pastoris*, obtaining active protein comprising about 3% of the total protein in microsome preparations (118). Similar expression levels could be observed in baker's yeast, yielding protein with ATPase activity, which could be stimulated by substrate (119). Additionally, a prokaryotic expression system has been reported employing the gram-positive bacterium *L. lactis* (120). Expression in another bacterial system, *E. coli,* did not yield functional protein (121).

Especially for the purification of ABCG2 from the membrane fraction of the expression host, a high yield is needed. Protein expression of the systems mentioned previously in this section was tested on the ability to obtain high yields of ABCG2 after purification. Solubilization of ABCG2 using different detergents showed the best results with the use of lysophosphatidylcholine (LPC) and n-dodecyl-β-D-maltoside (β-DDM) for *P. pastoris* membranes and FosCholine-14 and -16 for ABCG2 expressed in insect cells (122,123). Also used for solubilization of protein from insect cell membranes was CHAPS (117). Purification steps of the amino-terminal-histidine-tagged protein in all cases yielded sufficiently pure protein after immobilized metal-ion affinity chromatography (IMAC). Because of weak binding to the affinity resin, further purification steps were necessary when the insect cell expression system was used. These included ion exchange and size exclusion chromatography (117). ABCG2 retained ATPase activity and substrate binding after its purification.

Because of its ability to efflux a broad variety of substrates, multiple drug binding sites have been proposed for ABCG2. Clark et al. investigated this with heterologous displacement assays. [³H]daunomycin binding constants were measured in the presence of other known substrates of the ABCG2 gain-of-function mutant R482G. Three distinct binding sites were proposed, which are interlinked by allosteric communication (124). Several

studies of the substrate specificity and drug binding could be obtained by employing fluorescent substrates of ABCG2. For example 1,4-dihydropyridines could be identified as ABCG2 substrates by photoaffinity labeling with [^{125}I]Iodoarylazidoprazosin (IAAP) and [^{3}H]azidopine (125).

Since the translocation process is ATP-dependent, the ATPase activity of ABCG2 has been measured to confirm its physiologic activity. The majority of kinetic parameters were obtained on membrane preparations containing other ATPases. The wild type and the R482G isoform are capable of hydrolyzing ATP in the absence of any substrate (98). Since the influence of substrates and inhibitors on hydrolytic activity is an indicator of interaction with the protein, the measurement of ATPase activity is the focus of several studies. ATPase activity has been used as a readout in order to identify cholesterol content of the membrane as a major factor in ABCG2 activity (126). Cholesterol loading and depletion experiments showed stimulation of ATPase activity by substrates and improved drug transport in cholesterol-loaded membranes. In contrast, ATPase activity could not be stimulated in cholesterol-depleted membranes, indicating an essential role of membrane cholesterol.

Another tool to gain further knowledge about the topology of ABCG2 is epitope insertion mutagenesis. One study employed hemagglutinin (HA) tags to probe the predicted hydrophilic regions of ABCG2 via immunofluorescence (127). The results supported a model of six transmembrane helices with the amino and carboxy termini located intracellularly. A later study investigated a current homology model of ABCG2 by epitope insertion and found significant differences in the location of the predicted transmembrane segments (122).

To date there are no high-resolution structural data on ABCG2, although some attempts to obtain such data have been undertaken. These include negative-stain electron cryomicroscopy of purified protein. ABCG2 overexpressed in insect cells was solubilized and retained its stimulated ATPase activity. Analysis of the electron microscopy data revealed large particles (~170 Å in diameter) with a noticeable fourfold symmetry, in agreement with a higher oligomer as postulated by biochemical analysis. The final three-dimensional structure with an estimated resolution of ~18 Å could be accurately fitted with homology models of ABCG2, forming a tetramer. Data from size-exclusion chromatography and blue native PAGE supported the idea that ABCG2 forms a higher-order oligomeric species under the tested conditions (122).

Rosenberg et al. used purified ABCG2 expressed in *P. pastoris* to obtain two-dimensional crystals. The substrate mitoxantrone had a noticeable effect on the crystal shape. Analysis showed a significant change in unit cell dimensions, indicating a conformational change upon drug binding. A new homology model verified by epitope insertion mutagenesis supported the structural data by showing rigid body motion of two transmembrane helices, leading to a more compact conformation of the transporter in the drug-bound state. However, a three-dimensional structure is still a long way off. Eventually that will provide a detailed look at the function of ABCG2 at the molecular level.

2.5 Concluding remarks

Several interesting transporters are expressed within the hepatocytes, contributing to the function of these cells. This chapter focused on three of them, BSEP, MDR3, and

ABCG2. A wealth of information is to be derived from studies of their expression and mutation in different mammalian cell lines – studies focusing on their localization, trafficking, and activity. Although such studies have revealed extremely valuable and often essential information, the next step must be to achieve a molecular understanding of these transport mechanisms. Here, the first prerequisite is to elucidate the overexpression of these transporters, which will lead to their characterization directly in isolated membranes and/or after subsequent solubilization and purification in detergent solution. As described and summarized, the expression of membrane proteins is by no means trivial and often hampered by a too low expression in homo- or heterologous expression systems. However, if overexpression can be achieved for BSEP, MDR3, or ABCG2, the gain in knowledge derived from localization studies as well as mutational analysis will shed much light on the molecular mechanism of transport of a large variety of substrates with ATP only as an energizing molecule. Truly it will be a long way to go, but the information obtained will be worth the effort.

2.6 References

1. Schmitt L, R Tampe. Structure and mechanism of ABC transporters. Current opinion in structural biology 2002;12(6):754–60.
2. Zaitseva J, et al. A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB. Biochemical Society transactions 2005;33 (Pt 5):990–5.
3. Schmitt L, et al. Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. Journal of molecular biology 2003;330(2)333–42.
4. Jardetzky O. Simple allosteric model for membrane pumps. Nature 1966;211(5052):969–70.
5. Higgins CF, KJ Linton. The ATP switch model for ABC transporters. Nature structural & molecular biology 2004;11(10): 918–26.
6. Russell DW. Fifty years of advances in bile acid synthesis and metabolism. Journal of lipid research 2009:S120–5.
7. Hofmann AF. Bile acids: trying to understand their chemistry and biology with the hope of helping patients. Hepatology 2009;49(5):1403–18.
8. Hofmann AF, LR Hagey. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. Cellular and molecular life sciences: CMLS 2008;65(16):2461–83.
9. Kullak-Ublick GA, B Stieger, PJ Meier. Enterohepatic bile salt transporters in normal physiology and liver disease. Gastroenterology 2004;126(1):322–42.
10. Hofmann AF. Bile acids: the good, the bad, and the ugly. News in physiological sciences: an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society 1999;14:24–9.
11. Meier PJ, B Stieger. Bile salt transporters. Annual review of physiology 2002;64:635–61.
12. Kullak-Ublick GA, et al. Hepatic transport of bile salts. Seminars in liver disease 2000;20(3):273–92.
13. Hagenbuch B, P Dawson. The sodium bile salt cotransport family SLC10. Pflugers Archiv: European journal of physiology 2004;447(5):566–70.
14. Meier PJ, et al. Substrate specificity of sinusoidal bile acid and organic anion uptake systems in rat and human liver. Hepatology 1997;26(6):1667–77.
15. Agellon LB, EC Torchia. Intracellular transport of bile acids. Biochimica et biophysica acta 2000;1486(1):198–209.

42   2 Structure and function of hepatic ABC transporters

16. Arrese M. M Ananthanarayanan. The bile salt export pump: molecular properties, function and regulation. Pflugers Archiv: European journal of physiology 2004;449(2):123–31.
17. Stieger B, Y Meier, PJ Meier. The bile salt export pump. Pflugers Archiv: European journal of physiology 2007;453(5):611–20.
18. Meier PJ, AS Meier-Abt, JL Boyer. Properties of the canalicular bile acid transport system in rat liver. The Biochemical journal 1987;242(2):465–9.
19. Weinman SA, J Graf, JL Boyer. Voltage-driven, taurocholate-dependent secretion in isolated hepatocyte couplets. The American journal of physiology 1989;256(5 Pt 1):G826–32.
20. Adachi Y, et al. ATP-dependent taurocholate transport by rat liver canalicular membrane vesicles. Hepatology 1991;14(4 Pt 1):655–9.
21. Muller M, et al. ATP-dependent transport of taurocholate across the hepatocyte canalicular membrane mediated by a 110-kDa glycoprotein binding ATP and bile salt. The Journal of biological chemistry 1991;266(28):18920–6.
22. Brown RS, Jr., et al. Enhanced secretion of glycocholic acid in a specially adapted cell line is associated with overexpression of apparently novel ATP-binding cassette proteins. Proceedings of the National Academy of Sciences of the United States of America 1995;92(12):5421–5.
23. Childs S, et al. Identification of a sister gene to P-glycoprotein. Cancer research 1995;55(10):2029–34.
24. Gerloff T, et al. The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver. The Journal of biological chemistry 1998;273(16):10046–50.
25. Strautnieks SS, et al. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nature genetics 1998;20(3):233–8.
26. Green RM, F Hoda, KL Ward. Molecular cloning and characterization of the murine bile salt export pump. Gene 2000;241(1):117–23.
27. Lecureur V, et al. Cloning and expression of murine sister of P-glycoprotein reveals a more discriminating transporter than MDR1/P-glycoprotein. Molecular pharmacology 2000;57(1):24–35.
28. Noe J, et al. Characterization of the mouse bile salt export pump overexpressed in the baculovirus system. Hepatology 2001;33(5):1223–31.
29. Xu G, et al. Removal of the bile acid pool upregulates cholesterol 7alpha-hydroxylase by deactivating FXR in rabbits. Journal of lipid research 2002;43(1):45–50.
30. Yabuuchi H, et al. Cloning of the dog bile salt export pump (BSEP; ABCB11) and functional comparison with the human and rat proteins. Biopharmaceutics & drug disposition 2008;29(8):441–8.
31. Byrne JA, et al. The human bile salt export pump: characterization of substrate specificity and identification of inhibitors. Gastroenterology 2002;123(5):1649–58.
32. Noe J, B Stieger, PJ Meier. Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 2002;123(5):1659–66.
33. Ballatori N, et al. Bile salt excretion in skate liver is mediated by a functional analog of Bsep/Spgp, the bile salt export pump. American journal of physiology. Gastrointestinal and liver physiology 2000;278(1):G57–63.
34. Cai SY, et al. Bile salt export pump is highly conserved during vertebrate evolution and its expression is inhibited by PFIC type II mutations. American journal of physiology. Gastrointestinal and liver physiology 2001;281(2):G316–22.
35. Moitra K, M Dean. Evolution of ABC transporters by gene duplication and their role in human disease. Biological chemistry 2011;392(1–2):29–37.
36. Higgins CF. ABC transporters: from microorganisms to man. Annual review of cell biology 1992;8:67–113.
37. Dean M, A Rzhetsky, R Allikmets. The human ATP-binding cassette (ABC) transporter superfamily. Genome research 2001;11(7):1156–66.

38. Allikmets R, et al. Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. Human molecular genetics 1996;5(10):1649–55.

39. Gadsby DC, P Vergani, L Csanady. The ABC protein turned chloride channel whose failure causes cystic fibrosis. Nature 2006;440(7083):477–83.

40. Davit-Spraul A, et al. The spectrum of liver diseases related to ABCB4 gene mutations: pathophysiology and clinical aspects. Seminars in liver disease 2010;30(2):134–46.

41. Ernst R, et al. Engineering ATPase activity in the isolated ABC cassette of human TAP1. The Journal of biological chemistry 2006;281(37):27471–80.

42. Aller SG, et al. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 2009;323(5922):1718–22.

43. Byrne JA, et al. Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology 2009;49(2):553–67.

44. Vu K, et al. The functional expression of toxic genes: lessons learned from molecular cloning of CCH1, a high-affinity Ca2+ channel. Analytical biochemistry 2009;393(2):234–41.

45. Ma H, et al. Plasmid construction by homologous recombination in yeast. Gene 1987;58(2–3):201–16.

46. Oldenburg KR, et al. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. Nucleic acids research 1997;25(2):451–2.

47. Stindt J, et al. Heterologous overexpression and mutagenesis of the human bile salt export pump (ABCB11) using DREAM (Directed REcombination-Assisted Mutagenesis). PloS one 2011;6(5): e20562.

48. Schlegel S, et al. Revolutionizing membrane protein overexpression in bacteria. Microbial biotechnology 2010;3(4):403–11.

49. Kunji ER, et al. Eukaryotic membrane protein overproduction in Lactococcus lactis. Current opinion in biotechnology 2005;16(5):546–51.

50. Hayashi H, et al. Transport by vesicles of glycine- and taurine-conjugated bile salts and taurolithocholate 3-sulfate: a comparison of human BSEP with rat Bsep. Biochimica et biophysica acta 2005;1738(1–3):54–62.

51. Kubitz R, et al. Trafficking of the bile salt export pump from the Golgi to the canalicular membrane is regulated by the p38 MAP kinase. Gastroenterology 2004;126(2):541–53.

52. Hayashi H, et al. Two common PFIC2 mutations are associated with the impaired membrane trafficking of BSEP/ABCB11. Hepatology 2005;41(4):916–24.

53. Mita S, et al. Vectorial transport of unconjugated and conjugated bile salts by monolayers of LLC-PK1 cells doubly transfected with human NTCP and BSEP or with rat Ntcp and Bsep. American journal of physiology. Gastrointestinal and liver physiology 2006;290(3):G550–6.

54. Kis E, et al. Effect of membrane cholesterol on BSEP/Bsep activity: species specificity studies for substrates and inhibitors. Drug metabolism and disposition: the biological fate of chemicals 2009;37(9):1878–86.

55. Mochizuki K, et al. Two N-linked glycans are required to maintain the transport activity of the bile salt export pump (ABCB11) in MDCK II cells. American journal of physiology. Gastrointestinal and liver physiology 2007;292(3):G818–28.

56. Mao Q, GA Scarborough. Purification of functional human P-glycoprotein expressed in Saccharomyces cerevisiae. Biochimica et biophysica acta 1997;1327(1):107–18.

57. Lee SH, GA Altenberg. Expression of functional multidrug-resistance protein 1 in Saccharomyces cerevisiae: effects of N- and C-terminal affinity tags. Biochemical and biophysical research communications 2003;306(3):644–9.

44 | 2 Structure and function of hepatic ABC transporters

58. Chloupkova M, et al. Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 2007;46(27):7992–8003.
59. Mita S, et al. Inhibition of bile acid transport across Na+/taurocholate cotransporting polypeptide (SLC10A1) and bile salt export pump (ABCB 11)-coexpressing LLC-PK1 cells by cholestasis-inducing drugs. Drug metabolism and disposition: the biological fate of chemicals 2006;34(9):1575–81.
60. Stieger B. The role of the sodium-taurocholate cotransporting polypeptide (NTCP) and of the bile salt export pump (BSEP) in physiology and pathophysiology of bile formation. Handbook of experimental pharmacology 2011;(201):205–59.
61. Yamaguchi K, et al. Measurement of the transport activities of bile salt export pump using LC-MS. Analytical sciences: the international journal of the Japan Society for Analytical Chemistry 2009;25(9): 1155–8.
62. Hirano M, et al. Bile salt export pump (BSEP/ABCB11) can transport a nonbile acid substrate, pravastatin. The Journal of pharmacology and experimental therapeutics 2005;314(2):876–82.
63. Pauli-Magnus C, PJ Meier, B Stieger. Genetic determinants of drug-induced cholestasis and intrahepatic cholestasis of pregnancy. Seminars in liver disease 2010;30(2):147–59.
64. (EMA), T.E.M.A., Guideline on the Investigation of Drug Interactions 2010: 38.
65. Stapelbroek JM, et al. Liver disease associated with canalicular transport defects: current and future therapies. Journal of hepatology 2010;52(2):258–71.
66. Ortiz DF, et al. Identification of HAX-1 as a protein that binds bile salt export protein and regulates its abundance in the apical membrane of Madin-Darby canine kidney cells. The Journal of biological chemistry 2004;279(31):32761–70.
67. van der Bliek, AM, et al. Sequence of mdr3 cDNA encoding a human P-glycoprotein. Gene 1988;71(2):401–11.
68. Gros P, E Buschman. The mouse multidrug resistance gene family: structural and functional analysis. International review of cytology 1993;137C:169–97.
69. Smit JJ, et al. Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease. Cell 1993;75(3):451–62.
70. Smith AJ, et al. Hepatocyte-specific expression of the human MDR3 P-glycoprotein gene restores the biliary phosphatidylcholine excretion absent in Mdr2 (-/-) mice. Hepatology 1998;28(2):530–6.
71. Smith AJ, et al. The human MDR3 P-glycoprotein promotes translocation of phosphatidylcholine through the plasma membrane of fibroblasts from transgenic mice. FEBS letters 1994;354(3):263–6.
72. van Helvoort A, et al. MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. Cell 1996;87(3):507–17.
73. Higgins CF, MM Gottesman. Is the multidrug transporter a flippase? Trends in biochemical sciences 1992;17(1):18–21.
74. Elferink RP, et al. Class III P-glycoproteins mediate the formation of lipoprotein X in the mouse. The Journal of clinical investigation 1998;102(9):1749–57.
75. Oude Elferink RP, AK Groen. Mechanisms of biliary lipid secretion and their role in lipid homeostasis. Seminars in liver disease 2000;20(3):293–305.
76. Gros P, et al. Cloning and characterization of a second member of the mouse mdr gene family. Molecular and cellular biology 1988;8(7):2770–8.
77. Buschman E, P Gros. Functional analysis of chimeric genes obtained by exchanging homologous domains of the mouse mdr1 and mdr2 genes. Molecular and cellular biology 1991;11(2):595–603.

78. Buschman E, P Gros. The inability of the mouse mdr2 gene to confer multidrug resistance is linked to reduced drug binding to the protein. Cancer research 1994;54(18):4892–8.

79. Raymond M, et al. Physical mapping, amplification, and overexpression of the mouse mdr gene family in multidrug-resistant cells. Molecular and cellular biology 1990;10(4):1642–51.

80. Kino K, et al. Aureobasidin A, an antifungal cyclic depsipeptide antibiotic, is a substrate for both human MDR1 and MDR2/P-glycoproteins. FEBS letters 1996;399(1–2):29–32.

81. Smith AJ, et al. MDR3 P-glycoprotein, a phosphatidylcholine translocase, transports several cytotoxic drugs and directly interacts with drugs as judged by interference with nucleotide trapping. The Journal of biological chemistry 2000;275(31):23530–9.

82. Ruetz S, P Gros. Phosphatidylcholine translocase: a physiological role for the mdr2 gene. Cell 1994;77(7):1071–81.

83. Sleight RG, RE Pagano. Transbilayer movement of a fluorescent phosphatidylethanolamine analogue across the plasma membranes of cultured mammalian cells. The Journal of biological chemistry 1985;260(2):1146–54.

84. Margolles A, et al. The purified and functionally reconstituted multidrug transporter LmrA of Lactococcus lactis mediates the transbilayer movement of specific fluorescent phospholipids. Biochemistry 1999;38(49):16298–306.

85. van Genderen I, G van Meer. Differential targeting of glucosylceramide and galactosylceramide analogues after synthesis but not during transcytosis in Madin-Darby canine kidney cells. The Journal of cell biology 1995;131(3):645–54.

86. Baykov AA, OA Evtushenko, SM Avaeva. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. Analytical biochemistry 1988;171(2):266–70.

87. Kornberg A, WE Pricer, Jr. Enzymatic phosphorylation of adenosine and 2,6-diaminopurine riboside. The Journal of biological chemistry 1951;193(2):481–95.

88. Lindsley JE. Use of a real-time, coupled assay to measure the ATPase activity of DNA topoisomerase II. Methods in molecular biology 2001;95:57–64.

89. Doyle LA, et al. A multidrug resistance transporter from human MCF-7 breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America 1998;95(26):15665–70.

90. Miyake K, et al. Molecular cloning of cDNAs which are highly overexpressed in mitoxantrone-resistant cells: demonstration of homology to ABC transport genes. Cancer research 1999;59(1):8–13.

91. Allikmets R, M Dean. Cloning of novel ABC transporter genes. Methods in enzymology 1998;292:116–30.

92. Kage K, T Fujita, Y Sugimoto. Role of Cys-603 in dimer/oligomer formation of the breast cancer resistance protein BCRP/ABCG2. Cancer science 2005;96(12):866–72.

93. Henriksen U, et al. Identification of intra- and intermolecular disulfide bridges in the multidrug resistance transporter ABCG2. The Journal of biological chemistry 2005;280(44):36926–34.

94. Wakabayashi K, et al. Intramolecular disulfide bond is a critical check point determining degradative fates of ATP-binding cassette (ABC) transporter ABCG2 protein. The Journal of biological chemistry 2007;282(38):27841–6.

95. Polgar O, et al. Mutational studies of G553 in TM5 of ABCG2: a residue potentially involved in dimerization. Biochemistry 2006;45(16):5251–60.

96. Motamedi K, et al. Villonodular synovitis (PVNS) of the spine. Skeletal radiology 2005;34(4):185–95.

97. Honjo Y, et al. Acquired mutations in the MXR/BCRP/ABCP gene alter substrate specificity in MXR/BCRP/ABCP-overexpressing cells. Cancer research 2001;61(18):6635–9.

46 | 2 Structure and function of hepatic ABC transporters

98. Ozvegy C, A Varadi, B Sarkadi. Characterization of drug transport, ATP hydrolysis, and nucleotide trapping by the human ABCG2 multidrug transporter. Modulation of substrate specificity by a point mutation. The Journal of biological chemistry 2002;277(50):47980–90.

99. Robey RW, et al. Mutations at amino-acid 482 in the ABCG2 gene affect substrate and antagonist specificity. British journal of cancer 2003;89(10):1971–8.

100. Ejendal KF, et al. The nature of amino acid 482 of human ABCG2 affects substrate transport and ATP hydrolysis but not substrate binding. Protein science: a publication of the Protein Society 2006;15(7):1597–607.

101. Woodward OM, et al. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. Proceedings of the National Academy of Sciences of the United States of America 2009;106(25):10338–42.

102. Furukawa T, et al. Major SNP (Q141K) variant of human ABC transporter ABCG2 undergoes lysosomal and proteasomal degradations. Pharmaceutical research 2009;26(2):469–79.

103. Polgar O, RW Robey, SE Bates. ABCG2: structure, function and role in drug response. Expert opinion on drug metabolism & toxicology 2008;4(1):1–15.

104. Merino G, et al. Breast cancer resistance protein (BCRP/ABCG2) transports fluoroquinolone antibiotics and affects their oral availability, pharmacokinetics, and milk secretion. Drug metabolism and disposition: the biological fate of chemicals 2006;34(4):690–5.

105. Merino G, et al. Transport of anthelmintic benzimidazole drugs by breast cancer resistance protein (BCRP/ABCG2). Drug metabolism and disposition: the biological fate of chemicals 2005;33(5):614–18.

106. Wang X, et al. Breast cancer resistance protein (BCRP/ABCG2) induces cellular resistance to HIV-1 nucleoside reverse transcriptase inhibitors. Molecular pharmacology 2003;63(1):65–72.

107. Wang X, et al. Induction of cellular resistance to nucleoside reverse transcriptase inhibitors by the wild-type breast cancer resistance protein. Biochemical pharmacology 2004;68(7):1363–70.

108. Chen ZS, et al. Transport of methotrexate, methotrexate polyglutamates, and 17beta-estradiol 17-(beta-D-glucuronide) by ABCG2: effects of acquired mutations at R482 on methotrexate transport. Cancer research 2003;63(14):4048–54.

109. Litman T, et al. Use of peptide antibodies to probe for the mitoxantrone resistance-associated protein MXR/BCRP/ABCP/ABCG2. Biochimica et biophysica acta 2002; 1565(1):6–16.

110. Fetsch JF, WB Laskin, M Miettinen. Nerve sheath myxoma: a clinicopathologic and immunohistochemical analysis of 57 morphologically distinctive, S-100 protein- and GFAP-positive, myxoid peripheral nerve sheath tumors with a predilection for the extremities and a high local recurrence rate. The American journal of surgical pathology 2005;29(12):1615–24.

111. Rabindran SK, et al. Fumitremorgin C reverses multidrug resistance in cells transfected with the breast cancer resistance protein. Cancer research 2000;60(1):47–50.

112. Allen JD, et al. Potent and specific inhibition of the breast cancer resistance protein multidrug transporter in vitro and in mouse intestine by a novel analogue of fumitremorgin C. Molecular cancer therapeutics 2002;1(6):417–25.

113. Qadir M, et al. Cyclosporin A is a broad-spectrum multidrug resistance modulator. Clinical cancer research: an official journal of the American Association for Cancer Research 2005;11(6):2320–6.

114. de Bruin M, et al. Reversal of resistance by GF120918 in cell lines expressing the ABC half-transporter, MXR. Cancer letters 1999;146(2):117–26.

115. Zhou S, et al. Increased expression of the Abcg2 transporter during erythroid maturation plays a role in decreasing cellular protoporphyrin IX levels. Blood 2005;105(6):2571–6.

116. Ozvegy C, et al. Functional characterization of the human multidrug transporter, ABCG2, expressed in insect cells. Biochemical and biophysical research communications 2001;285(1):111–17.

117. Pozza A, et al. Purification of breast cancer resistance protein ABCG2 and role of arginine-482. Cellular and molecular life sciences: CMLS 2006;63(16):1912–22.

118. Mao Q, et al. Functional expression of the human breast cancer resistance protein in Pichia pastoris. Biochemical and biophysical research communications 2004;320(3):730–7.

119. Jacobs A, et al. Recombinant synthesis of human ABCG2 expressed in the yeast Saccharomyces cerevisiae: an experimental methodological study. The protein journal 2011;30(3):201–11.

120. Janvilisri T, et al. Sterol transport by the human breast cancer resistance protein (ABCG2) expressed in Lactococcus lactis. The Journal of biological chemistry 2003;278(23):20645–51.

121. Pozza A, JM Perez-Victoria, A Di Pietro. Insect cell versus bacterial overexpressed membrane proteins: an example, the human ABCG2 transporter. Methods in molecular biology 2010;654:47–75.

122. Rosenberg MF, et al. The human breast cancer resistance protein (BCRP/ABCG2) shows conformational changes with mitoxantrone. Structure 2010;18(4):482–93.

123. McDevitt CA, et al. Purification and 3D structural analysis of oligomeric human multidrug transporter ABCG2. Structure 2006;14(11):1623–32.

124. Clark R, ID Kerr, R Callaghan. Multiple drugbinding sites on the R482G isoform of the ABCG2 transporter. British journal of pharmacology 2006;149(5):506–15.

125. Shukla S, et al. The calcium channel blockers, 1,4-dihydropyridines, are substrates of the multidrug resistance-linked ABC drug transporter, ABCG2. Biochemistry 2006;45(29):8940–51.

126. Pal A, et al. Cholesterol potentiates ABCG2 activity in a heterologous expression system: improved in vitro model to study function of human ABCG2. The Journal of pharmacology and experimental therapeutics 2007;321(3):1085–94.

127. Wang H, et al. Membrane topology of the human breast cancer resistance protein (BCRP/ABCG2) determined by epitope insertion and immunofluorescence. Biochemistry 2008;47(52):13778–87.

Chapter 4

Heterologous Overexpression and Mutagenesis of the Human Bile Salt Export Pump (ABCB11) Using DREAM (Directed REcombination-Assisted Mutagenesis)

PLoS one

Heterologous Overexpression and Mutagenesis of the Human Bile Salt Export Pump (ABCB11) Using DREAM (Directed REcombination-Assisted Mutagenesis)

Jan Stindt[1.], Philipp Ellinger[1.], Claudia Stross[2], Verena Keitel[2], Dieter Häussinger[2], Sander H. J. Smits[1], Ralf Kubitz[2], Lutz Schmitt[1]*

1 Institute of Biochemistry, Heinrich-Heine-University, Düsseldorf, Germany, 2 Clinic for Gastroenterology, Hepatology and Infectiology, Heinrich-Heine-University, Düsseldorf, Germany

Abstract

Homologous recombination in Saccharomyces cerevisiae is a well-studied process. Here, we describe a yeast-recombination-based approach to construct and mutate plasmids containing the cDNA of the human bile salt export pump (BSEP) that has been shown to be unstable in E. coli. Using this approach, we constructed the necessary plasmids for a heterologous overexpression of BSEP in the yeast Pichia pastoris. We then applied a new site-directed mutagenesis method, DREAM (Directed REcombination-Assisted Mutagenesis) that completely bypasses E. coli by using S. cerevisiae as the plasmid host with high mutagenesis efficiency. Finally, we show how to apply this strategy to unstable non-yeast plasmids by rapidly turning an existing mammalian BSEP expression construct into a S. cerevisiae-compatible plasmid and analyzing the impact of a BSEP mutation in several mammalian cell lines.

Introduction

Recombinant protein expression is a frequent necessity for biochemical studies of proteins, which cannot be obtained in high amounts from their natural source. Among these are many membrane proteins. Extensive expression screening is a vital initial step in the study of membrane proteins, and this stage involves substantial work with recombinant DNA to create the necessary expression constructs. Standard techniques of molecular biology, however, become limiting when working with gene sequences that are unstable in Escherichia coli. This is especially encountered in the case of mammalian membrane proteins [1,2,3]. Comprehensive studies that address the actual cloning, propagation, and manipulation of constructs containing these unstable DNA sequences are rare and often lack detailed descriptions of the difficulties and, more importantly, of the solutions. In the case of the human bile salt export pump (BSEP, ABCB11), the cloning of the cDNA into an expression vector only succeeded after a tremendous amount of work [2]. This process resulted in a plasmid with several point mutations in the coding sequence, six of which changed the sequence on the protein level. A general issue encountered during propagation and targeted mutagenesis of BSEP-containing plasmids is the loss of various parts of the BSEP cDNA sequence in E. coli [2]. This loss of sequences has been observed for other proteins as well [4]. The corresponding DNA sequences have generally been termed

"unstable" or even "toxic", because the presence of the intact plasmid ultimately resulted in bacterial cell death.

One approach to create plasmids containing these "toxic" DNA fragments is to assemble it by homologous recombination in Saccharomyces cerevisiae (S. cerevisiae) thereby circumventing E. coli [5,6,7]. S. cerevisiae is able to recombine several overlapping fragments into one circular plasmid containing the desired cDNA. By incorporation of a suitable origin of replication (Ori) as well as a selection marker virtually any plasmid can be created for usage of recombination-based cloning by S. cerevisiae. A fragment containing both ORI and selection marker can be added together with the toxic target cDNA in a single recombination step yielding an intact and most importantly stable expression plasmid.

We have used S. cerevisiae to create such an expression plasmid containing the "toxic" coding sequence of human BSEP which was subsequently used for BSEP expression in S. cerevisiae and Pichia pastoris (P. pastoris). Severe hereditary diseases of the liver are directly associated with mutations in the BSEP transporter [8,9]. Previously heterologous BSEP expression was only demonstrated in insect cells [2], which hampered a detailed analysis of function of both wild-type BSEP and clinically relevant mutations in vitro. Besides creating an expression plasmid for BSEP in P. pastoris the method described here is also used to directly create BSEP mutants in the yeast plasmid for subsequent expression in mammalian cell lines. This highlights the applicability of this

98

method to both "simple" expression systems like the yeast based as well as more sophisticated expression in mammalian cell lines.

Results

Cloning and expression of BSEP

The unicelluar eukaryote S. cerevisiae was initially chosen because of three advantages: (i) it can perform efficient homologous recombination [5,7]; (ii) expression of other eukaryotic ABC transporters has been successfully reported [10]. For example, S. cerevisiae has been used to express the BSEP homologue MDR1 [11,12]. (iii) Transformants resulting from in vivo homologous recombination can immediately be tested for target protein expression. We used these advantages for BSEP, but expression levels in S. cerevisiae were very low and not sufficient for subsequent purification or activity studies (Figure 1B, left panel).

Therefore, we changed the expression system from S. cerevisiae to P. pastoris. Here, expression is driven from the strong inducible AOX1 promoter. In addition, this yeast strain can reach high cell densities and thereby lead to substantial amounts of membrane protein [13,14,15]. Furthermore, Chloupkova et al. were able to express 25 human ABC transporters in P. pastoris [15], however BSEP was not among them. We used the P. pastoris integration vector pPIC3.5, which was prepared for manipulation in S. cerevisiae by integrating the relevant sequence that is necessary for maintenance (ORI) and selection in this yeast. A PCR product containing the S. cerevisiae 2 micron ORI and a leucine prototrophy marker and a second PCR product containing BSEP with an C-terminal his_8 tag (kind gift of Dr. Kenneth Linton) were simultaneously recombined into pPIC3.5 in vivo in S. cerevisiae (Figure 1A). The resulting derivative pPIC3.5-C_{his}BSEP (Figure S1) is identical to the construct that would be obtained by
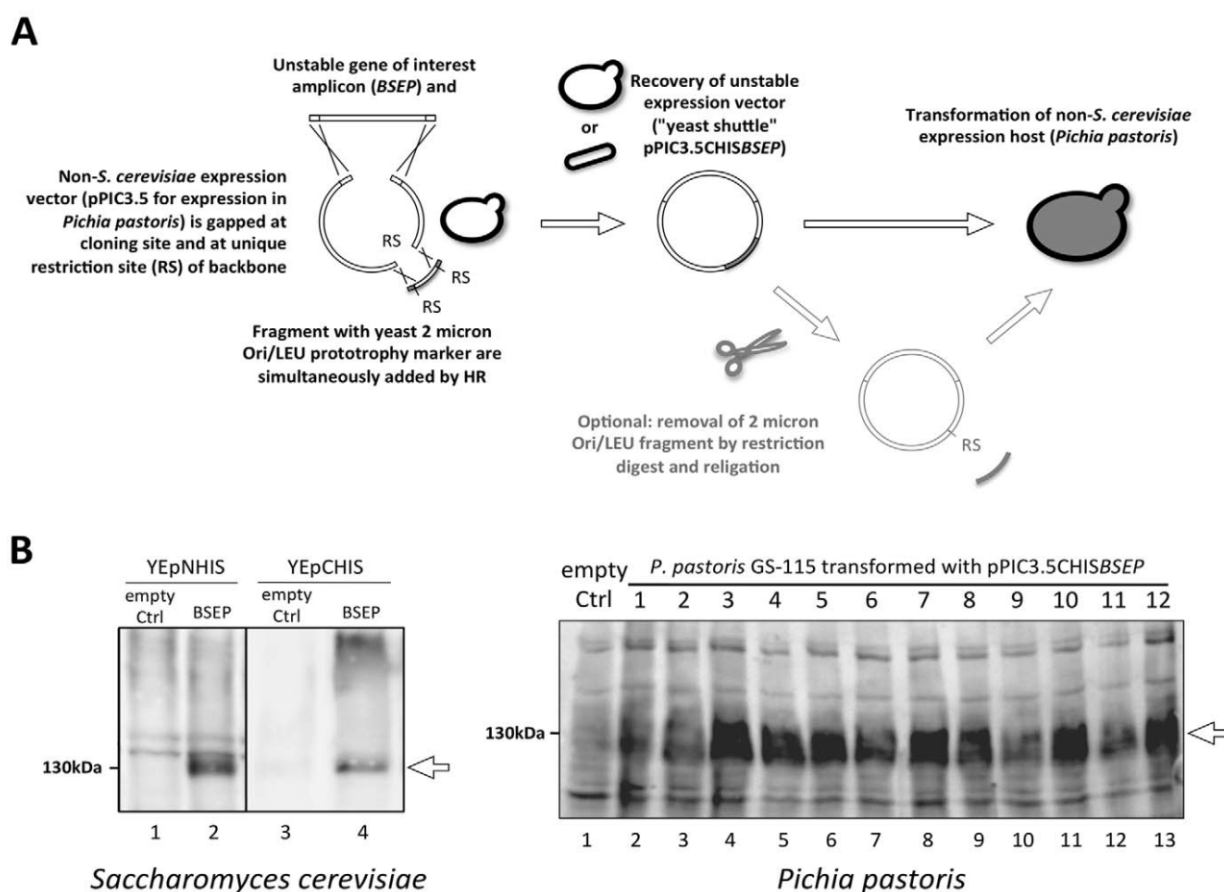

Figure 1. Heterologous overexpression of BSEP in *Saccharomyces cerevisiae* and *Pichia pastoris*. A, Toxic or unstable expression plasmids can be constructed for any system in S. cerevisiae by adding the necessary sequence to the plasmid backbone. In order to clone BSEP into the Pichia pastoris expression cassette on pPIC3.5, the recombination vector was double-digested to allow the simultaneous insertion of both the unstable BSEP coding sequence and a PCR-generated fragment of the YEpHIS plasmid carrying the 2 micron origin (Ori) of replication and the leucine (LEU) prototrophy marker by homologous recombination (RS=NdeI). This plasmid was recovered from S. cerevisiae and obtained in preparative amounts from E. coli by strict cultivation at 30uC under suitable conditions. B, Expression of human BSEP in S. cerevisiae and P. pastoris. Equal amounts of whole yeast cell extracts were resolved on SDS-PAGE, electroblotted and probed with the polyclonal BSEP antiserum K168. Left panel, Homologous recombination was used to construct both BSEP expression vectors directly in S. cerevisiae. NHIS/CHIS, N- or C-terminal his tag position; empty Ctrl, strain transformed with corresponding empty YEpHIS expression plasmid. Right panel, pPIC3.5-CHISBSEP was constructed as described in A and used to transform P. pastoris strain GS-115 by electroporation. Empty Ctrl, P. pastoris GS-115 strain transformed with the empty pPIC3.5 integration vector.
doi:10.1371/journal.pone.0020562.g001

conventional bacterial cloning, with the exception of the introduced ORI and selection marker. The plasmid was used to transform P. pastoris. Abundant colonies were obtained and twelve of these were subsequently analyzed for BSEP expression. As demonstrated by Western blot analysis, all of the tested clones expressed BSEP at similar levels (Figure 1B, right panel). A strain transformed with the empty integration vector was used as a negative control. The yield of recombinant BSEP achieved in P. pastoris is substantially higher than in S. cerevisiae allowing further purification and subsequent biochemical analysis.

DREAM - A site-directed mutagenesis method for unstable and toxic plasmids

Several severe hereditary diseases are known to be associated with human ABC transporter genes [16]. To date, 146 BSEP mutations have been reported in the Human Gene Mutation Database [17], The vast majority of which are associated with liver diseases. One of the most frequently used methods to generate specific mutations is the site-directed mutagenesis (SDM) procedure [18]. This method relies on the usage of E. coli to turn the linear product obtained by an in vitro mutagenesis into a circular plasmid via nick repair. However, since the cDNA BSEP is toxic for E.coli, this standard method is not applicable. Therefore, we used S. cerevisiae as host.

Classic SDM relies on the removal of non-mutated template plasmid achieved by DpnI digestion, which recognizes and cleaves only methylated DNA template. Thereby only the mutated plasmid can give colonies. Plasmids prepared directly from S. cerevisiae, however are, unmethylated and the template plasmid cannot be removed by DpnI digestion [19,20,21]. Thereby, the efficiency of mutagenesis is too low. To obtain positive clones, several hundred nanograms of mutated plasmid are needed and a substantial amount of time is required to pick and analyze several clones to find the correctly mutated plasmid [22,23,24]. By changing the mutagenesis primer design from a complete to a partial, 5′-overlap of the primer pair (Figure 2A, step 1), the linear in vitro mutagenesis step is turned into an exponential polymerase chain reaction: due to this primer shift a product is generated which carries priming sites that serve as a template in the subsequent reaction cycles (step 2). The usefulness of such a primer shift was previously reported, although in a different context [25]. The reaction product is therefore endowed with homologous double-stranded ends that allow the precise recirculation by homologous recombination into an intact plasmid (step 3; for a detailed comparison of both, classic and DREAM mutagenesis, see Figure S2). A change of primers to mutated plasmid DNA used for transformation thereby increases the probability of picking positive mutated clones.

We analyzed the efficiency of this yeast-based mutagenesis method by introducing a mutation into BSEP on a S. cerevisiae plasmid that resulted in an additional recognition site for the restriction enzyme BstBI (Figure 2B). The mutagenesis reaction gave , 100 colonies after transformation. Twenty-five of the resulting transformants were picked and the mutated region of the plasmid was amplified by colony PCR. The product of 1240 bp was then subjected to BstBI digestion, which - assuming that the site directed mutagenesis has been successful - should result in two bands of 960 and 280 bp, respectively. Since no BstB1 site was present in the amplified DNA, the PCR product of colonies not bearing the mutation cannot be digested by this restriction enzyme. From 25 clones picked, 19 carried the introduced mutation, corresponding to a mutagenesis efficiency of 76%. Five of these mutated plasmids where sequenced and all of them were positive for the mutation. The efficiency of two other mutations

generated by this strategy was 74%. This demonstrates that our method is comparable to the 80% efficiency reported for the E. coli-dependent, classical SDM [26].

Rapid DREAM mutagenesis and expression analysis of a yeast-enabled mammalian BSEP expression vector

Many eukaryotic membrane proteins are expressed and studied in mammalian cell lines [27,28]. In order to extend the usage of the DREAM method to human cell culture, a yeast-compatible derivative of the mammalian BSEP expression vector pEYFP-N1-BSEP [29] (Figure S1) was created by introducing the necessary S. cerevisiae ORI and selection marker(in analogy to the P. pastoris vector modification shown in Figure 1A). After introduction and sequence verification, the obtained plasmid construct was successfully transfected into HEK293 cells as visualized by detection of the fluorescent YFP tag at the C-terminus of the BSEP fusion protein (Figure 3). Comparison with the parental non-yeast expression vector showed that in both cases BSEP is localized at the plasma membrane (Figure 3A and B). Additionally, flow cytometric analyses of HEK293 cells transfected with equimolar amounts of BSEP in pEYFP-N1 and in pEYFP-N1-OriLeu, respectively, were performed to determine transfection and expression rates of both plasmids (Figure S3). Here, transfection rates and BSEP expression from the larger yeast shuttle vector and the smaller, parental vector were found to be comparable (21% and 32%, respectively). These experiments, show that the presence of the additional ORI and selection marker does not influence the expression of BSEP in mammalian cell cultures.

We introduced a BstBI mutation by the E. coli-free DREAM method as described above. After plasmid recovery, expression was analyzed by transfection in different mammalian cell lines (Figure 3C). The green fluorescence of the BSEP mutant fusion protein could be detected in HEK293, MDCK and HepG2 cells. Here, a mislocalization of the mutated ABC transporter from the plasma membrane to intracellular compartments could be observed (Figure 3C). This mislocalization has also been observed in a liver biopsy from a patient carrying this BSEP mutation (Kubitz and Häussinger, unpublished data).

Discussion

Here we present a complete, yeast-based workflow to create and manipulate unstable or toxic genes for any expression system. By using this method, we were able to express the "toxic" human wild-type BSEP cDNA both in S. cerevisiae and P. pastoris. Furthermore it is possible to rapidly generate and study BSEP mutations and to use the resultant plasmids in yeast and mammalian expression systems without further modifications. We describe and provide proof of principle for DREAM, a new site-directed mutagenesis method that allows the site-specific manipulation of plasmids in S. cerevisiae. This method can be used for many plasmids by simple addition of the sequence necessary for S. cerevisiae plasmid maintenance. Moreover, these plasmids can also be obtained in sufficient quantity and quality directly from S. cerevisiae cultures [30]. The presented approach can rely on S. cerevisiae as the single organism used in all cloning steps from plasmid construction and propagation as well as its mutagenesis.

Homologous recombination has other additional advantages: since the recombination process is completely independent of restriction sites, the in-frame fusion of an insert to e.g. a plasmid-encoded tag never results in unwanted additional amino acids on the recombinant protein, while this is often the case in restriction- or LIC-based cloning [31,32] (ligation-independent cloning). At
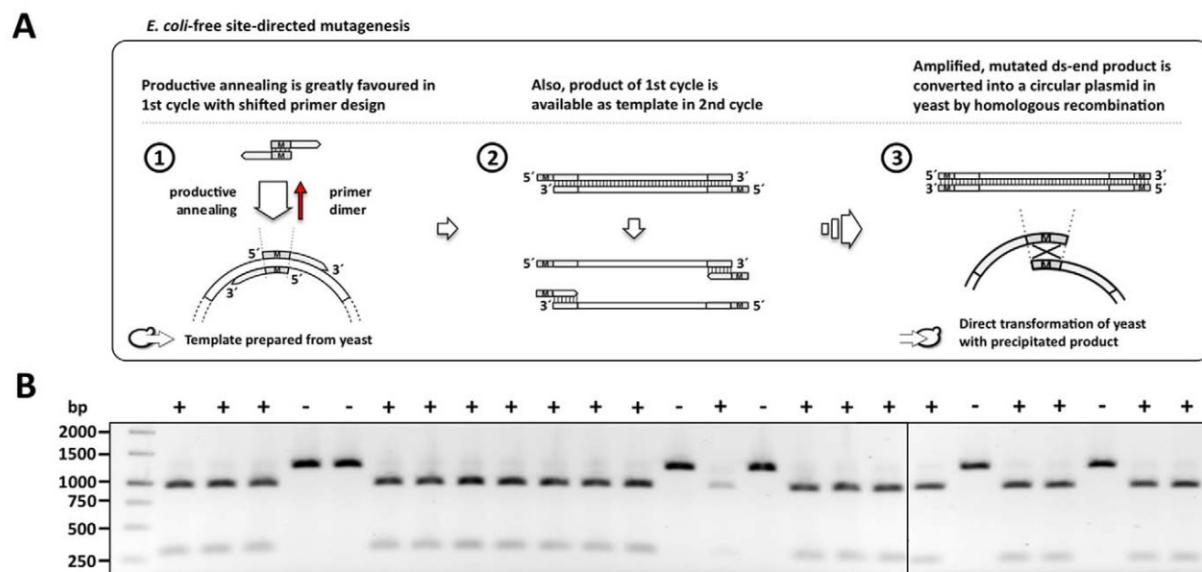
Figure 2. The DREAM method allows site-directed mutagenesis of an unstable *BSEP*-carrying plasmid without the need for *E. coli*. A, The product yield can be increased significantly by shifting the primer binding sites from a complete to a partial overlap that allows binding of the primers to product in subsequent cycles. The minute amount (10 ng) of non-mutated template used in the reaction is easily outnumbered by the yield of mutagenesis product, which is double-stranded (ds) due to the modified primer design (see also Figure S2). This allows for the direct transformation of *S. cerevisiae* without the need of prior template removal. The mutagenesis product is then recircularized by homologous recombination of the double-stranded ends. B, Analysis of *S. cerevisiae* transformants obtained from the modified SDM protocol shows a high mutagenesis efficiency. Successful mutagenesis in this case results in the addition of an BstBI restriction site into the BSEP coding sequence. Colony PCR of the resulting transformants was performed with primers surrounding the mutagenesis site, and the resulting product was digested with BstBI. 19 of the 25 tested clones carried the additional restriction site (+), corresponding to a mutagenesis efficiency of 76%.
doi:10.1371/journal.pone.0020562.g002

the same time, homologous recombination is generally as high-throughput-compatible as LIC.

DREAM - Directed REcombination-Assisted Mutagenesis

The modified primer design of the yeast-based DREAM mutagenesis method is simple and like the classic SDM covers all aspects of mutagenesis: mutation, insertion and deletion (see Figure S2). The DREAM mutagenesis primer pair can easily be designed by the following basic rules: each oligonucleotide should be around 50 bases long, with 20 bases of 59 primer-to-primer overlap for an efficient recombination of the mutated plasmid ends, carrying the mutation, deletion or insertion in their middle, and 30 bases of 39-sequence for template annealing. The 20 bp overlap on both ends of the linear mutagenesis product results in efficient homology-based gap repair in *S. cerevisiae* to form the circular, mutated plasmid (Figure 2). To minimize the occurrence of errors, the number of thermal cycles is restricted to 18 as recommended for the conventional SDM kit [26]. Accordingly, all analyzed DREAM clones so far replicated the template plasmid sequence i.e. apart from the introduced mutation the whole BSEP coding sequence was found to be unchanged.

The new mutagenesis method permits the rapid realization of patient-derived BSEP mutations for immediate study in cell cultures. Using DREAM, we could show that a BSEP mutation identified in a patient with progressive familial intrahepatic cholestasis type 2 results in a trafficking defect of the mutant protein that prevents BSEP from being correctly incorporated into the plasma membrane. Future mutations can be generated quickly for their study in mammalian cell lines and/or in vitro on the

isolated recombinant protein. This is a major advantage since the realization of for example BSEP mutations was previously a work-intensive and time-consuming task. Glycosylation has been shown to be irrelevant for the function of other human ABC transporters expressed in yeast in the past [12,33]. Furthermore, BSEP expressed in Sf9 cells, which also harbors a glycosylation pattern different from the human pattern, was functional [2]. Thus, it is very likely that the glycosylation state and/or pattern of BSEP is not relevant for its function.

Concluding remarks

We have established and validated a complete new workflow for the cloning and manipulation of "unstable or toxic" DNA. We believe that in particular the new E. coli-independent DREAM mutagenesis strategy will be helpful for proteins where functionally relevant mutants of proteins could not be studied due to limitations of their bacterial cloning.

Materials and Methods

Yeast strains and growth conditions

The *Saccharomyces cerevisiae* strain used in this study was the S288C derivative YRE1001 (MATa; ura3–52; trp1-1; leu2–3,112; his3–11, 15; ade2-1; pdr1–3;DPDR5,DPDR5prom::TRP1) [34]. For expression in *Pichia pastoris*, strain GS-115 from Invitrogen was used. Yeast cells were maintained either on YPD agar (1% (w/v) yeast extract, 2% (w/v) peptone from caseine, 2% (w/v) glucose, and 2% (w/v) agar for solid media) or synthetic complete (SC) minus leucine media [35] at 30uC, and liquid cultures were shaken at 200 (S. cerevisiae) or 250 rpm (P. pastoris), respectively.
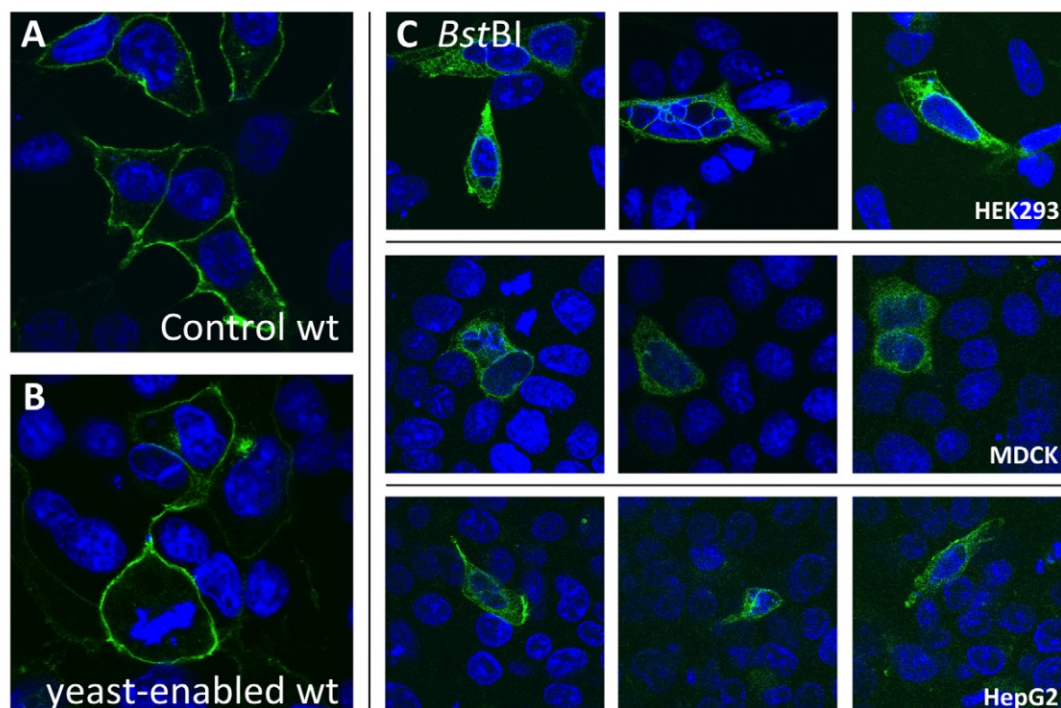
Figure 3. The yeast-maintainable *BSEP* expression plasmid is functionally indistinguishable from the non-yeast parental plasmid in mammalian cell culture. A, HEK293 cells were transfected with pEYFP-N1-BSEP. B, HEK293 cells were transfected with pEYFP-N1-OriLeu-BSEP. Equimolar amounts of both constructs showed comparable transfection efficiencies and expression levels (established by FACS analysis, see Figure S3). C, HEK293 (upper panel), MDCK (middle), and HepG2 cells (lower panel) were transfected with the BstBI mutation in pEYFP-N1-OriLeu-BSEP that was introduced with the E coli-free site-directed mutagenesis method. All cells were transfected with equimolar amounts of the respective constructs via Lipofectamine according to the manufacturer's guidelines. After fixation, nuclei were stained with Hoechst 34580 (blue), and the fluorescence (green) of the YFP tag in the BSEP fusion protein was observed.
doi:10.1371/journal.pone.0020562.g003

Molecular biology reagents, kits, and oligonucleotides

All polymerase-based reactions were performed with the Phusion HHigh Fidelity DNA polymerase (NEB) and the supplied HF buffer. Restriction enzymes were from NEB and Fermentas. Oligonucleotides were ordered from Eurofins MWG Operon (Ebersberg, Germany). Plasmid miniprep kits were obtained from Qiagen (Hilden, Germany), and the Nucleobond HXtra midiprep kit from Macherey-Nagel was used (Düren, Germany).

Generation of PCR products and preparation of vectors for homologous recombination in S cerevisiae

The Clone Manager Suite 6 (Sci-Ed Software) was used to design all oligonucleotides and calculate annealing temperatures of the primer pairs. All oligonucleotides and plasmids used in this study can be found in Table S1 and S2, respectively. All PCR reactions were performed with the PhusionH DNA polymerase in HF buffer according to the recommendations of the manufacturer. 50 ml reactions contained 1.5 ml of DMSO. PCR conditions were: 2 min initial denaturation, 50 sec cycle denaturation, 50 sec annealing, 20 sec per kbp of extension for 35 cycles, followed by 7 min of final extension. For homologous recombination, all vectors were linearized by restriction digest as indicated during which they also were dephosphorylated with calf intestinal alkaline phosphatase (Fermentas). YEpMDR1HIS was double-digested with BamHI and BsmI, YEpHIS was cut with BamHI and MluI to remove the C-terminal his tag [12]. The

oligonucleotides YEpNHISFor and YEpNHISRev encoding an N-terminal his_{14} tag followed by a factor X_a cleavage site were mixed in equimolar amounts, heated to 95uC for 5 min and allowed to anneal by slow cooling to room temperature. After T4 polynucleotide kinase (NEB) treatment, the phosphorylated synthetic insert was ligated into the gel-purified linearized YEpHIS plasmid. BSEP PCR products with fitting overlaps were generated with the primer pairs BSEP-YEpHISN-S1/-S2 and BSEP-YEpHISC-S1/-S2, respectively.

Transformation of competent S cerevisiae cells

Competent cells were generated as described elsewhere [24]. Briefly, a 5 ml overnight culture was used to inoculate 50 ml YPD to an OD_{600} of 0.2. Cells were harvested at an OD_{600} of 0.8 to 1 and washed once with 50 ml of sterile water. The cell pellet was then washed once in 1 ml of LATE buffer (0.1 M lithium acetate, 10 mM Tris-HCl pH 8, 1 mM EDTA) and adjusted to 50 ml LATE per 10 OD_{600} equivalents of cells. Cells were directly used or stored for a maximum of two days at 4uC. Competent yeast cells were transformed either with intact plasmid or equimolar amounts of DNA fragments for homologous recombination using the Lithium-Acetate/PEG method [24] without single-stranded DNA, and transformants were selected on solid SC minus leucine media. The DNA for yeast transformation was salt/ethanol-precipitated, briefly dried and directly redissolved in the yeast suspension. Resulting transformants were directly used to

inoculate 5 ml overnight cultures. From these, 5 ml cultures were inoculated to an OD_{600} of 0.3 and harvested in the early logarithmic growth stage (OD_{600} of 1).

Analysis of BSEP expression in Pichia pastoris

The primer pair OriLeu-pPIC3.5-NdeI-S1/-S2 was used to amplify the region of the YEpHIS plasmid containing the 2 micron ori and the leucine prototrophy marker (the sequence of YEpLacl81, the parental vector of YEpHIS1 containing the ori and marker information can be found under Genbank acc. no. X75460.1), and BSEP was amplified with primers BSEP-pPIC3.5-S1/-S2. pPIC3.5 was double-digested with NdeI and BamHI, and all fragments were pooled in equimolar amounts and used for transformation of S. cerevisiae. After plasmid recovery in E. coli XL1blue, the Pichia pastoris strain GS-115 was transformed according to the Invitrogen guidelines. Clones obtained from transformation were subjected to a second round of selection on MD plates without histidine. From these, 10 ml MGY cultures were grown overnight at 30uC and 250 rpm shaking. Cells were pelleted at an OD_{600} between 2 to 6 (1500 g, 5 min, 4uC) and resuspended in MMY medium to a final OD_{600} of 1. After additional growth for 24 h, 10 OD_{600} equivalents were harvested and used to make whole cell lysates. Cells were washed once with water, resuspended in 1 ml of water, and lysed on ice for 10 min by addition of 150 ml of YEX buffer (1.5 M NaOH, 7.5% (v/v) b-mercaptoethanol [36]). The solution was precipitated by addition of 150 ml of 50% (w/v) trichloroacetic acid for 10 min on ice, and the precipitate was harvested by centrifugation (10 min, 14.000 rpm, 4uC). After complete removal of supernatants, the pellets were dried briefly at room temperature and resuspended in 100 ml of sample buffer (for composition, see [36]). Samples were incubated for 10 min at 65uC and spun down briefly. 0.5 OD_{600} equivalents were then loaded onto 7% SDS gels and separated at 150 V. After semi-dry electroblotting (Biorad) onto nitrocellulose, the membrane was blocked for 30 min in TBS-T with 5% non-fat dried milk and then probed with a 1:2500 dilution of the polyclonal rabbit antiserum K168 directed against human BSEP [28].

For S. cerevisiae, 2 OD_{600} equivalents of cells were taken and processed. the pellets were resuspended in 80 ml of sample buffer, and 20 ml (0.5 OD_{600} equivalents) were resolved on a 7% SDS gel.

The E coli-free site-directed mutagenesis (DREAM)

For introduction of the missense, BstBI mutation into BSEP constructs, the mutagenesis primer pair BSEP-BstBImut-S1/-S2 was used. Cycling conditions were as in the classic SDM protocol, with 18 cycles to avoid PCR-induced errors, and with the annealing temperature strictly being kept at 60uC to ensure the generation of ds-ended mutagenesis product (also see Figure S1). Reactions were set up with the Phusion HDNA polymerase, as this proofreading enzyme has a low error rate practically identical to the enzyme used in the Stratagene SDM kits (according to the manufacturers' datasheets: $4.46 \ 10^{27}$ for Phusion H High Fidelity Polymerase, $4.36 \ 10^{27}$ for PfuUltra H HF Polymerase). Extension time was 1 min per kb to allow for complete extension. The recommended elongation time for Phusion polymerase is 15–30 seconds per kbp, and an extended incubation at the elongation step was found to be important for a successful exponential product generation, probably because it ensures the quantitative integrity of the ends that serve as priming sites in subsequent cycles. 10 ng of template was used, and the reaction was initiated after heating the reaction to 98uC (hot start) with 0.5 ml of Phusion polymerase. The reaction was precipitated with sodium acetate and ethanol, and directly resuspended in 20 ml (= 4 OD_{600} equivalents) of fresh competent yeast cells in LATE buffer. This

assures an optimal ratio of mutagenesis product to yeast cells. The transformation was carried out as described above by addition of 120 ml PLATE buffer (40% (v/v) PEG 4000, 0.1 M lithium acetate, 10 mM Tris-HCl pH 8, 1 mM EDTA).

Yeast colony PCR

Freshly growing transformant colonies were picked and restreaked onto selective agar plates. The remaining of the material was treated with Zymolyase for 45 min at 37uC in 50 ml reaction volume (0.1 M sodium phosphate buffer pH 7.4, 2 mM DTT, and 5–10 mg/ml of Zymolyase T-100 (ICN)). After heating the reactions to 95uC for 10 min, the material was frozen at 2 20uC for 10 min and thawed again. 5 ml of this were used as template for colony PCR (30 ml reaction in Phusion HF buffer: 2 mM final conc. of $MgCl_2$, 20 pmol/primer, 200 mM dNTPs, 1.5 units Phusion H DNA polymerase). 10 ml of the reaction was directly digested with BstBI (15 ml reaction, 8 units restriction enzyme) for 1 h at 37uC and then resolved on a 1% agarose gel.

Plasmid recovery from yeast

Small- (5–10 ml for E. coli transformation) or large-scale (0.5–1 l for obtaining preparative plasmid amounts) overnight yeast liquid cultures were harvested, washed in cold water once, and then resuspended in Zymolyase incubation buffer (see above, with 1.2 M sorbitol). The yeast cell wall was digested for 30 min at 37uC, cells were then lysed by alkaline lysis as described in [30]. 0.5–1 ml of this preparation was used to transform chemically competent E. coli strain XL1blue. After heat shock, bacteria were shaken for 1 h at 30uC and 200 rpm. This low temperature is crucial to prevent the loss of unstable construct. After plating out on low salt LB media containing 50 mg/ml Carbenicillin, plates were incubated at 30uC. Carbenicillin allows for a tighter and longer lasting selection as it hydrolyzes much slower than ampicillin. 150 ml low salt LB with Carbenicillin were then directly inoculated with a single colony and allowed to grow for 36 to 48 hours at 30uC and 200 rpm. Cells were harvested and plasmid prepared from these. The plasmids were sequence-verified.

Transfection of cell lines with wt and BstBI-mutated BSEP-YFP and immunofluorescence analyses

pEYFP-N1-BSEP was made yeast-compatible as described for pPIC3.5 by addition of an Ori/Leu PCR product made from YEpHIS with primer pair OriLeu-pEYFP-AflII-S1/-S2.

HepG2 cells (ATCC, ordering number HB-8065) were cultured in Dulbecco's modified Eagle's medium Nutrimix F12 (DMEM-F12; Invitrogen), HEK293 cells (ATCC, ordering number CRL-1573) were cultured in DMEM and MDCK cells were cultured in MEM with Earle's Salts, each containing 10% fetal calf serum (PAA, Coelbe, Germany), in a humidified, 5% CO_2 atmosphere at 37uC. The indicated BSEP-YFP plasmid DNA was transfected using FuGENE H HD (Roche) according to the manufacturer's guidelines. For fluorescence microscopy (LSM 510, Zeiss, Oberkochen, Germany) cells were fixed and permeabilized with methanol (100%, 4uC, 1 min) and nuclei were stained with Hoechst 34580 (Invitrogen).

Flow cytometric analysis

HEK293 cells transiently transfected with equimolar amounts of BSEPwt in pEYFP-N1 (1 mg) and in pEYFP-N1-OriLeu (1.38 mg), respectively, and untransfected cells were cultured in 12-well culture plates until subconfluence. For flow cytometric analyses they were washed with ice-cold phosphate buffered saline (PBS) and incubated with Accutase at 37uC. Cells were

transferred into 1.5 ml tubes, centrifuged for 30 sec at 45006 g and resuspended in FACS buffer (PBS+5% (v/v) FCS+0,1 % (w/v) NaN₃). Cell size, granularity and fluorescence intensities were measured by a FACSCanto Flow Cytometer (Becton Dickinson, Heidelberg, Germany) with excitation at 488 nm. EYFP fluorescence was measured at 5306 30 nm. Transfected cells were gated by comparison with the fluorescence of untransfected control cells.

Supporting Information

Figure S1 Maps of the "yeast-enabled" plasmids used in this study for expression and DREAM mutagenesis of the unstable BSEP cDNA. A, pPIC3.5-OriLeu-CHISBSEP for heterologous expression in Pichia pastoris. B, pEYFP-N1-OriLeu-BSEP for expression in mammalian cell culture. Plasmid features used for propagation in E. coli, S. cerevisiae, and the organism used for BSEP expression are indicated in red, green, and blue, respectively.
(DOC)

Figure S2 A simple modification of the classic site-directed mutagenesis protocol allows the mutagenesis of toxic or unstable plasmids without the need for E. coli. The classic site-directed mutagenesis (left cartoon side) results in a mutated and linear plasmid with single-stranded 5′-overhangs formed by the mutagenesis primers. Since the primers are absolutely complementary to each other, the product of the SDM reaction exists de facto in a non-covalently closed circular form that is nick-repaired after transformation into E. coli. The polymerase-involving mutagenesis reaction is, in contrast to standard PCR, non-exponential: the mutagenesis primers completely overlap (step 1), so the only primer binding sites on the generated SDM product would be at its very ends. These, however, are single-stranded (step 2), and cannot, in the second cycle, serve to further amplify the product of the first cycle (step 3). Instead, in each cycle the primers bind to the original template and generate a linear product with themselves forming the single-stranded 5′-overhangs (step 4). Along with the use of a proofreading DNA polymerase, this assures a minimum of PCR-introduced mutations, as only the original plasmid is copied and thus mutated in each of the 18 reaction cycles. However, because of the low product yield resulting from the linear template amplification, it is also necessary to remove the unmutated template to minimize the chance of picking wild type clones after transformation. The restriction enzyme DpnI is generally used to recognize and digest both the methylated plasmid template and hemimethylated heteroduplex strands while leaving intact the unmethylated mutagenesis product. Native yeast DNA, however, is unmethylated. We solved both the problem of low product yield and false positives by changing the mutagenesis primer design from a complete to a partial, 5′-overlap of the pair (right cartoon side, step 1). A shift of primer positions has previously been reported in a different context [4] and this changes the mutant strand synthesis reaction into a true exponential PCR reaction,

because with this primer design a product is generated (step 2) that carries binding sites for the primers and can serve as template in the subsequent reaction cycles (step 3). In classic SDM, the template has to compete with the oligonucleotides for priming, because they allow for just as much base pairing as itself, if the primer-template mismatch due to the mutation is not taken into account (left cartoon side, step 1). In contrast, the reduction of the primer complementarity to 20 bases greatly favors the 30 plus 20 bases annealing to the template in the first cycle (right cartoon side, step 2). More importantly yet, it still outcompetes in subsequent cycles, when the primers can anneal to the plasmid template with 50 bases and to the product template with 30 bases as compared to the 20 bases of unproductive primer-primer annealing. While the 10 additional bases do not drastically favor mutagenesis product priming in the second cycle, already at the end of this cycle the first product molecules with double-stranded ends appear, which subsequently are drastically preferred priming targets offering the full 50 bases. To make sure that the ends of the final reaction product are double-stranded, the annealing temperature is strictly kept at 60uC, which is well above the melting temperature for a 20-mer and prevents the protection of single-stranded ends by unproductive primer annealing.
(DOC)

Figure S3 FACS analysis of the unmodified and "yeast-enabled" mammalian BSEP expression vector pEYFP-N1-BSEP. The data indicate that while pEYFP-N1-OriLeu is transfected at a somewhat reduced yet comparable efficiency as the unmodified construct, while the mean fluorescence is even slightly higher. Taken together with Figure 3, this clearly shows that the addition of the Ori/Leu segment to the vector backbone does not compromise construct performance. The data shown here is representative for three independent transfection experiments.
(DOC)

Table S1 PCR primers used in this study.
(DOC)

Table S2 Plasmids used in this study.
(DOC)

Author Contributions

Conceived and designed the experiments: JS PE CS RK SHJS LS. Performed the experiments: JS PE CS. Analyzed the data: JS PE CS RK SHJS LS. Contributed reagents/materials/analysis tools: VK DH RK SHJS LS. Wrote the paper: JS PE CS VK DH SHJS RK LS.

References

1. Byrne JA, Strautnieks SS, Mieli-Vergani G, Higgins CF, Linton KJ, et al. (2002) The human bile salt export pump: characterization of substrate specificity and identification of inhibitors. Gastroenterology 123: 1649–1658.
2. Noe J, Stieger B, Meier PJ (2002) Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 123: 1659–1666.
3. Byrne JA, Strautnieks SS, Ihrke G, Pagani F, Knisely AS, et al. (2009) Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt

export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology 49: 553–567.
4. Vu K, Bautos J, Hong M-P, Gelli A (2009) The functional expression of toxic genes: Lessons learned from molecular cloning of CCH1, a high-affinity Ca2+ channel. Anal Biochem 393: 234–241.
5. Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58: 201–216.

104

6. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. Nucleic Acids Res 37: 6984–6990.

7. Oldenburg KR, Vo KT, Michaelis S, Paddon C (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. Nucleic Acids Res 25: 451–452.

8. Stieger B, Meier Y, Meier PJ (2007) The bile salt export pump. Pflugers Arch 453: 611–620.

9. Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, et al. (1998) A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet 20: 233–238.

10. Evans GL, Ni B, Hrycyna CA, Chen D, Ambudkar SV, et al. (1995) Heterologous expression systems for P-glycoprotein: E. coli, yeast, and baculovirus. J Bioenerg Biomembr 27: 43–52.

11. Kuchler K, Thorner J (1992) Functional expression of human mdr1 in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci USA 89: 2302–2306.

12. Figler RA, Omote H, Nakamoto RK, Al-Shawi MK (2000) Use of chemical chaperones in the yeast Saccharomyces cerevisiae to enhance heterologous membrane protein expression: high-yield expression and purification of human P-glycoprotein. Arch Biochem Biophys 376: 34–46.

13. Urbatsch IL, Wilke-Mounts S, Gimi K, Senior AE (2001) Purification and characterization of N-glycosylation mutant mouse and human P-glycoproteins expressed in Pichia pastoris cells. Arch Biochem Biophys 388: 171–177.

14. Wang Z, Stalcup LD, Harvey BJ, Weber J, Chloupkova M, et al. (2006) Purification and ATP Hydrolysis of the Putative Cholesterol Transporters ABCG5 and ABCG8{. Biochemistry 45: 9929–9939.

15. Chloupková M, Pickert A, Lee J-Y, Souza S, Trinh YT, et al. (2007) Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 46: 7992–8003.

16. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 11: 1156–1166.

17. Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, et al. (2009) The Human Gene Mutation Database: 2008 update. Genome Med 1: 13.

18. Braman J, Papworth C, Greener A (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods Mol Biol 57: 31–44.

19. Feher Z, Kiss A, Venetianer P (1983) Expression of a bacterial modification methylase gene in yeast. Nature 302: 266–268.

20. Proffitt JH, Davie JR, Swinton D, Hattman S (1984) 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol 4: 985–988.

21. Hattman S, Kenny C, Berger L, Pratt K (1978) Comparative study of DNA methylation in three unicellular eucaryotes. J Bacteriol 135: 1156–1157.

22. Gietz RD, Woods RA (2001) Genetic transformation of yeast. Bio Techniques 30: 816–820, 822-816828 passim.

23. Ito H, Fukuda Y, Murata K, Kimura A (1983) Transformation of intact yeast cells treated with alkali cations. J Bacteriol 153: 163–168.

24. Schiestl RH, Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. Curr Genet 16: 339–346.

25. Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol 8: 91.

26. QuikChange II XL. Site-Directed Mutagenesis Kit Instruction Manual Revision D Agilent Technologies Inc.

27. Tate CG (2001) Overexpression of mammalian integral membrane proteins for structural studies. FEBS Letters 504: 94–98.

28. Keitel V, Burdelski M, Warskulat U, Kühlkamp T, Keppler D, et al. (2005) Expression and localization of hepatobiliary transport proteins in progressive familial intrahepatic cholestasis. Hepatology 41: 1160–1172.

29. Keitel V, Burdelski M, Vojnisek Z, Schmitt L, Häussinger D, et al. (2009) De novo bile salt transporter antibodies as a possible cause of recurrent graft failure after liver transplantation: A novel mechanism of cholestasis. Hepatology 50: 510–517.

30. Singh MV, Weil PA (2002) A method for plasmid purification directly from yeast. Anal Biochem 307: 13–17.

31. Aslanidis C, de Jong PJ (1990) Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res 18: 6069–6074.

32. Geertsma ER, Poolman B (2007) High-throughput cloning and expression in recalcitrant bacteria. Nat Methods 4: 705–707.

33. Lee SH, Altenberg GA (2003) Expression of functional multidrug-resistance protein 1 in Saccharomyces cerevisiae: effects of N- and C-terminal affinity tags. Biochem Biophys Res Commun 306: 644–649.

34. Ernst R, Kueppers P, Klein CM, Schwarzmueller T, Kuchler K, et al. (2008) A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. Proc Natl Acad Sci USA 105: 5069–5074.

35. Kaiser C, Michaelis S, Mitchell A (1994) Methods in yeast genetics - a laboratory course manual. Cold Spring Harbor, NY.: Cold Spring Harbor Laboratory Press.

36. Mamnun YM, Schüller C, Kuchler K (2004) Expression regulation of the yeast PDR5 ATP-binding cassette (ABC) transporter suggests a role in cellular detoxification during the exponential growth phase. FEBS Lett 559: 111–117.

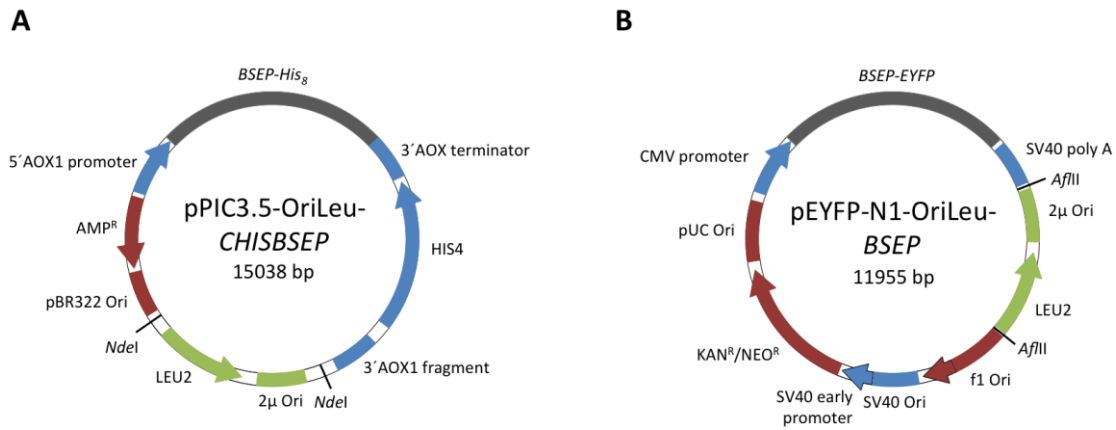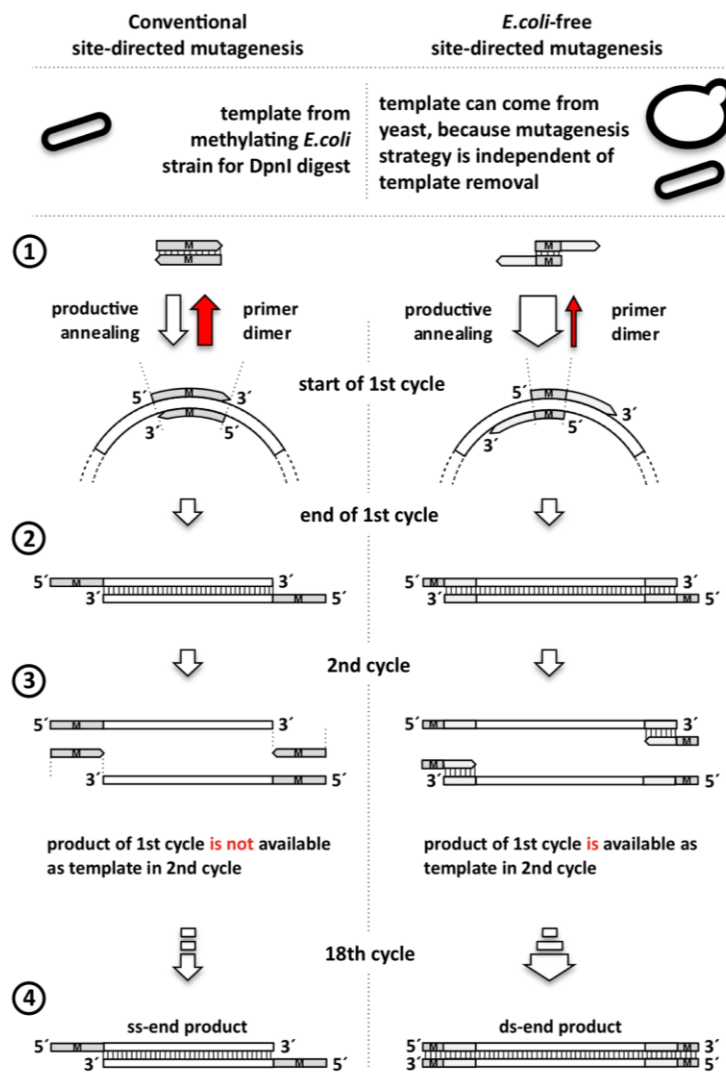# Supporting Information SI

Fig. S1

A

B

Fig. S2

Fig. S3

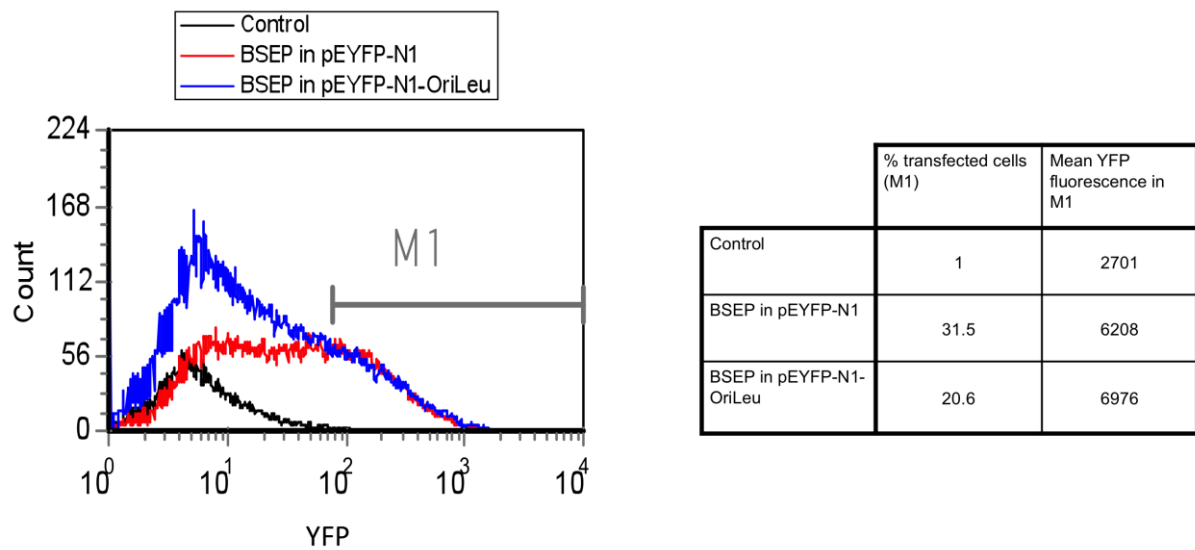|	% transfected cells (M1)	Mean YFP fluorescence in M1
Control	1	2701
BSEP in pEYFP-N1	31.5	6208
BSEP in pEYFP-N1-OriLeu	20.6	6976

Table S1: PCR primers used in this study

Oligonucleotide	Sequence 5´-3´ (relevant restriction sites in underlined italics and point mutations indicated in bold letter)	Used in
YEpNHISFor	GATCCTTTAATTATCAAACAATATCAATATGCATCATCACCATCATCACCATCATCACCATCATCATCACCATGGTGGTGGTATTGAAGGTAGACCCGGGTAGA	Supplementary figure 1
YEpNHISRev	CGCGTCTACCCGGGTCTACCTTCAATACCACCACCATGGTGATGATGATGGTGATGATGGTGATGATGGTGATGATGCATATTGATATTGTTTGATAATTAAAG	Supplementary figure 1
BSEP-YEpHISN-S1	TCACCATCATCATCACCATGGTGGTGGTATTGAAGGTAGATCTGACTCAGTAATTCTTCGAAGTATAAAG	Supplementary figure 1
BSEP-YEpHISN-S2	GAATAAGGTAAACATGGTAGCGATGTCGACCTCGAGACGCGTCTAACTGATGGGGGATCCAGTGGTGACT	Supplementary figure 1
BSEP-YEpHISC-S1	ATAAGAAGATAGGATCCTTTAATTATCAAACAATATCAATATGTCTGACTCAGTAATTCTTCGAAGTAT	Supplementary figure 1
BSEP-YEpHISC-S2	CGATGTCGACCTCGAGACGCGTCTAATGGTGATGGTGATGGTGATGGTGACCACTGATGGGGGATCCAGTGGTGACT	Supplementary figure 1
OriLeu-pPIC3.5-*Nde*I-S1	AACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCAC*CATATG*CGAGGCCCTTTCGTCTTCAAGAATTAACTGTGGGA	Figure 1
OriLeu- pPIC3.5-*Nde*I-S2	GTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCG*CATATG*ATCTGTGCGGTATTTCACACCGCATATATCG	Figure 1
BSEP-pPIC3.5-S1	AATTATTCGAAGGATCCTACGTAGAATTCCCTAGGGCGGCCGCATGTCTGACTCAGTAATTCTTCGAAGTATAAAGAAAT	Figure 1
BSEP-pPIC3.5-S2	TGAGGAACAGTCATGTCTAAGGCGAATTAATTCGCGGCCGCCTAATGGTGATGGTGATGGTGACCACTGATG	Figure 1
BSEP-*Bst*BImut-S1	AGCTC*TTCGAA*GAGCCTTCTCTTACACCCCAAGTTATGCAAAAGCTAAA	Figures 2 and 3
BSEP-*Bst*BImut-S2	AGAAGGCTC*TTCGAA*GAGCTGTTGCACTCAGTACAACTGCAGAGATCAC	Figures 2 and 3
OriLeu-pEYFP-*Afl*II-S1	TCTAGTTGTGGTTTGTCCAAACTCATCAATGTAT*CTTAAG*CGAGGCCCTTTCGTCTTCAAGAATTAACTGTGGGA	Figure 3
OriLeu-pEYFP-*Afl*II-S2	ATTTTAACAAAATATTAACGCTTACAATTTACGC*CTTAAG*ATCTGTGCGGTATTTCACACCGCATATATCG	Figure 3

Table S2: Plasmids used in this study

Plasmid	Purpose	Reference
YEpHIS and YEpMDR1HIS	constitutive; C-terminal his_8 tag. YEpMDR1HIS contains human *MDR1* cDNA	[67]
YEpNHIS	constitutive; N-terminal his_{14} tag	This study
YEpNHIS-*BSEP*	constitutive BSEP expression with N-terminal his_{14} tag	This study
YEpHIS-*BSEP*	constitutive BSEP expression with C-terminal his_8 tag	This study
pPIC3.5-OriLeu-*CHISBSEP*	yeast-shuttle, *P. pastoris* integration vector for chromosomal BSEP expression with C-terminal his_8 tag	This study
pEYFP-N1-*BSEP*	mammalian BSEP expression plasmid	[68]
pEYFP-N1-OriLeu-*BSEP*	yeast-shuttle, mammalian BSEP expression plasmid	This study

Chapter 5

Detergent Screening and Purification of the Human Liver ABC Transporters BSEP (ABCB11) and MDR3 (ABCB4) expressed in the yeast *Pichia pastoris*

Published in: *PLOS ONE*

Impact factor: 4.092

Own Proportion

 to this work: 60 %

                    Expression of BSEP in *P. pastoris*

                    Detergent screen with dot blot and FSEC

                    Purification of BSEP and binding studies

                    Writing of the manuscript

PLOS | ONE

Detergent Screening and Purification of the Human Liver ABC Transporters BSEP (ABCB11) and MDR3 (ABCB4) Expressed in the Yeast *Pichia pastoris*

Philipp Ellinger[9], **Marianne Kluth**[9], **Jan Stindt**[¤], **Sander H. J. Smits, Lutz Schmitt***

Institute of Biochemistry, Heinrich Heine University, Düsseldorf, Germany

Abstract

The human liver ATP-binding cassette (ABC) transporters bile salt export pump (BSEP/ABCB11) and the multidrug resistance protein 3 (MDR3/ABCB4) fulfill the translocation of bile salts and phosphatidylcholine across the apical membrane of hepatocytes. In concert with ABCG5/G8, these two transporters are responsible for the formation of bile and mutations within these transporters can lead to severe hereditary diseases. In this study, we report the heterologous overexpression and purification of human BSEP and MDR3 as well as the expression of the corresponding C-terminal GFP-fusion proteins in the yeast *Pichia pastoris*. Confocal laser scanning microscopy revealed that BSEP-GFP and MDR3-GFP are localized in the plasma membrane of *P. pastoris*. Furthermore, we demonstrate the first purification of human BSEP and MDR3 yielding ~1 mg and ~6 mg per 100 g of wet cell weight, respectively. By screening over 100 detergents using a dot blot technique, we found that only zwitterionic, lipid-like detergents such as Fos-cholines or Cyclofos were able to extract both transporters in sufficient amounts for subsequent functional analysis. For MDR3, fluorescence-detection size exclusion chromatography (FSEC) screens revealed that increasing the acyl chain length of Fos-Cholines improved monodispersity. BSEP purified in n-dodecyl-β-D-maltoside or Cymal-5 after solubilization with Fos-choline 16 from *P. pastoris* membranes showed binding to ATP-agarose. Furthermore, detergent-solubilized and purified MDR3 showed a substrate-inducible ATPase activity upon addition of phosphatidylcholine lipids. These results form the basis for further biochemical analysis of human BSEP and MDR3 to elucidate the function of these clinically relevant ABC transporters.

Introduction

ATP-binding cassette (ABC) transporters constitute one of the largest families of membrane transport proteins present in all three kingdoms of life. They transport a wide variety of different substrates ranging from small ions to large proteins across biological membranes using ATP as energy source [1,2]. ABC transporters are composed of two transmembrane domains (TMDs) and two highly conserved nucleotide-binding domains (NBDs). TMDs determine the substrate specificity and the NBDs fuel the transport by binding and hydrolyzing ATP. In eukaryotes, the TMDs and NBDs are encoded on one gene and build up either a full-size transporter (one gene encoding two TMDs and two NBDs) or a half-size transporter (one gene encoding one TMD and one NBD), which hetero- or homodimerize to form the functional unit.

Within the human genome 48 genes encode for ABC proteins, which are involved mainly in transport [3]. Mutations in these ABC protein genes can lead to severe diseases such as cystic fibrosis, X-linked Adrenoleukodystrophy or Tangier disease. Beside this, ABC transporters are also involved in processes like

multidrug resistance of cancer cells [4,5,6,7]. In hepatocytes, eleven ABC transporters are expressed. Except for the transport of different cyclic nucleotides, glucuronide and glutathione conjugates through MRPs (MRP 1–6, note that MRP1 is detected only in fetal hepatocytes) [8] and the transport of endo- and xenobiotics by MDR1 (P-gp) [9] and ABCG2 [10], one of the main function of ABC transporters in the liver is the formation of bile depending on the ABC transporters BSEP (ABCB11), MDR3 (ABCB4) and ABCG5/8 [11]. Bile is essential for the digestion of fat as well as for the absorption of lipids and fat-soluble vitamins originating from food ingestion in the small intestine. In the intestine the main components of bile, bile salts and phosphatidylcholine are recycled via the enterohepatic circulation [12]. Bile salts, phosphatidylcholine and cholesterol form mixed micelles in the canaliculus, which dampen the detergent effect of the amphiphatic bile salts as well as prevent the formation of cholesterol crystals. Bile formation is dependent on the three ABC transporters BSEP (ABCB11), MDR3 (ABCB4) and ABCG5/8 [13].

The bile salt export pump (BSEP) is the main bile salt transporter in humans and is localized in the apical membrane of hepatocytes [14]. It is a 1321 amino acid large, glycosylated

110

full-size ABC transporter and mediates the ATP-dependent bile flow by transporting monovalent bile salts like taurine and glycine conjugates of primary and secondary bile salts (e.g. tauro- and glycocholate or taurodeoxycholate) into the canaliculus [15]. The human multidrug resistance protein 3 (MDR3) is a close homologue of MDR1 (P-glycoprotein, ABCB1) with an amino acid sequence identity of nearly 80%. However, MDR3 exclusively translocates phosphatidylcholine from the inner to the outer leaflet of the apical membrane [16]. MDR3 is like BSEP a glycosylated full-size transporter composed of 1288 amino acid [17]. The heterodimeric ABC transporter ABCG5/G8 completes the bile forming machinery by transporting cholesterol [18,19].

Mutations within the *BSEP* and *MDR3* gene can lead to different cholestatic diseases, e.g. progressive familiar intrahepatic cholestasis type 2 and 3 (PFIC2 and PFIC3) [20,21,22], benign recurrent intrahepatic cholestasis type 2 (BRIC2) [23] or intrahepatic cholestasis of pregnancy (ICP) [24] and low-phospholipid associated cholestasis (LPAC) [25]. Therapy for cholestatic disease includes treatment with e.g. ursodeoxycholic acid or surgical biliary diversion [26]. If none of those treatments is successful, the only alternative therapy is liver transplantation. New successful forms of therapy include treatment with chemical chaperones like 4-phenylbutyrate for misfolded BSEP mutants [27].

Because of their high clinical interest, MDR3 and especially BSEP have been characterized extensively in cell culture as well as animal models [22,28,29,30,31,32]. A well-established system for investigating BSEP are for example insect cell-based vesicles, which allow to perform transport studies and to study kinetics, inhibitors or mutants [13,33]. Less is known about MDR3, because of the difficulty to establish a robust activity assay. Together, all these assays are performed in whole cells or membranes and not with the isolated proteins.

To investigate the function of BSEP and MDR3 in its isolated form, a substantial expression of these proteins is required. To date, no reports regarding the purification of both proteins from cell culture systems or other expression systems have been reported. An alternative to cell culture is the use of yeast expression systems such as *Saccharomyces cerevisiae* or *Pichia pastoris*, which also harbor the eukaryotic protein processing machinery and can be grown to high cell densities. Chloupková *et al.* tested 25 human ABC transporters for expression in *P. pastoris* [34], but BSEP and MDR3 were not included in this study, while for example MRP2, another human liver ABC transporter, could not be expressed.

In general, *S. cerevisiae* has been used frequently to express eukaryotic membrane proteins [35]. After successful establishment of an expression system, the purification of a membrane protein requires first of all its solubilization with detergents from the membrane of the expression host. However, finding an adequate detergent for extraction and purification that preserves the membrane protein in a stable and functional form is an empirical process. High throughput methods have been developed in order to screen the influence of detergents on stability and monodispersity of the purified membrane protein [36,37,38,39,40,41,42]. One of these approaches is fluorescence-detection size exclusion chromatography (FSEC) based on the fluorescence of a green fluorescent protein (GFP) tag fused to the membrane protein. In this approach solubilized crude membranes are loaded on a size exclusion column and the elution is monitored via the fluorescence of GFP. Thereby only the membrane-GFP fusion protein is visible and the result can be evaluated based on the shape of the elution peak [43,44]

In this study, we established the heterologous overexpression in the yeast *P. pastoris* and the subsequent solubilization and purification of human BSEP and MDR3. To achieve this, we applied a dot blot technique and FSEC to identify the most suitable detergent for BSEP and MDR3. The purified protein could be isolated in a functional state as judged by substrate-induced ATPase activity of MDR3 and ATP binding in the case of BSEP.

Materials and Methods

Materials
All detergents were obtained from Affymetrix with the exception of Digitonin, which was purchased from Sigma. Lipids were from Sigma or Avanti Polar Lipids.

Routine Procedures
SDS-PAGE on 7% gels used the Bio-Rad Minigel system. Immunoblotting followed standard procedures using the monoclonal anti-P-gp C219 antibody in case of MDR3 (Abcam), the F-6 anti-BSEP antibody (Santa Cruz Biotechnology) or an anti-GFP antibody (Sigma). Protein concentration was estimated by the Bradford method using a Coomassie Plus Assay (Pierce).

Cloning of human BSEP and MDR3 and GFP fusion expression constructs for *Pichia pastoris*
The general cloning procedure is described in detail in Stindt *et al.* [45]. The *P. pastoris* expression vector pSGP18 was made compatible for *Saccharomyces cerevisiae* by introducing a 2μ origin of replication into its backbone. The 2μ origin of replication was PCR-amplified from the YEpHIS vector with the primer pairs 2μ for pPIC S1 and 2μ for pPIC S2 (for oligonucleotide sequences see Table 1). The resulting PCR product and the pSGP18 vector were digested with *Pci*I and ligated yielding pSGP18-2μ. The coding sequences for human BSEP and MDR3 (NCBI accession code: NM_003742.2 and NM_000443.3) were PCR-amplified with the primer pairs BSEP-HR-PP-S1 and BSEP-HR-PP-S2 and MDR3-HR-PP-S1 and MDR3-HR-PP-S2, respectively. For *Pichia* expression of the GFP-tagged transporters, the respective coding sequences were amplified either with the primer pair BSEP-PP-HR-S1 and YEpN14HIS-BSEP-S2 or with MDR3-PP-HR-S1 and YEpN14HIS-MDR3-S2. The S65T-GFP sequence of pFA6a-GFP(S65T)-kanMX6 [27] was either amplified with primer pair GFP-BSEP-HR-S1 and GFP-PP-HR-S2 or primer pair GFP-MDR3-HR-S1 and GFP-PP-HR-S2. This includes the necessary homologous overlaps to the PCR products for in-frame recombination into pSGP18-2μ. pSGP18 contains a 3C protease cleavage site, a calmodulin binding peptide (CBP) tag and a RGS-6xhis-tag C-terminal to the proteins in the multiple cloning site [34]. For expression of the GFP-fusion proteins, tags were replaced by GFP in the process of recombination. The *Bsm*BI linearized pSGP18-2μ vector and the PCR fragments were gel-purified, mixed in equimolar amounts (either with BSEP or MDR3 or each together with GFP) and transformed into *S. cerevisiae* [45]. The ATP hydrolysis deficient mutant of MDR3 was generated by introduction of two point mutations in the conserved NBD. Therefore, we replaced Glu 558 and Glu 1207 of the Walker B motif to Gln using the QuikChange® XL Site-Directed Mutagenesis Kit (Agilent Technologies). The sequence of all constructs were verified by DNA sequencing.

Transformation of *P. pastoris*
BSEP and *MDR3* expression constructs were transformed into competent *P. pastoris* X33 (Invitrogen) cells using standard

Table 1. PCR oligonucleotides used in this study.

Oligonucleotide	Sequence 5′→ 3′
pSGP18-2μ-ori-S1	TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTAAATATTGCGAATACCGCTTCCACAAACATTG
pSGP18-2μ-ori-S2	AACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTATTTCACACCGCATATATCGGATCGTACT
BSEP-HR-PP-S1	ATCAAAAAACAACTAATTATTCGAACGAGGTAAAAGAATGTCTGACTCAGTAATTCTTCGAAGT ATA
BSEP-HR-PP-S2	ACGTTTGGACCTTGGAAAAGACTTCTAAGGAGTTGGAGGCACTGATGGGGGATCCAGTGGTGACTAGTTT
MDR3-HR-PP-S1	ATCAAAAAACAACTAATTATTCGAACGAGGTAAAAGAATGGATCTTGAGGCGGCAAAGAACGGAACA
MDR3-HR-PP-S2	ACGTTTGGACCTTGGAATAAGACTTCTAAGGAGTTGGAGGCTAAGTTCTGTGTTCCAGCCTGGACACTGACCATTGAAAAATAG
YEpN14HIS-BSEP-S2	GAATAAGGTAAACATGGTAGCGATGTCGACCTCGAGACGCGTCTAACTGATGGGGGATCCAGTGGTGACT
YEpN14HIS-MDR3-S2	GAATAAGGTAAACATGGTAGCGATGTCGACCTCGAGACGCGTCTATAAGTTCTGTGTCCCAGCCTGGACACTGACCATT
GFP-BSEP-HR-S1	AGCCTACTACAAACTAGTCACCACTGGATCCCCCATCAGTGGTGGTGGTCGACGGATCCCCGGGTTA
GFP-PP-HR-S2	ACGTTTGGACCTTGGAATAAGACTTCTAAGGAGTTGGAGGCTATTATTTGTATAGTTCATCCATGCCATGT
GFP-MDR3-HR-S1	TTTCAATGGTCAGTGTCCAGGCTGGAACAAAGAGACAAGGTGGTGGTCGACGGATCCCCGGGTTA
MDR3-E558Q S1	GATCCTTCTGCTGGATCAAGCCACGTCAGCATTGGACAC
MDR3-E558Q S2	GTGTCCAATGCTGACGTGGCTTGATCCAGCAGAAGGATC
MDR3-E1207Q S1	CAAATCCTCCTGTTGGATCAAGCTACATCAGCTCTGGATAC
MDR3-E1207Q S2	GTATCCAGAGCTGATGTAGCTTGATCCAACAGGAGGATTTG

doi:10.1371/journal.pone.0060620.t001

procedures (Invitrogen). Briefly, 10-20 μg DNA of the expression construct were linearized using PmeI (New England Biolabs) to facilitate homologous recombination at the AOX1 locus, extracted by phenol/chloroform, re-suspended in 10 μl sterile H_2O and transformed into 80 μl electro-competent P. pastoris cells by electroporation (1.500 V, 5 ms). Cells were incubated in 1 M sorbitol without shaking for 1 h at 30°C, 1 ml YPD was subsequently added and cells were shaken for 2 h at 200 rpm and 30°C. 100 μl of this suspension was plated onto YPDS plates containing 100 μg/ml Zeocin or higher and incubated for 30°C until colonies appeared. 10 to 20 colonies were re-streaked on YPD plates containing Zeocin and used for expression studies.

Expression screening of BSEP and MDR3 transformed P. pastoris cells

Small-scale expression screens of BSEP or MDR3 P. pastoris clones were performed similarly as described by Wang et al. [46]. 50 ml cultures were grown overnight in MGY medium (1.34% (w/v) yeast nitrogen base, 1% (v/v) glycerol and 4×10^{-5}% (w/v) biotin) at 30°C and 220 rpm, harvested by centrifugation, re-suspended in 50 ml MMY (1.34% (w/v) yeast nitrogen base, 0.5% (v/v) methanol and 4×10^{-5}% (w/v) biotin) and incubated for another 24 h to induce protein expression. 2 ml of these cells were harvested, washed in 2 ml of homogenization buffer (50 mM Tris-HCl, pH 8.0, 0.33 M sucrose, 75 mM NaCl, 1 mM EDTA, 1 mM EGTA, 100 mM 6-Aminocaproic acid, 2 mM β-Mercaptoethanol) supplemented with protein inhibitor cocktail (Roche) and re-suspended in 500 μl of homogenization buffer. Cells were lysed with 1 ml of acid-washed zirconia beads (Roth) by vortexing 6 times for 1 min with 1 min breaks on ice. Disrupted cells were centrifuged for 5 min, 12.000 xg, 4°C and the supernatant was adjusted to 10 mM $MgCl_2$ and incubated on ice for 15 min. Precipitated membranes were harvested by centrifugation for 30 min, 20.000 xg, 4°C and the resulting pellet was re-suspended in SDS sample buffer and loaded onto a 7% SDS-PAGE. Expression was visualized by immuno blotting.

Fermentation of BSEP and MDR3

For large-scale expression, BSEP, BSEP-GFP and MDR3 expressing clones were fermented in a 15 liter table-top glass fermentor (Applikon Biotechnology) according to the Invitrogen Pichia fermentation guidelines [26] using the basal salt media. Typically a volume of 6 l media was inoculated with 1 l of an overnight culture grown in MGY (1.34% yeast nitrogen base, 1% glycerol and 4×10^{-5}% biotin) media. Aeration was kept above 20% O_2 saturation and the glycerol fed-batch was performed for 5 h feeding ~500 ml of 50% (v/v) glycerol. Protein expression was induced by addition of 3.6 ml/h l (~1000 ml) methanol for 48 h. Cells were harvested by centrifugation (5.000 xg, 10 min, 4°C), flash-frozen in liquid nitrogen and stored at −80°C until further use. Under these conditions approximately 1–1.4 kg of wet cell mass could be obtained.

Expression of GFP fusion proteins in shaking flask cultures

Clones either expressing BSEP-GFP or MDR3-GFP were inoculated in 2 l shaking flasks containing 0.5 l of MGY media and shaken overnight at 30°C and 220 rpm. Protein expression was induced with methanol by harvesting the cells in sterile centrifuge buckets (5.000 xg, 10 min, 4°C) and re-suspended in 0.5 l methanol-containing media (MMY). 24 h after induction, methanol was added to a final concentration of 0.5% and after 48 h the cells were harvested (5.000 xg, 10 min, 4°C), flash-frozen in liquid nitrogen and stored at −80°C until further usage.

Confocal fluorescence microscopy of GFP fusion proteins

P. pastoris cells expressing either BSEP-GFP or MDR3-GFP were directly spotted onto microscope slides coated with poly-L-lysine (Thermo Scientific) from shaking flasks and mounted with a coverslip. Images were acquired using an Olympus FV1000 confocal laser scanning microscope equipped with a 60× UPLSAPO objective (N.A. 1.35). GFP was excited at 488 nm and emission was recorded at 500 nm–600 nm.

112

Preparation of crude membrane vesicles for protein purification

100 g batches of *P. pastoris* cells expressing BSEP or MDR3 were thawed on ice, washed with ddH$_2$O and re-suspended at a concentration of 0.5 g cells/ml in homogenization buffer containing protease inhibitor cocktail (Roche). Cells were disrupted by two passages through a pre-cooled TS Series Cell Disrupter (Constant Systems) at 2.5 kbar. After cell debris was spun down by two centrifugation steps (15 min at 5,000 xg, 4°C and 30 min at 15,000 xg, 4°C), crude membrane vesicles were prepared by ultracentrifugation for 1 h at 125,000 xg, 4°C. Membrane vesicles were re-suspended in buffer A (50 mM Tris-HCl pH 8.0, 75 mM NaCl, 30% (v/v) glycerol) and flash frozen in liquid N$_2$.

Solubilization screen via the Dot Blot technique

Membranes were thawed on ice and solubilized in 200 µl buffer A. Membrane concentration was kept at 5 mg/ml during solubilization and detergents were used at a concentration of 1% (w/v) or higher according to their critical micellar concentration (cmc). A complete list of the used detergents is provided in Table S1 in File Supplementary Information. Samples were solubilized for 1 h at 4°C on a rotator, centrifuged (100.000 xg, 30 min, 4°C) and the supernatant was supplemented with SDS sample buffer. The samples were heated to 65°C for 10 min and 3 µl were spotted onto a dry nitrocellulose membrane. After extensive drying of the sample, the membrane was blocked for 1 h in TBS-T with 5% (w/v) milk powder and then probed with a 1:2000 dilution of the respective primary antibody. Dot blots were quantified using the GeneTools software (Syngene).

Fluorescence-detection size-exclusion chromatography (FSEC)

BSEP-GFP or MDR3-GFP containing membranes were solubilized in detergents based on the results of the dot blot analysis. 100 µl of the solubilized sample was applied to a Biosep SEC-S4000 size-exclusion chromatography column (Phenomenex) connected to a HPLC system (Hitachi) equipped with a fluorescence detector (L-2485, Hitachi), which was equilibrated in running buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 15% (v/v) glycerol and 0.02% (w/v) β-DDM). The UV absorption of the proteins was followed at 280 nm and for online fluorescence detection, the GFP tag was excited at $\lambda_{ex} = 470$ nm to improve the signal to noise ratio and fluorescence emission was detected at $\lambda_{em} = 512$ nm.

Solubilization and Purification of MDR3 and BSEP

The purification of MDR3 and BSEP was performed by tandem-affinity purification (TAP) consisting of an immobilized metal ion affinity chromatography (IMAC) step followed by a calmodulin binding peptide affinity purification (CBP). All procedures were carried out at 4°C. Crude membrane vesicles equivalent to 100 g wet cells were thawed at 4°C, diluted to a final concentration of 5 mg/ml total protein with buffer A as determined by the Coomassie Plus Assay (Pierce) and solubilized in 1% (w/v) of Fos-choline-16 or other detergents for 1 h at 4°C (for cmc values see Table S1 in File Supplementary Information). Non-solubilized membrane vesicles were removed by centrifugation at 100.000 xg, 4°C for 1 h. The supernatant supplemented with 20 mM imidazole was loaded onto a Ni^{2+}-loaded HiTrap Chelating column (5 ml, GE Healthcare) and washed with 10 column volumes of buffer A supplemented with 20mM imidazole and typically 2.5×cmc of detergent. Proteins were eluted in one step with buffer B (50 mM Tris-HCl pH 8.0, 75 mM NaCl,

200 mM imidazole, 20% (v/v) glycerol) supplemented with 2.5×cmc detergent. The IMAC eluate was diluted 5-times with CaCl$_2$ binding buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM MgCl$_2$, 2 mM CaCl$_2$ and 20% (v/v) glycerol) containing 2.5×cmc detergent, applied to 4 ml calmodulin affinity resin equilibrated in CaCl$_2$ binding buffer and incubated with the calmodulin resin over night at 4°C on a rotator. The resin was transferred into a gravity flow column and washed with 10 column volumes of CaCl$_2$ binding buffer containing 2.5×cmc detergent. The proteins were eluted with 3 bed volumes of EGTA elution buffer (2 mM EGTA, 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and 20% (v/v) glycerol) supplemented with 2.5×cmc detergent. The purified protein was directly used for ATPase activity or further concentrated using an Amicon Ultra-15 filter (Millipore) with a cut-off of 100 kDa, aliquoted, snap frozen in liquid nitrogen and stored at −80°C. Aliquots of the sample were analyzed by Coomassie blue stained SDS-PAGE and immunoblotting.

ATP Agarose binding assay of BSEP

To test the ability of detergent solubilized BSEP to bind ATP, 25 µl of a 1:1 slurry of C8-linked ATP-agarose resin (Sigma) equilibrated in buffer A was added to 20 µg of purified BSEP in the detergent to be examined and incubated at 4°C on a rotator. After 1 h, the resin was pelleted by centrifugation (8200 xg, 2 min, 4°C) and the resin was washed three more times with 250 µl of buffer A supplemented with 2.5×cmc of the detergent. Bound proteins were eluted in SDS sample buffer by heating the resin to 65°C for 20 min. The pellet samples were subjected to SDS-PAGE and analyzed by immunoblotting.

ATPase activity measurements of MDR3

The ATPase activity of MDR3 was examined with the malachite green assay by determination of released free inorganic orthophosphate as described previously [47]. Reactions were performed in a total volume of 100 µl in buffer C (50 mM Tris-HCl pH 7.4, 50 mM NaCl, 15% (v/v) glycerol) containing 2.5×cmc detergent and 10 mM MgCl$_2$. 5 – 20 µg purified, detergent-soluble MDR3 was used. The reaction was started by typically adding 2 mM ATP at 37°C and stopped at appropriate time points by the addition of 25 µl of the reaction into 175 µl of 20 mM ice-cold H$_2$SO$_4$. Subsequently, 50 µl dye solution (0.096% (w/v) malachite green, 1.48% (w/v) ammonium molybdate, and 0.173% (w/v) Tween-20 in 2.36 M H$_2$SO$_4$) was added. After 15 min the amount of free phosphate was quantified spectroscopically by measuring the absorption at 595 nm. For subsequent data evaluation, all appropriate controls were performed and subtracted. For calibration of free phosphate concentrations a Na$_2$HPO$_4$ standard curve was used. For substrate stimulated ATPase activity, purified MDR3 was incubated with the equal volume of 2–5 mM lipid stock solution at room temperature for 20 min and sonified for 30 s to facilitate the incorporation of lipids into the detergent-protein micelles. The lipid-protein sample was stored on ice until further usage.

Results

Cloning and Expression of human BSEP and MDR3 in *P. pastoris*

For the expression of human BSEP and MDR3 in the methylotrophic yeast *P. pastoris* we used the expression plasmid pSGP18, which was used before to express 25 human ABC transporters in *P. pastoris* [34]. BSEP and MDR3 were not included in this study likely due to the inherent toxicity of the cDNAs, which hampers the cloning procedure and often results in

113

the failure of obtaining suitable expression plasmids [13]. We custom modified the plasmid by introducing a 24 origin of replication for *S. cerevisiae* in its backbone and cloned the human *BSEP* and *MDR3* cDNA via homologous recombination into pSGP18-2μ. After transformation in *P. pastoris*, ten clones were tested for expression. A clone for each transporter was chosen for fermentation, which yielded about 1.0–1.4 kg of wet cell weight (wcw) in a typical fermentation. As can be seen by immunoblotting both wild-type proteins were expressed in *P. pastoris* (Fig. 1A and C, middle lanes). The wild-type proteins exhibited a distinct protein band at ~130 kDa that cross-reacted with monoclonal antibodies against BSEP or MDR3. No signal was obtained using the empty plasmid as a control (see Fig. 1, neg ctrl).

Localization and judging the quality of BSEP and MDR3 in *P. pastoris*

For the determination of the trafficking and localization of human BSEP and MDR3 in *P. pastoris* cells, we generated and expressed the corresponding GFP-fusion proteins, BSEP-GFP and MDR3-GFP. The C-terminal GFP-tag was confirmed by immunoblot analysis against GFP (Fig. 1B and D) as well as by a shift to a higher molecular weight visualized by antibodies against BSEP and MDR3, respectively (Fig. 1A and C, right lane). Both the fusion proteins migrated at ~160 kDa. The correct trafficking of the GFP-fusion proteins to the plasma membrane of *P. pastoris* was checked by confocal laser scanning microscopy (Fig. 2, upper row). Induced cells expressing BSEP-GFP or MDR3-GFP showed clear ring-shaped fluorescence at the plasma membrane, which co-localized with the cell surrounding of the differential interference contrast (DIC) scan (Fig. 2, bottom row, merged pictures). As control only GFP was expressed in *P. pastoris* and the fluorescence

was distributed homogenously within the cell, which leads to the conclusion that BSEP and MDR3 are processed and trafficked correctly in *P. pastoris*. Non-induced cells did not show any fluorescence (data not shown). We also employed sucrose density centrifugation of whole cell membranes containing BSEP or MDR3, which demonstrated co-localization of a plasma membrane marker with BSEP or MDR3, respectively (data not shown).

Solubilization Screen via the Dot Blot technique

To find an appropriate detergent for membrane protein extraction, we tested over 100 different detergents for their ability to solubilize BSEP and MDR3 via dot blot analysis. These detergents covered all four classes: non-ionic (N), anionic (A), cationic (C) as well as zwitterionic (Z) (Table S1 in File Supplementary Information). Most of the detergents were used at a concentration of 1% (w/v). However, depending on the critical micellar concentration (cmc) other concentrations were also chosen when necessary (see Table S1 in File Supplementary Information).

Membranes were solubilized for 1 h at 4°C, subsequently centrifuged and the supernatant was spotted on the dot blot membrane. For BSEP, we tested solubilization of the wild-type protein as well as the GFP-fusion protein, to investigate if the GFP-tag had any influence on the solubilization. Therefore, the BSEP-GFP fusion protein was fermented the same way as the wild-type BSEP protein for comparison. As seen in Fig. 3A and 3B, BSEP-GFP could be extracted more efficiently than BSEP by maltosides and glucosides (D-I 1-5). Furthermore, some differences can be seen in a more efficient extraction of BSEP-GFP in Fos-choline-unisat-11-10 and Fos-choline-8 (G8 and G9). Despite this, there are large similarities between BSEP and BSEP-GFP, in fact only
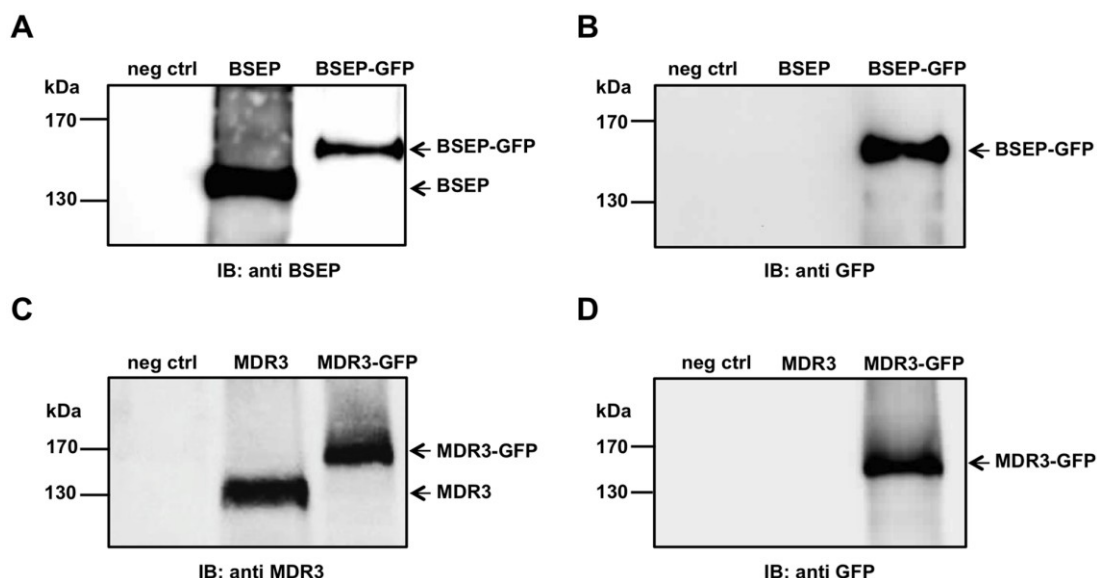

Figure 1. Human BSEP and MDR3 expression in *Pichia pastoris*. A 5 μg of membranes derived from *P. pastoris* cells carrying the empty expression plasmid pSGP18 (neg ctrl), BSEP or BSEP-GFP were subjected to SDS-PAGE and immunoblotting (lanes from left to right). The negative control (left lane) did not react with the monoclonal antibody (F-6), while BSEP (middle lane) and BSEP-GFP (right lane) could be detected by the same antibody. **B** Identical samples were probed with a monoclonal GFP antibody. The negative control (left lane) as well as BSEP (middle lane) showed no signal with anti-GFP antibody, while BSEP-GFP could be detected (right lane). **C** In case of MDR3 the negative control (left lane) showed no signal with the monoclonal antibody C219; MDR3 (middle lane) as well as MDR3-GFP (right lane) could be detected with the monoclonal antibody C219. **D** Identical MDR3 samples were probed with a monoclonal GFP antibody. The negative control (left lane) as well as MDR3 (middle lane) showed no signal with anti-GFP antibody, while MDR3-GFP could be detected (right lane). The position of the molecular weight markers are shown on the left.
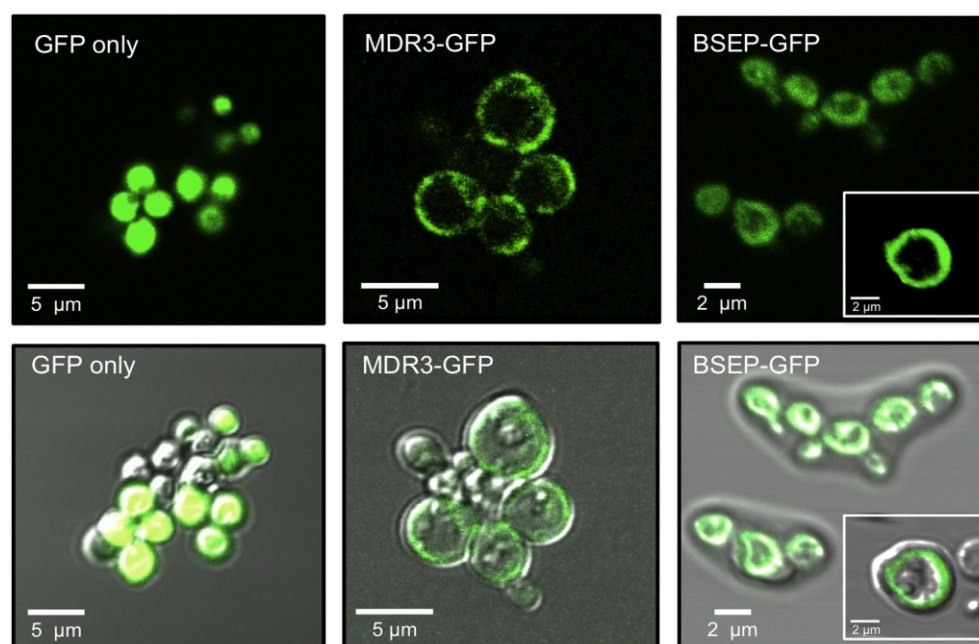doi:10.1371/journal.pone.0060620.g001

114

Figure 2. Fluorescence microscopy of BSEP-GFP and MDR3-GFP expressing *Pichia pastoris* **cells.** *P. pastoris* cells expressing GFP, BSEP-GFP or MDR3-GFP were harvested 48 h after induction and examined for GFP fluorescence (upper row) by confocal LSM. BSEP-GFP as well as MDR-GFP was located in the plasma membrane of *P. pastoris* cells in contrast to soluble GFP, which was homogenously distributed within the cell. Bottom row: merge of the GFP fluorescence and the Differential Interference Contrast (DIC) scans.
doi:10.1371/journal.pone.0060620.g002

the Fos-choline and Cyclofos detergents were able to solubilize both proteins in large quantities (Fig. 3A and B, D7-10, E7-10, F7-9). Also the Anapoe detergents (A-C 1-5) solubilized BSEP, but to a lesser extent. Furthermore, the anionic detergent dodecanoyl sarcosine (A9) as well as the zwitterionic detergents Anzergent® 3–14 (C8) and 2-carboxy-ω-heptadecenamidopropyldimethylamine (J8) resulted in strong signals in the dot blot.

In contrast to BSEP, we observed that only lipid-like detergents like the Fos-choline series (Fig. 3C, E9-E10, F6-F9, G8) and Cyclofos series (Fig. 3C, D8-D10) were able to solublize MDR3 in high amounts. In addition, the anionic detergents sodium dodecanoyl sarcosine (A9) and n-dodecyl-β-iminodipropionic acid (A10) were also able to solubilize MDR3. Furthermore, very low amounts of MDR3 were solubilized by Anzergent® 3-14 (C8) and 2-carboxy-ω-heptadecenamidopropyldimethylamine (J8). None of the Anapoes except Anapoe-58 (A4), none of the glucosides, none of the thio-maltosides, none of the maltosides or any other series of detergents showed a signal indicating that MDR3 was completely resistant to solubilization. In the case of MDR3, we did not analyze the GFP-fusion protein, because wild-type MDR3 displayed a substrate-induced ATPase activity (see below). Thus, the dot blot based solubilization screen revealed that only the lipid-like and more "harsh" detergents of the Fos-choline and Cyclofos series were able to solubilize both, BSEP and MDR3, in a near quantitative manner. For a quantification of the dot blots see Figure S1 in File S1.

Fluorescence-detection Size Exclusion Chromatography of selected detergents

Based on this analysis, the result of selected detergents used for the solubilization of BSEP and MDR3 were examined by fluorescence-detection size exclusion chromatography (FSEC). SEC is a common tool for monitoring the monodispersity and stability of proteins. In combination with a fluorescence detector, we were able to ascertain a high number of detergents using the GFP fusion proteins as reporter. This strategy requires only nanogram quantities of non-purified GFP-fusion protein by directly using solubilized membrane proteins in the detergent to be investigated. Our criteria for FSEC profiles in terms of monodispersity and stability were a sharp and symmetrical peak, no or only a small peak in the void volume or no signal corresponding to free GFP, which would indicate degradations of the fusion protein (for a FSEC profile of free GFP see Figure S2 in File S1).

The Fos-choline series as well as some maltosides and other detergents (see Figure S3 in File S1) solubilized BSEP-GFP, although the latter only resulted a weak signal in the dot blot. Fos-choline 8 and 9 did not give a significant signal in FSEC. A reliable signal was only obtained in the case of Fos-cholines containing long acyl chains. The signal increased with increasing acyl chain length from 10 to 16 carbon atoms (Fig. 4A). BSEP-GFP eluted to a certain portion in the void volume in Fos-choline detergents (10–11 min retention time), especially in Fos-choline-12 (Fig. 4A) indicating aggregated protein. The main BSEP-GFP peak (between 16 and 17 minutes) became more non-symmetrical and more BSEP-GFP degradation product (~20–21 min retention time, free GFP) was detected for detergents with longer acyl chains (Fig. 4A and Fig S3 in File S1). On the other hand, the maltosides gave sharp and symmetrical FSEC chromatograms and only very little aggregation was detected. This was very pronounced for β-DM, β-DDM and Cymal5. This implies that the protein was monodisperse and stable. Other detergents tested such as the
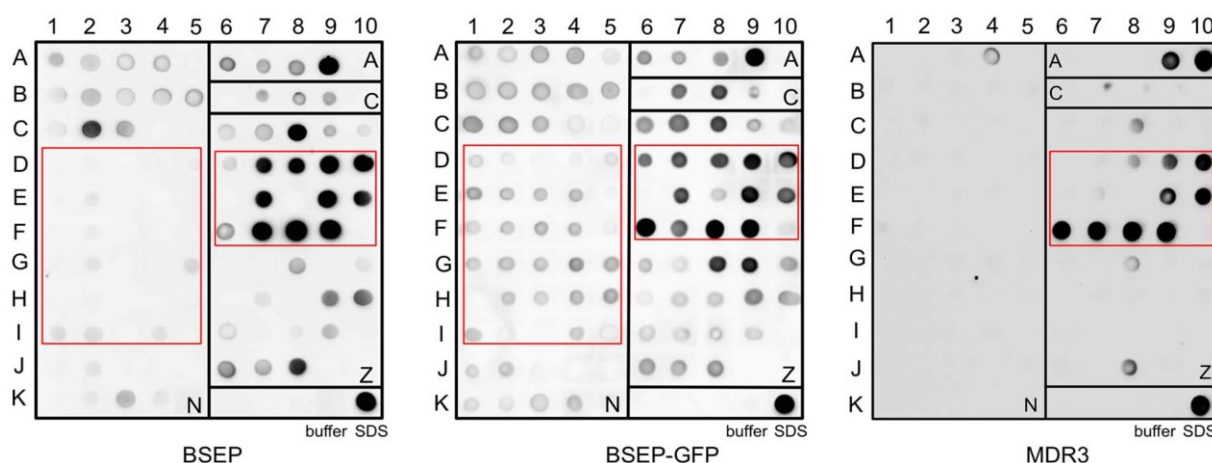
Figure 3. Solubilization screen of human BSEP and MDR3 using Dot Blot. Solubilization screen of *P. pastoris* membranes containing BSEP (**A**), BSEP-GFP (**B**) or MDR3 (**C**) with over 100 different detergents were analysed regarding the solubilization efficiacy. The solubilized protein was spotted onto a nitrocellulose membrane and examined via dot blotting with BSEP or MDR3 specific monoclonal antibodies (F-6 and C219). Compared areas are marked with a red box. All dot blots were performed in duplicate.
doi:10.1371/journal.pone.0060620.g003

anionic detergent sodium dodecanoyl sarcosine resulted in a non-symmetrical peak (Figure S3 in File S1). These observations, suggested that the length of the acyl chain of either group of detergents had a profound influence on the monodispersity and that an acyl chain length between 10 to 13 carbon atoms preserved the monodispersity of the transporters.

The FSEC profiles obtained for MDR3-GFP using the aforementioned detergents are summarized in Figure 4B and Figure S2 in File S1. None of the detergents showed perfect monodisperse peaks. The anionic detergent n-dodecyl-β-iminodipropionic acid was able to solubilize MDR3 (Figure S2 in File S1), however, the FSEC peak resulted in a major signal in the void volume of the SEC (molecular weight>1 MDa). This suggested aggregation. Sodium dodecanoyl sarcosine and the group of Cyclofos detergents showed a very inhomogeneous SEC profile (Figure S2 in File S1). The lipid-like Fos-cholines solubilized MDR3 with high efficiency nearly to the same extent as the SDS sample, which was used as control for solubilization efficiency (Fig. 4B). Importantly, the length of the acyl chain had again an impact on the monodispersity and stability of MDR3-GFP. The longer the acyl chain became, the more symmetrically the MDR3-GFP peak was observed (FC-16>FC-15>FC-14>FC-13>FC-12). The most promising result of solubilization efficiency and monodispersity was obtained for FC-16, so that all further experiments such as purification and ATPase acitivity were performed in this particular detergent.

Purification of the human ABC transporter MDR3 and BSEP

For functional analysis, we purified both transporter in the detergents, which showed the most promising results in the dot blot and FSEC analysis. The procedure we applied for the purification of human BSEP and MDR3 was established by Wang *et al.* based on the purification of the human ABC-transporters ABCG5/G8 and ABCC3 expressed in *P. pastoris* [34,46] and is described in detail in "Materials and Methods". MDR3 and BSEP both contain a tandem affinity tag consisting of a calmodulin binding-peptide tag (CBP-tag) and a 6xhis-tag at their C-termini. Briefly, BSEP and MDR3 were purified by immobilized metal-ion

affinity chromatography (IMAC) and calmodulin affinity resin (CBP) after solubilization of crude membranes in the appropriate detergent isolated from fermenter cultures.

We chose Fos-choline-16 as detergent of choice for solubilization of BSEP, because of its high efficacy. During the purification process, we exchanged the detergent on the CBP affinity column to maltoside detergents (e.g. β-DDM and Cymal-5), which according to the FSEC profiles corresponded to monodisperse protein (Fig. 4A). BSEP could be purified and yielded ~1 mg of protein from solubilized membranes of 100 g (wcw) of *Pichia* cells with a purity of roughly 75% (Fig. 5A).

MDR3 was solubilized with Fos-choline-16 and purified via an identical tandem affinity approach. The MDR3 transporter was visualized on a Coomassie blue-stained SDS-gel and further identified by immunoblot analysis (Fig. 6A). We obtained ~6 mg of highly purified protein from 100 g yeast cells with a purity of more than 90% as judged by SDS-PAGE analysis.

Binding of solubilized human BSEP to ATP-Agarose

BSEP was tested for ATPase activity in detergent solubilized state, but no reliable activity could be detected. Therefore, we investigated the capability of BSEP to bind to ATP coupled to agarose beads (ATP-beads) in the detergent-solubilized state, which would indicate that the protein is in a state where the nucleotide can bind, but the conformation is likely locked in a non-productive state, which inhibits hydrolysis. As shown in Figure 5B BSEP purified in Fos-choline-16 was not eluted from the ATP-beads after incubation suggesting that BSEP cannot bind to ATP in Fos-choline. Maltosids are known as mild detergents and often find usage to preserve the functionality of the membrane protein such as LmrA [47]. Accordingly, we solubilized BSEP with Fos-choline-16 and exchanged the detergent to β-DDM or Cymal-5 during purification. In these two detergents, BSEP bound to the ATP-beads. This result is in agreement with the FSEC results in those detergents (Fig. 4A).

ATPase Activity of purified human MDR3

We further examined whether purified MDR3 exhibits ATPase activity that could be stimulated by its natural substrate
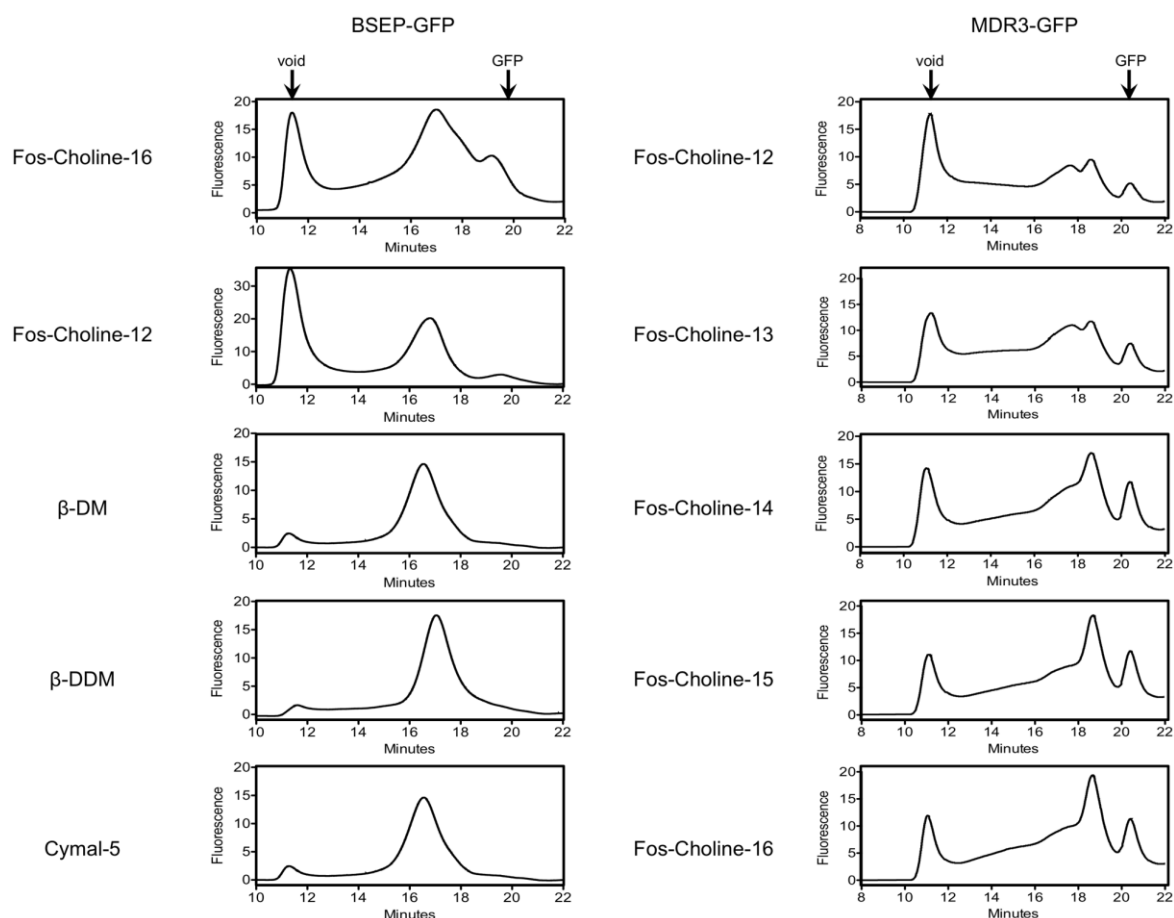
116

Figure 4. Detergent screening utilizing FSEC. FSEC analysis of BSEP-GFP (**A**) in five representative detergents and MDR3-GFP in five detergents (**B**). The arrows indicate the estimated elution position of the void volume and free GFP. Additional FSEC profiles are summarized in the supplementary material.
doi:10.1371/journal.pone.0060620.g004

phosphatidylcholine (PC) lipids. For this purpose we added two synthetic PC lipids (DPPC and DOPC) to the purified protein and measured the ATPase activity at 37°C up to 60 min (Fig. 6B). Under these conditions, we observed an approximately 2.5 fold stimulation of ATPase activity. Because co-purification of contaminating ATPases cannot be excluded, we cloned an ATP hydrolysis deficient mutant by introducing two point mutations and purified the mutant as described for the wild-type protein. The exchange of Glu to Gln in the highly conserved Walker B motif (ΦΦΦΦDE, where Φ can be every hydrophobic amino acid) of MDR3 prevents hydrolysis of ATP. The ATPase inactive mutant (E558Q, E1207Q, further called EQ/EQ mutant) exhibited basal ATPase activity comparable to the wild-type protein. This suggested that the observed activity was derived from co-purified ATPases. However and most important, no stimulation of activity was observed in the presence of PC lipids. ATPase activity of ABC transporters is often stimulated after addition of lipids. To demonstrate that the increased ATPase activity of MDR3 is caused by a substrate-specific and not by a conformational stabilization effect of PC lipids, we added DPPE and DOPE lipids to MDR3 wild-type as well as to the ATPase-deficient EQ/EQ mutant. The MDR3 wild-type ATPase activity is slightly

increased by a factor of 1.4 for DPPE and 1.6 for DOPE, whereas the ATPase activity of the EQ/EQ mutant are not increased compared to PC added ATPase activity. The data demonstrates a substrate-specific ATPase activity of 15 nmol/min per mg MDR3 wild-type in comparison to the DOPE-stimulated ATPase activity. We ascertained that the stimulation of ATPase activity is MDR3 specific by PC and indicated that MDR3 is functional in the detergent-solubilized state with respect to its capability to bind and hydrolyze ATP.

Discussion

In this study, we presented a high-throughput detergent screening and purification approach for the human liver-localized ABC transporters BSEP and MDR3 expressed in the methylotrophic yeast *Pichia pastoris*. This expression host has all the advantages of other eukaryotic expression systems, such as post-translational modifications or trafficking machinery. However, the overexpression per cell is only moderate and therefore requires fermentation to compensate this by high biomass. This system was used before for expression trials of human ABC transporters, which showed its general applicability for this class of transporter.
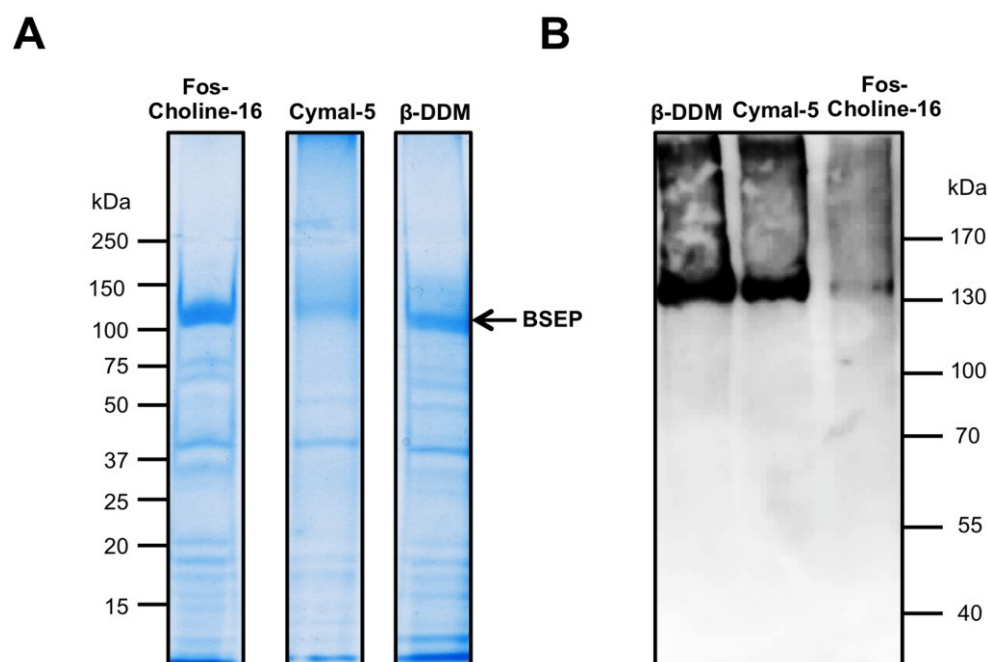
117

Figure 5. Purification and nucleotide binding of human BSEP. A Coomassie Brilliant Blue-stained SDS-PAGE of purified BSEP solubilized in Fos-choline-16 or in β-DDM and Cymal5, which were exchanged after solubilization. Molecular weight markers are indicated on the left. **B** Purified BSEP in all three detergents was incubated with ATP-agarose and bound protein was eluted in SDS sample buffer and examined with immunoblotting with a monoclonal antibody (F-6). BSEP signals could be detected in β-DDM and Cymal5, but not in Fos-Choline-16, indicating only binding to ATP in maltosides.
doi:10.1371/journal.pone.0060620.g005

Especially ABC transporters of the liver like MDR1 (P-gp, ABCB1), ABCG2, ABCG5/G8 or ABCC1, ABCC3 and ABCC6 (MRP 1, 3 and 6) could be expressed and partially purified [34,46,48,49,50,51]. Since BSEP and MDR3 were not included in this expression screen, we cloned these genes into the expression vector pSGP18. The cDNA of BSEP and MDR3 is unstable and cannot be cloned by conventional cloning in *E. coli* [45]. Therefore, we modified the pSGP18 vector. Both transporters as

well as the GFP-fusion proteins were expressed without detectable degradation products (Fig. 1). To analyze whether processing and especially targeting of BSEP and MDR3 to the plasma membrane in *P. pastoris* occurs, we employed fluorescence microscopy. Fluorescence microscopy of heterologous expressed proteins, particularly with distinct destinations in the cell is a valuable tool to directly judge the quality of the overexpressed protein. These experiments revealed that both transporters were targeted
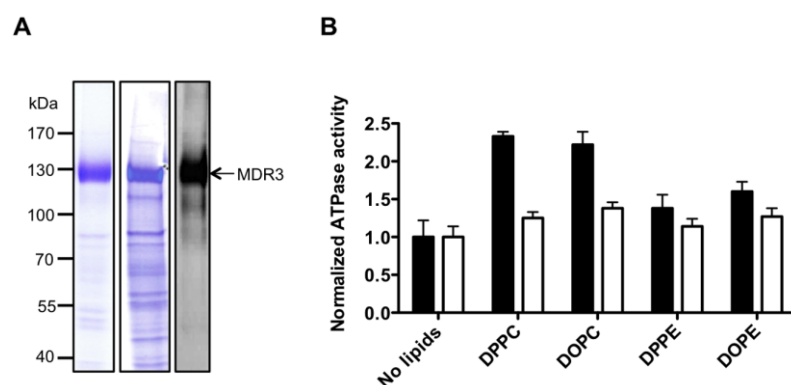
Figure 6. Characterization of purified human MDR3 in Fos-choline-16. A Coomassie Brilliant Blue-stained SDS-PAGE and immunoblot using an anti-MDR3 antibody of purified MDR3 wild-type and the MDR3 EQ/EQ-mutant via TAP. Molecular weight markers are shown on the left. **B** Normalized ATPase activity of MDR3 wild-type (black) and of an ATPase deficient mutant (E558Q E1207Q, white) in FC-16 without and with different phospholipids. The ATPase activity of three independent MDR3 purifications was determined ± SD (n = 3).
doi:10.1371/journal.pone.0060620.g006

118

correctly and no intracellular retention occurred. This adverts correct folding of BSEP and MDR3.

To date high-throughput methods are available to systematically screen a huge number of detergents in an appropriate time frame for their capability to solubilize the membrane protein of interest. We used a dot blot based solubilization screen on an analytical scale similar to approaches used for GPCRs heterologously expressed in *P. pastoris* or *E. coli* [36,52]. We analyzed more than 100 different detergents covering all four classes of detergents. Only detergents of the Fos-choline as well as Cyclofos series were able to solubilize BSEP and MDR3 in a nearly quantitative manner. None of the maltosides, thio-maltosides or glycosides were able to solubilize BSEP and MDR3. Surprisingly, BSEP-GFP could be solubilized to some degree by those detergents suggesting that the GFP tag enhances solubilization. The zwitterionic Fos-choline and Cyclofos series are lipid-like detergents and possess a head group consisting of phosphocholine, but differ in the hydrophobic part as Fos-cholines have a plain acyl chain with varying number of carbon atoms and Cyclofos detergents additionally contain a cyclohexane ring at the omega position of the acyl chain. This result is in contrast to other used detergents for liver ABC transporters expressed heterologously in *P. pastoris*. ABCC3 was solubilized in β-DDM like ABCG5/G8 [34,53]. MDR1 was solubilized in various detergents from *P. pastoris* membranes including β-DM, β-DDM, Lyso-PC, deoxycholic acid or Triton-X100 [48,54,55,56,57]. Despite the high degree of sequence identity between MDR1 and MDR3 (>85% homology to human MDR1, 80% to mouse MDR1), MDR3 behaves different, since it could not be solubilized with Triton-X100 (data not shown), which was used to crystallize mouse MDR1 [54]. Also ABCG2 was solubilized in β-DDM, but could only be solubilized in Fos-Choline-16 when expressed in *High Five* cells [46,49].

GFP fusion proteins cannot only be used as quality marker for heterologous expression, but also as a tool to screen the influence of a detergent to the membrane protein using FSEC. We employed this technique to investigate those detergents more in detail that were successfully identified in the dot blot screen. The Fos-Choline series displayed a clear dependence on the acyl chain length, i. e. increasing the acyl chain length increased the monodispersity of the protein sample. For BSEP-GFP, we also tested some maltoside detergents, although the solubilization efficacy was moderate for BSEP-GFP as judged from the dot blot. All tested maltosides showed very monodisperse FSEC profiles with less aggregation and a symmetrical peak, e.g. with Cymal-5 or β-DDM. β-DDM in general is believed to be a mild detergent and is often used for solubilization, purification and crystallization trails. On the other hand, the Fos-cholines showed a large aggregation peak and with increasing acyl chain length, a BSEP-GFP degradation product was more visible indicating instability of the membrane protein, and the peak became more unsymmetrically. Nonetheless, we decided to use Fos-choline-16 for BSEP and MDR3 for solubilization because of its efficacy and its use for other ABC transporter like LmrA [47], BmrC/D [58] and the aforementioned ABCG2 [46].

We were able to purify MDR3 and BSEP for the first time yielding ~6 mg and ~1 mg of protein per 100 g of cells, respectively. This is in good agreement with ABCB1 (~6 mg) or ABCC3 (~9 mg) [34,48]. BSEP is expressed at lower levels than MDR3 in *P. pastoris* and thus the yield is lower underlining the variance of expression of different proteins. Both transporters were purified by TAP from crude membranes, which resulted in a homogeneous preparation for MDR3 as judged by SDS-PAGE. In case of BSEP, the purity was not as high.

BSEP and MDR3 belong to the ABC transporter family and ATP hydrolysis drives translocation of bile salts or phosphatidylcholine. However, we could not detect any ATPase activity for BSEP in the detergent-solubilized state, neither basal nor substrate induced. To see whether BSEP was purified in a state, which at least allows binding of ATP, we employed ATP beads. Here, we could confirm that purified BSEP is able to bind ATP in the presence of β-DDM and Cymal-5. This indicates that at least the NBDs of BSEP are properly folded, which is a prerequisite for proper functioning and that Fos-choline-16 likely locks the protein in a binding-incompatible state, whereas β-DDM and Cymal-5 invert this state. For BSEP it is known, that its transport activity is depending on cholesterol [59]. Enrichment of *Sf9* cell membranes expressing BSEP with cholesterol drastically increases its transport activity [60]. If cholesterol is bound to the transporter itself or is just required as a membrane component has not been clarified yet. In the yeast expression host however, ergosterol is the predominant sterol instead of cholesterol like in other mammalian cells. Both sterols differ by two additional double bonds (in the ring and in the tail) in the case of ergosterol. The striking dependence of transport activity of BSEP on cholesterol and the absence of this steroid in yeast might explain the lack of ATPase activity. However, even in the absence of cholesterol, BSEP is able to bind to ATP in the detergent-solubilized state.

MDR3 displayed a substantial ATPase activity. This observation was sustained by analysis of an ATP hydrolysis deficient EQ double mutant (E558Q, E1207Q). Generally, a mutation of the glutamine of the Walker B motif renders ABC transporters ATPase inactive. Here, we generated the double mutant to ensure that the observed stimulation of ATPases was not due to contaminating ATPases. The substrate specificity for MDR3 was already investigated using *S. cerevisiae* secretory vesicles or cell-culture based methods [16,32,61]. All experiments showed, that MDR3 translocates short chain PC lipids (C_8) or long chain derivatives (C_{16}), but not PE, sphingomyelin or ceramides.

Here, we demonstrate for the first time that the ATPase activity of detergent-solubilized MDR3 in the presence of phosphatidylcholine lipids could be stimulated by a factor of almost 2.5, while the EQ/EQ mutant did not display any stimulation. Furthermore, it was shown that the specificity resulted from the phosphatidylcholine headgroup. We proved that MDR3 ATPase activity is specifically stimulated by PC lipids and not by PE lipids, which differ only in the headgroup.

In summary, we demonstrate for the first time the expression of two human ABC transporters, MDR3 and BSEP, in the yeast *P. pastoris* and their correct targeting to the plasma membrane. BSEP could bind to ATP in detergent, but no hydrolytic activity could be detected. Furthermore, we established a purification procedure for human MDR3, which resulted in purified and functional protein. This study provides the foundation for further investigations of the human liver ABC transporters BSEP and MDR3.

Supporting Information

File S1 Combined file of supporting figures and tables. Figure S1: Dot Blot quantification of BSEP (A), BSEP-GFP (B) and MDR3 (C). Average values from two independent dot blots are shown (n = 2) ± SD. Large errors for e.g. the Fos-Choline series resulted from saturation of the detector. The intensity of SDS was set to 100% and all other values were normalized to SDS. Black bars represent zwitter-ionic detergents, grey bars ionic detergents and white bars non-ionic detergents. Figure S2: FSEC profiles of free GFP and MDR3-GFP in selected detergents. The x-axis shows time in minutes, the y-axis fluorescence in arbitrary units.

Figure S3: FSEC profiles of BSEP-GFP in selected detergents. The x-axis shows time in minutes, the y-axis fluorescence in arbitrary units. Table S1: Used detergents for solubilization of BSEP and MDR3 and Dot Blot analysis; N: Non-ionic detergents; Z: Zwitterionic detergents; A: Anionic detergents; C: Cationic detergents.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: PE MK SS LS. Performed the experiments: PE MK JS. Analyzed the data: PE MK SS LS. Contributed reagents/materials/analysis tools: JS. Wrote the paper: PE MK SS LS.

References

1. Schmitt L, Tampe R (2002) Structure and mechanism of ABC transporters. Current opinion in structural biology 12: 754–760.
2. Davidson AL, Dassa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiology and molecular biology reviews: MMBR 72: 317–364, table of contents.
3. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome research 11: 1156–1166.
4. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066–1073.
5. Mosser J, Douar AM, Sarde CO, Kioschis P, Feil R, et al. (1993) Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. Nature 361: 726–730.
6. Bodzioch M, Orso E, Klucken J, Langmann T, Bottcher A, et al. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. Nature genetics 22: 347–351.
7. Brooks-Wilson A, Marcil M, Clee SM, Zhang LH, Roomp K, et al. (1999) Mutations in ABC1 in Tangier disease and familial high-density lipoprotein deficiency. Nature genetics 22: 336–345.
8. Keppler D (2011) Multidrug resistance proteins (MRPs, ABCCs): importance for pathophysiology and drug therapy. Handbook of experimental pharmacology: 299–323.
9. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM (2003) P-glycoprotein: from genomics to mechanism. Oncogene 22: 7468–7485.
10. Ni Z, Bikadi Z, Rosenberg MF, Mao Q (2010) Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). Current drug metabolism 11: 603–617.
11. Meier PJ, Stieger B (2002) Bile salt transporters. Annual review of physiology 64: 635–661.
12. Hofmann AF, Hagey LR (2008) Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. Cellular and molecular life sciences: CMLS 65: 2461–2483.
13. Noe J, Stieger B, Meier PJ (2002) Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 123: 1659–1666.
14. Kubitz R, Droge C, Stindt J, Weissenberger K, Haussinger D (2012) The bile salt export pump (BSEP) in health and disease. Clinics and research in hepatology and gastroenterology.
15. Stieger B (2011) The role of the sodium-taurocholate cotransporting polypeptide (NTCP) and of the bile salt export pump (BSEP) in physiology and pathophysiology of bile formation. Handbook of experimental pharmacology: 205–259.
16. van Helvoort A, Smith AJ, Sprong H, Fritzsche I, Schinkel AH, et al. (1996) MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine. Cell 87: 507–517.
17. Oude Elferink RP, Paulusma CC (2007) Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). Pflugers Archiv: European journal of physiology 453: 601–610.
18. Graf GA, Yu L, Li WP, Gerard R, Tuma PL, et al. (2003) ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. The Journal of biological chemistry 278: 48275–48282.
19. Small DM (2003) Role of ABC transporters in secretion of cholesterol from liver into bile. Proceedings of the National Academy of Sciences of the United States of America 100: 4–6.
20. Strautnieks SS, Kagalwalla AF, Tanner MS, Knisely AS, Bull L, et al. (1997) Identification of a locus for progressive familial intrahepatic cholestasis PFIC2 on chromosome 2q24. American journal of human genetics 61: 630–633.
21. Deleuze JF, Jacquemin E, Dubuisson C, Cresteil D, Dumont M, et al. (1996) Defect of multidrug-resistance 3 gene expression in a subtype of progressive familial intrahepatic cholestasis. Hepatology 23: 904–908.
22. Dzagania T, Engelmann G, Haussinger D, Schmitt L, Flechtenmacher C, et al. (2012) The histidine-loop is essential for transport activity of human MDR3. A novel mutation of MDR3 in a patient with progressive familial intrahepatic cholestasis type 3. Gene 506: 141–145.
23. van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, et al. (2004) Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 127: 379–384.
24. Glantz A, Marschall HU, Mattsson LA (2004) Intrahepatic cholestasis of pregnancy: Relationships between bile acid levels and fetal complication rates. Hepatology 40: 467–474.
25. Rosmorduc O, Hermelin B, Poupon R (2001) MDR3 gene defect in adults with symptomatic intrahepatic and gallbladder cholesterol cholelithiasis. Gastroenterology 120: 1459–1467.
26. Jacquemin E (2012) Progressive familial intrahepatic cholestasis. Clinics and research in hepatology and gastroenterology 36 Suppl 1: S26–35.
27. Gonzales E, Grosse B, Cassio D, Davit-Spraul A, Fabre M, et al. (2012) Successful mutation-specific chaperone therapy with 4-phenylbutyrate in a child with progressive familial intrahepatic cholestasis type 2. Journal of hepatology 57: 695–698.
28. Kubitz R, Sutfels G, Kuhlkamp T, Kolling R, Haussinger D (2004) Trafficking of the bile salt export pump from the Golgi to the canalicular membrane is regulated by the p38 MAP kinase. Gastroenterology 126: 541–553.
29. Mita S, Suzuki H, Akita H, Hayashi H, Onuki R, et al. (2006) Vectorial transport of unconjugated and conjugated bile salts by monolayers of LLC-PK1 cells doubly transfected with human NTCP and BSEP or with rat Ntcp and Bsep. American journal of physiology Gastrointestinal and liver physiology 290: G550–556.
30. Wang R, Chen HL, Liu L, Sheps JA, Phillips MJ, et al. (2009) Compensatory role of P-glycoproteins in knockout mice lacking the bile salt export pump. Hepatology 50: 948–956.
31. Hayashi H, Takada T, Suzuki H, Akita H, Sugiyama Y (2005) Two common PFIC2 mutations are associated with the impaired membrane trafficking of BSEP/ABCB11. Hepatology 41: 916–924.
32. Groen A, Romero MR, Kunne C, Hoosdally SJ, Dixon PH, et al. (2011) Complementary functions of the flippase ATP8B1 and the floppase ABCB4 in maintaining canalicular membrane integrity. Gastroenterology 141: 1927–1937.e1921–1924.
33. Byrne JA, Strautnieks SS, Mieli-Vergani G, Higgins CF, Linton KJ, et al. (2002) The human bile salt export pump: characterization of substrate specificity and identification of inhibitors. Gastroenterology 123: 1649–1658.
34. Chloupkova M, Pickert A, Lee JY, Souza S, Trinh YT, et al. (2007) Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 46: 7992–8003.
35. Junge F, Schneider B, Reckel S, Schwarz D, Dotsch V, et al. (2008) Large-scale production of functional membrane proteins. Cellular and molecular life sciences: CMLS 65: 1729–1755.
36. Zeder-Lutz G, Cherouati N, Reinhart C, Pattus F, Wagner R (2006) Dot-blot immunodetection as a versatile and high-throughput assay to evaluate recombinant GPCRs produced in the yeast Pichia pastoris. Protein expression and purification 50: 118–127.
37. Newstead S, Kim H, von Heijne G, Iwata S, Drew D (2007) High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 104: 13936–13941.
38. Mizutani K, Yoshioka S, Mizutani Y, Iwata S, Mikami B (2011) High-throughput construction of expression system using yeast Pichia pastoris, and its application to membrane proteins. Protein expression and purification 77: 1–8.
39. Drew D, Newstead S, Sonoda Y, Kim H, von Heijne G, et al. (2008) GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. Nature protocols 3: 784–798.
40. Eshaghi S (2009) High-throughput expression and detergent screening of integral membrane proteins. Methods in molecular biology 498: 265–271.
41. Gutmann DA, Mizohata E, Newstead S, Ferrandon S, Postis V, et al. (2007) A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay. Protein science: a publication of the Protein Society 16: 1422–1428.
42. Vergis JM, Purdy MD, Wiener MC (2010) A high-throughput differential filtration assay to screen and select detergents for membrane proteins. Analytical biochemistry 407: 1–11.

120

43. Hattori M, Hibbs RE, Gouaux E (2012) A fluorescence-detection size-exclusion chromatography-based thermostability assay for membrane protein precrystallization screening. Structure 20: 1293–1299.

44. Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14: 673–681.

45. Stindt J, Ellinger P, Stross C, Keitel V, Haussinger D, et al. (2011) Heterologous overexpression and mutagenesis of the human bile salt export pump (ABCB11) using DREAM (Directed REcombination-Assisted Mutagenesis). PloS one 6: e20562.

46. McDevitt CA, Collins R, Kerr ID, Callaghan R (2009) Purification and structural analyses of ABCG2. Advanced drug delivery reviews 61: 57–65.

47. Infed N, Hanekop N, Driessen AJ, Smits SH, Schmitt L (2011) Influence of detergents on the activity of the ABC transporter LmrA. Biochimica et biophysica acta 1808: 2313–2321.

48. Lerner-Marmarosh N, Gimi K, Urbatsch IL, Gros P, Senior AE (1999) Large scale purification of detergent-soluble P-glycoprotein from Pichia pastoris cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins. The Journal of biological chemistry 274: 34711–34718.

49. Rosenberg MF, Bikadi Z, Chan J, Liu X, Ni Z, et al. (2010) The human breast cancer resistance protein (BCRP/ABCG2) shows conformational changes with mitoxantrone. Structure 18: 482–493.

50. Cai J, Daoud R, Alqawi O, Georges E, Pelletier J, et al. (2002) Nucleotide binding and nucleotide hydrolysis properties of the ABC transporter MRP6 (ABCC6). Biochemistry 41: 8058–8067.

51. Cai J, Daoud R, Georges E, Gros P (2001) Functional expression of multidrug resistance protein 1 in Pichia pastoris. Biochemistry 40: 8307–8316.

52. Ren H, Yu D, Ge B, Cook B, Xu Z, et al. (2009) High-level production, solubilization and purification of synthetic human GPCR chemokine receptors CCR5, CCR3, CXCR4 and CX3CR1. PloS one 4: e4509.

53. Johnson BJ, Lee JY, Pickert A, Urbatsch IL (2010) Bile acids stimulate ATP hydrolysis in the purified cholesterol transporter ABCG5/G8. Biochemistry 49: 3403–3411.

54. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718–1722.

55. Beaudet L, Urbatsch IL, Gros P (1998) Mutations in the nucleotide-binding sites of P-glycoprotein that affect substrate specificity modulate substrate-induced adenosine triphosphatase activity. Biochemistry 37: 9073–9082.

56. Bai J, Swartz DJ, Protasevich, II, Brouillette CG, Harrell PM, et al. (2011) A gene optimization strategy that enhances production of fully functional P-glycoprotein in Pichia pastoris. PloS one 6: e22577.

57. Urbatsch IL, Beaudet L, Carrier I, Gros P (1998) Mutations in either nucleotide-binding site of P-glycoprotein (Mdr3) prevent vanadate trapping of nucleotide at both sites. Biochemistry 37: 4592–4602.

58. Galian C, Manon F, Dezi M, Torres C, Ebel C, et al. (2011) Optimized purification of a heterodimeric ABC transporter in a highly stable form amenable to 2-D crystallization. PloS one 6: e19677.

59. Paulusma CC, de Waart DR, Kunne C, Mok KS, Elferink RP (2009) Activity of the bile salt export pump (ABCB11) is critically dependent on canalicular membrane cholesterol content. The Journal of biological chemistry 284: 9947–9954.

60. Kis E, Ioja E, Nagy T, Szente L, Heredi-Szabo K, et al. (2009) Effect of membrane cholesterol on BSEP/Bsep activity: species specificity studies for substrates and inhibitors. Drug metabolism and disposition: the biological fate of chemicals 37: 1878–1886.

61. Ruetz S, Gros P (1994) Phosphatidylcholine translocase: a physiological role for the mdr2 gene. Cell 77: 1071–1081.

121

Supplementary Information

Detergent Screening and Purification of the Human Liver ABC Transporters BSEP (ABCB11) and MDR3 (ABCB4) expressed in the yeast *Pichia pastoris*
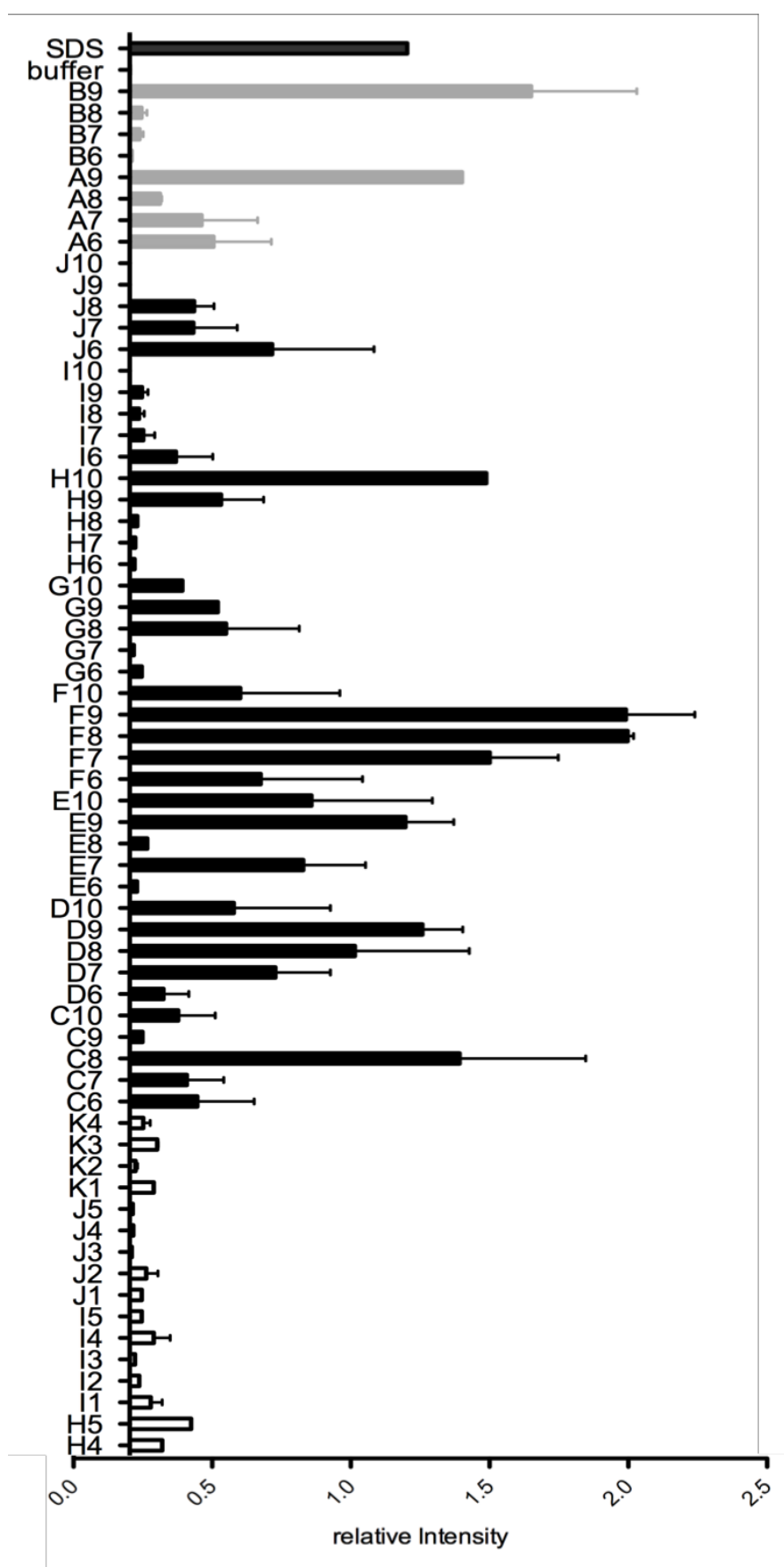
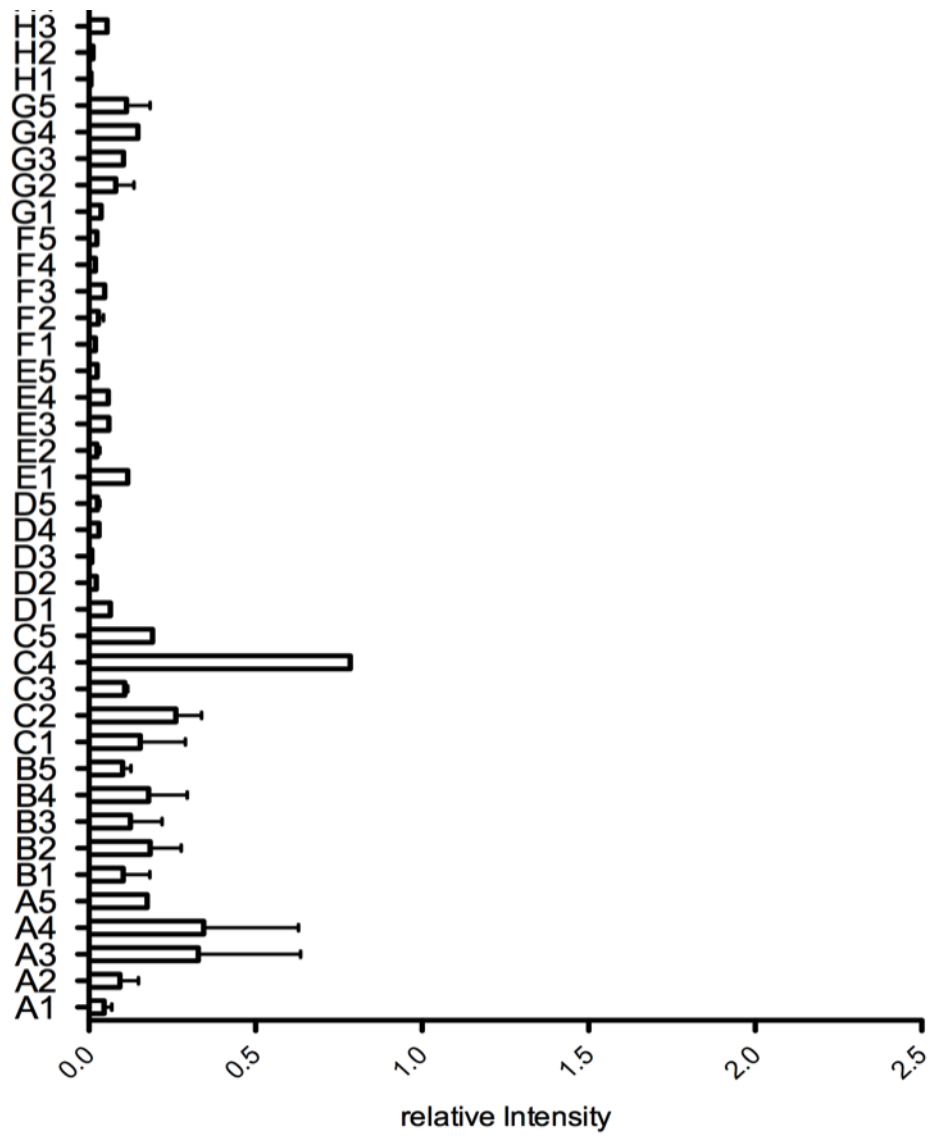Philipp Ellinger, Marianne Kluth, Jan Stindt, Sander H. Smits and Lutz Schmitt

Institute of Biochemistry, Heinrich Heine University, Düsseldorf. Germany
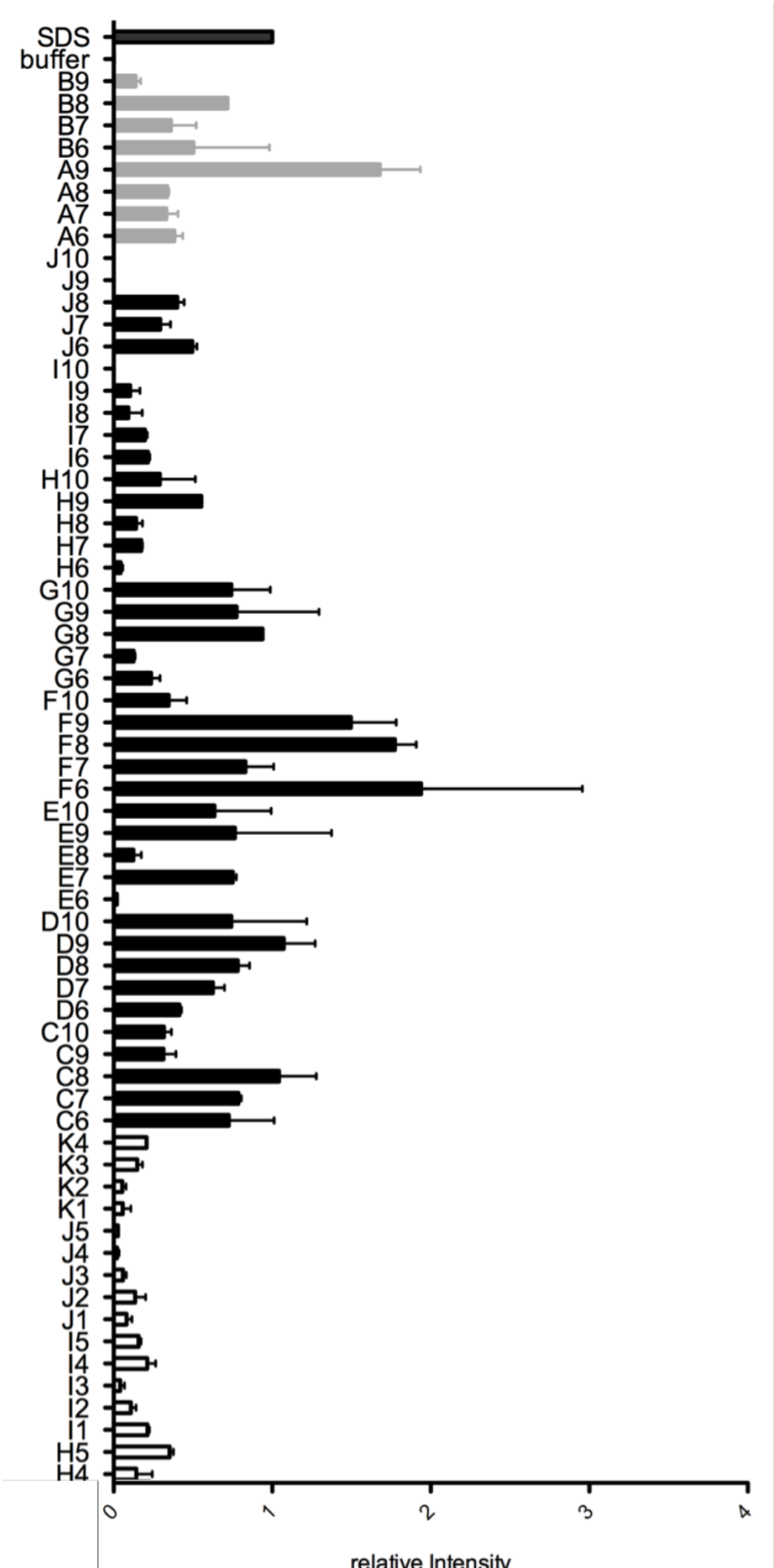
Supplementary Information comprises:
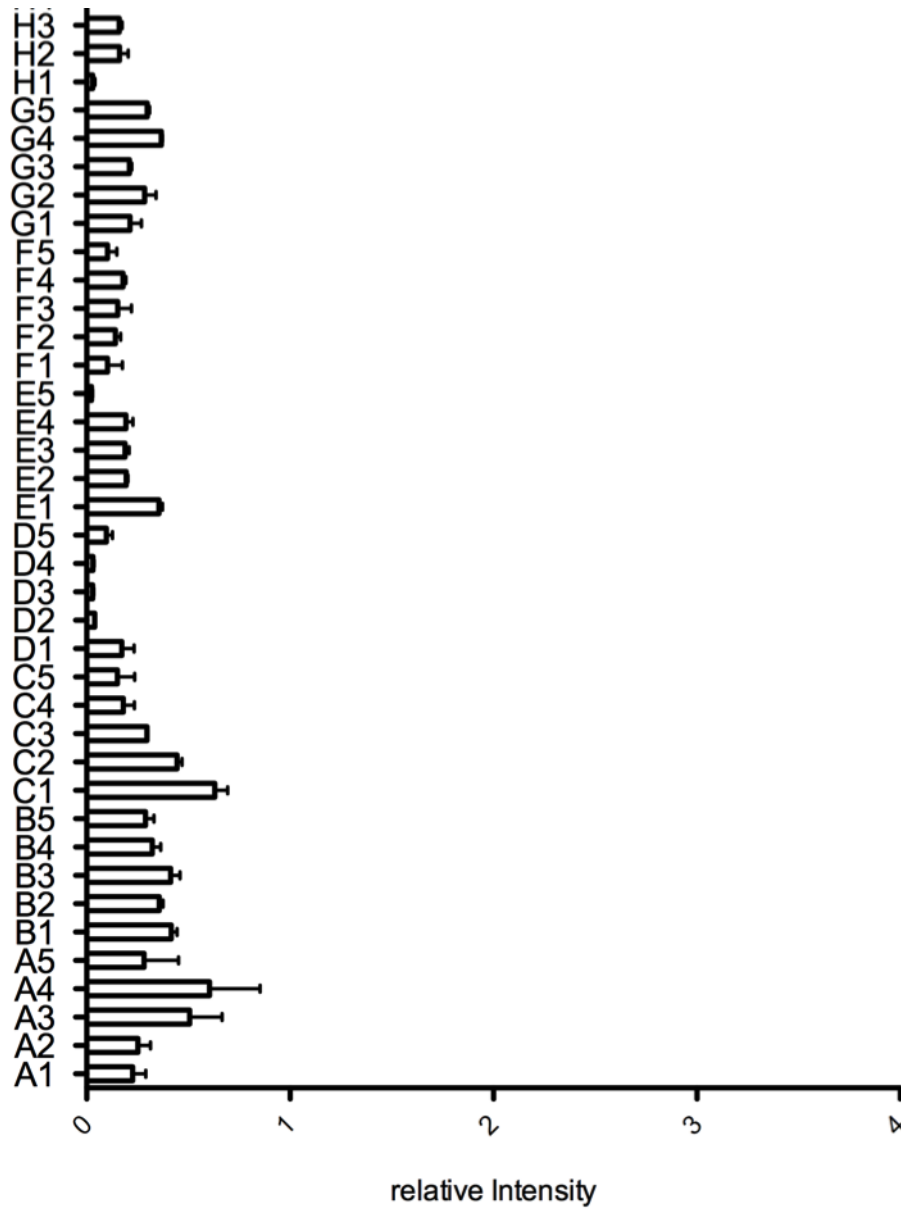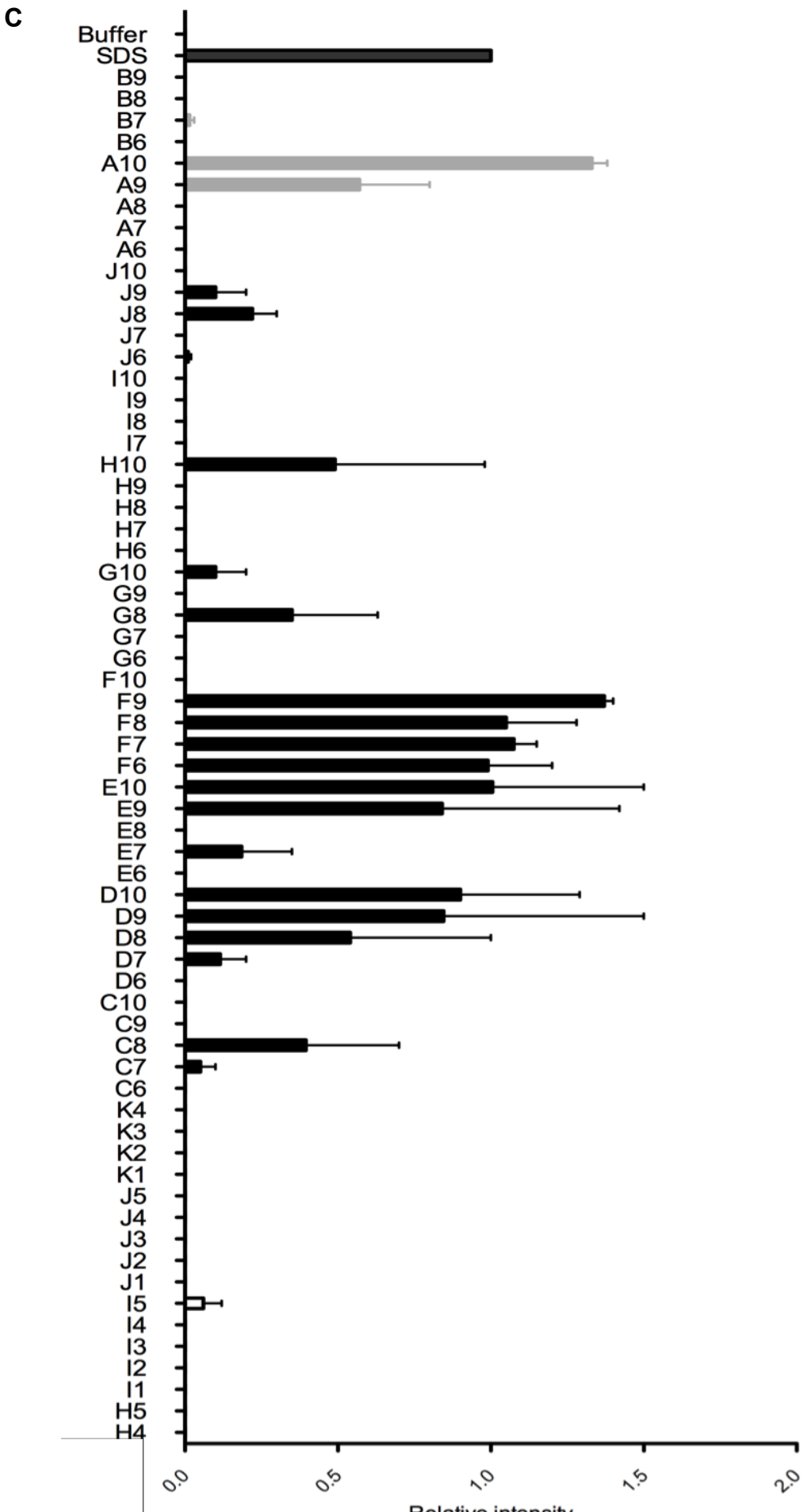
Supplementary Figures 1-3
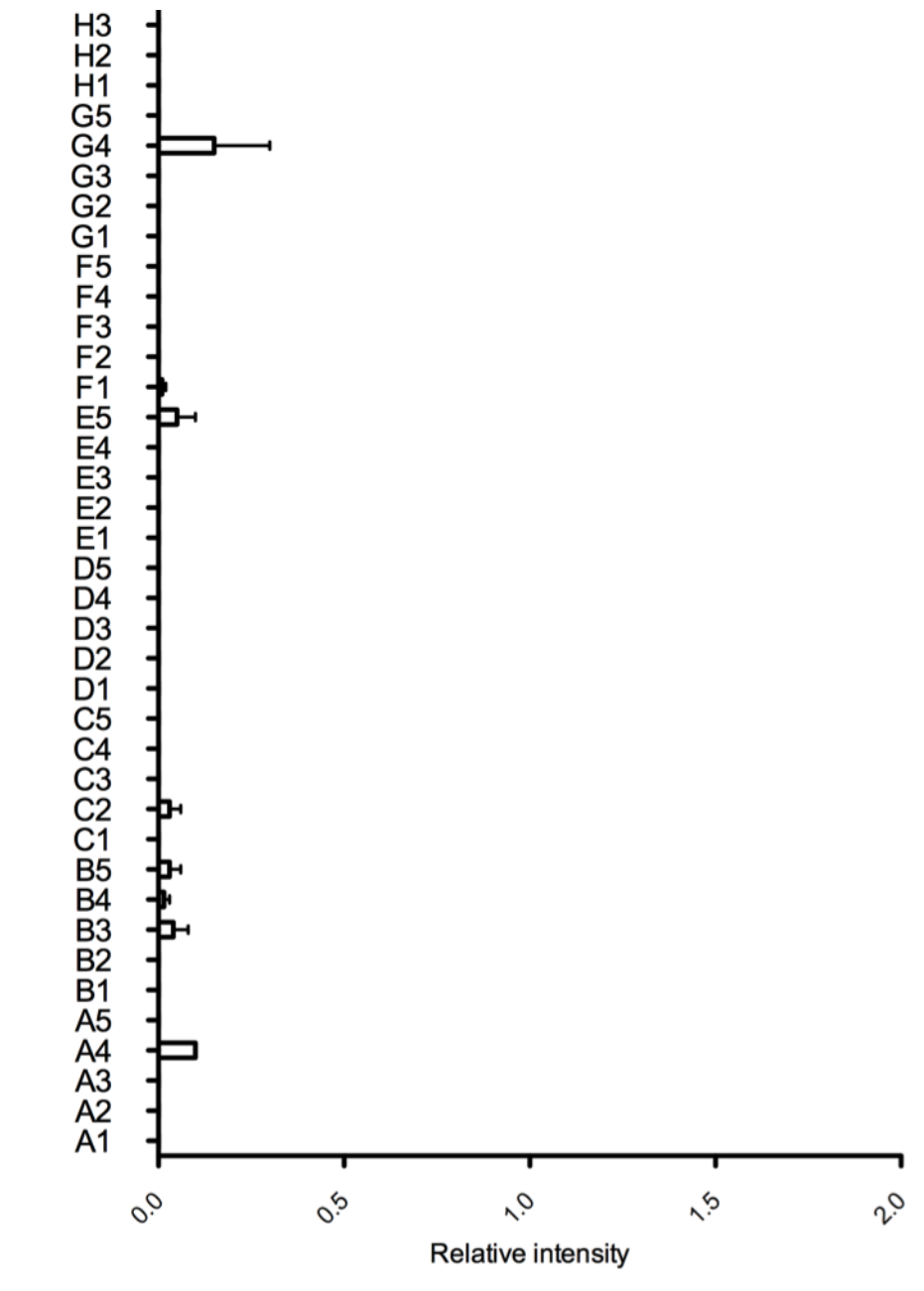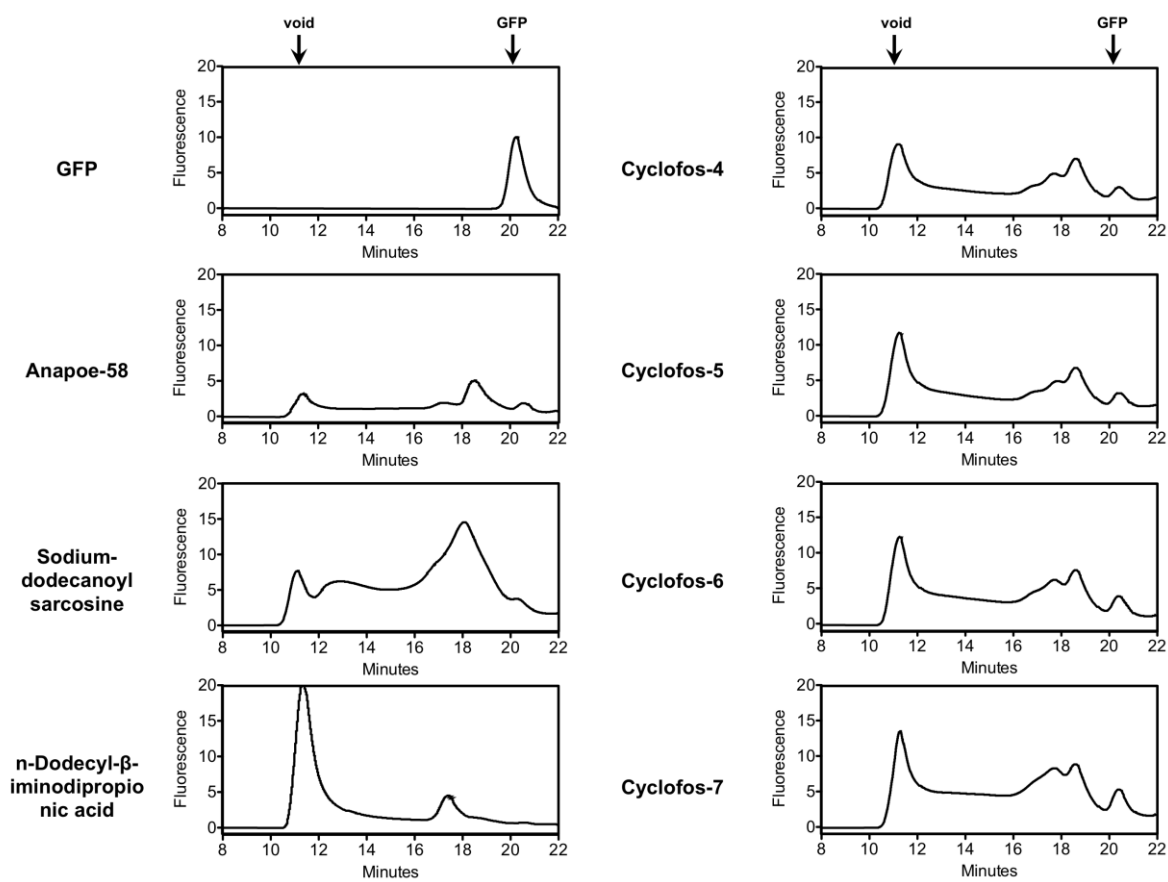
Supplementary Table 1

Supplementary Figure S1

A

B

Supplementary Figure S2
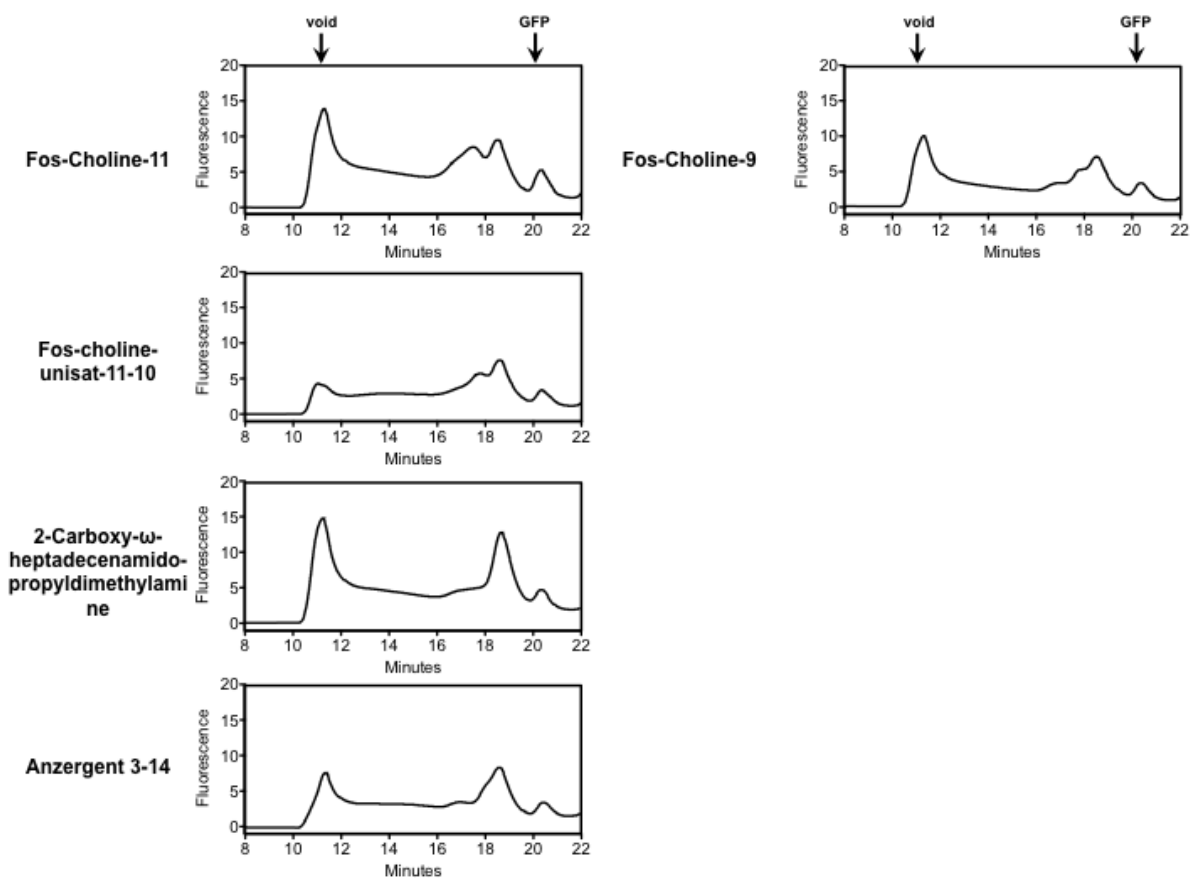
Supplementary Figure S3
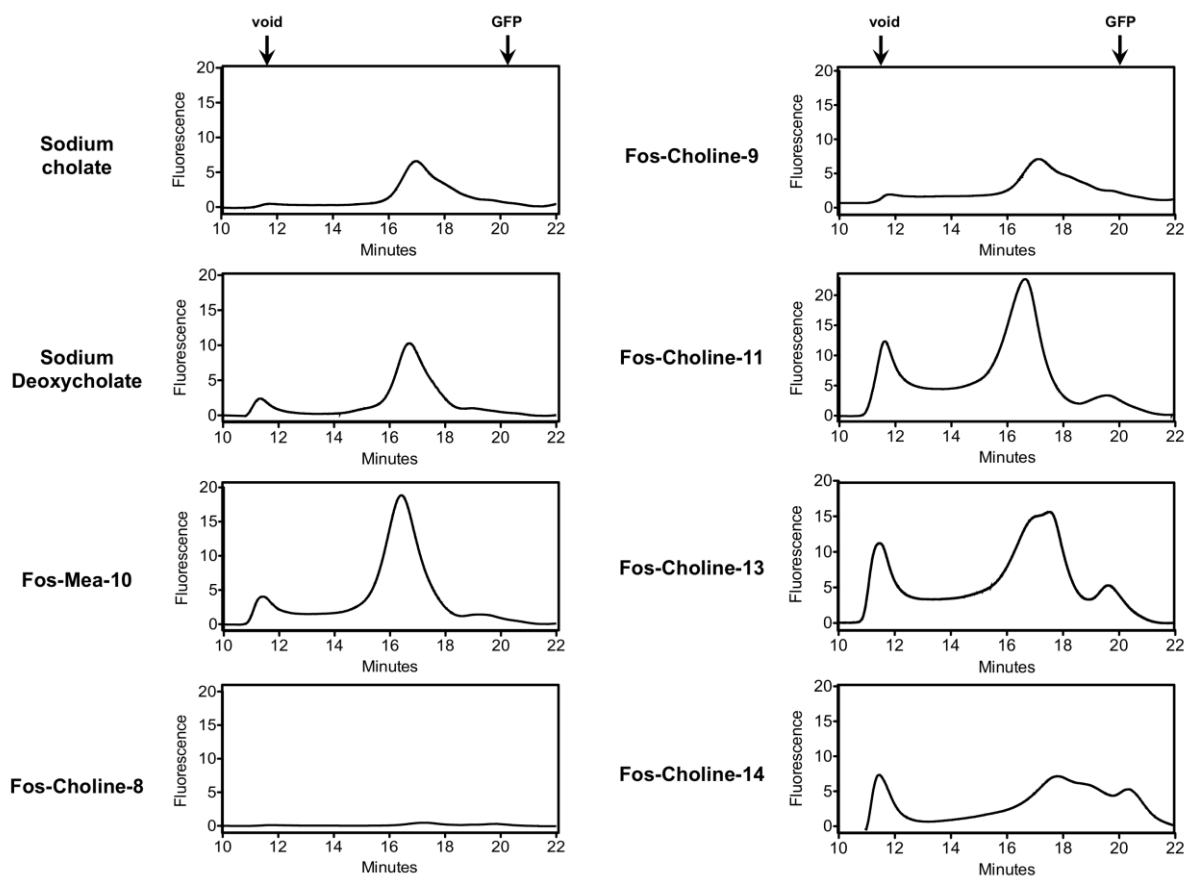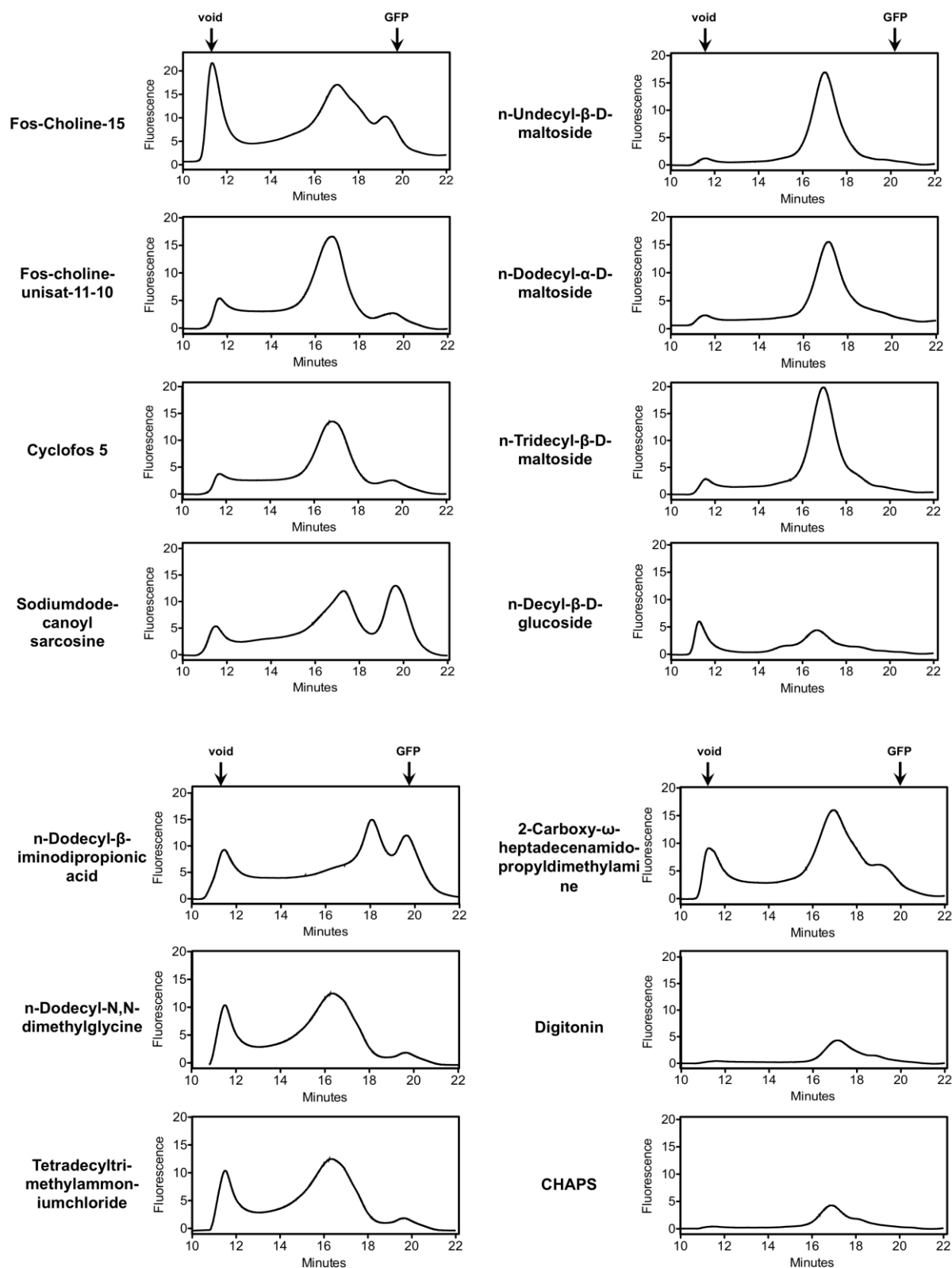
Supplementary Table S1

Position	Detergent	cmc %	used %	nature
A1	Anameg®-7	0.65	1%	N
A2	Anapoe®-20	0.0072	1%	N
A3	Anapoe®-35	0.001	1%	N
A4	Anapoe®-58	0.00045	1%	N
A5	Anapoe®-80	0.0016	1%	N
B1	Anapoe®-C10E6	0.025	1%	N
B2	Anapoe®-C10E9	0.053	1%	N
B3	Anapoe®-C12E8	0.0048	1%	N
B4	Anapoe®-C12E9	0.003	1%	N
B5	Anapoe®-C12E10	0.2	1%	N
C1	Anapoe®-C13E8	0.0055	1%	N
C2	Anapoe®-X-100	0.015	1%	N
C3	Anapoe®-X-114	0.011	1%	N
C4	Anapoe®-X-305	–	1%	N
C5	Anapoe®-X-405	0.16	1%	N
D1	Big CHAP	0.25	1%	N
D2	Big CHAP deoxy	0.12	1%	N
D3	CYGLU®-3	0.86	2%	N
D4	CYMAL®-1	15	2%	N
D5	CYMAL®-2	5.4	2%	N
E1	CYMAL®-3	0.37	1%	N
E2	2.6-Dimethyl-4-heptyl-β-D-maltose	1.2	2%	N
E3	2-propyl-1-pentyl maltose	1.9	2%	N
E4	MEGA-8	2.5	2%	N
E5	n-Octyl-β-D-glucoside	0.53	2%	N
F1	n-Nonyl-β-D-glucoside	0.2	1%	N
F2	n-Octyl-β-D-maltoside	0.89	2%	N
F3	n-Nonyl-β-D-maltoside	0.28	2%	N
F4	n-Decyl-α-D-maltoside	–	1%	N
F5	n-Tetradecyl-β-D-maltoside	0.00054	1%	N
G1	n-Undecyl-α-D-maltoside	0.029	1%	N
G2	n-Undecyl-β-D-maltoside	0.029	1%	N
G3	n-Dodecyl-α-D-maltoside	0.0076	1%	N
G4	n-Dodecyl-β-D-maltoside	0.0087	1%	N

G5	n-Tridecyl-β-D-maltoside	0.0017	1%	N
H1	n-Heptyl-β-D-thioglucoside	0.85	2%	N
H2	n-Octyl-β-D-thiomaltoside	0.4	2%	N
H3	n-Nonyl-β-D-thiomaltoside	0.15	1%	N
H4	n-Decyl-β-D-thiomaltoside	0.045	1%	N
H5	n-Undecyl-β-D-thiomaltoside	0.011	1%	N
I1	n-Dodecyl-β-D-thiomaltoside	0.0026	1%	N
I2	Pentaethylene glycol monodecylether(C10E5)	0.031	1%	N
I3	Tetraethylene glycol monoethylether(C8E4)	0.25	1%	N
I4	Sucrose monododecanoate	0.016	1%	N
I5	Dimethyldecylphosphine oxide	0.1	1%	N
J1	Tripglu	3.6	2%	N
J2				
J3	Decyl-β-D-glucoside	0.07	1%	N
J4				
J5	CYGLU®-4	0.058	2%	N
K1	CYMAL®-5	0.12	2%	N
K2	MEGA-10	0.21	2%	N
K3	NP40	0.05-0.3	1%	N
K4	Cyclohexyl-n-hexyl-β-D-maltoside	–	1%	N
C6	Anzergent® 3-10	1.2	2%	Z
C7	Anzergent®3-12	0.094	1%	Z
C8	Anzergent® 3-14	0.007	1%	Z
C9	CHAPS	0.49	2%	Z
C10	CHAPSO	0.5	2%	Z
D6	C-DODECAFOS™	0.77	2%	Z
D7	Cyclofos™-4	0.45	2%	Z
D8	Cyclofos™-5	0.15	1%	Z
D9	Cyclofos™-6	0.094	1%	N
D10	Cyclofos™-7	0.022	1%	Z
E6	Cyclofos™-2	7.5	1%	Z
E7	Fos-Choline®-9	1.2	2%	Z
E8	Cyclofos™-3	1.3	2%	Z
E9	Fos-Choline®-11	0.062	1%	Z
E10	Fos-Choline®-12	0.047	1%	Z
F6	Fos-Choline®-13	0.027	1%	Z

F7	Fos-Choline®-14	0.0046	1%	Z
F8	Fos-Choline®-15	0.0027	1%	Z
F9	Fos-Choline®-16	0.00053	1%	Z
F10	Fos-Choline®-Iso-9	0.99	2%	Z
G6	Fos-Choline®-Iso-11	0.9	2%	Z
G7	Fos-Choline®-Iso-11-6U	0.87	2%	Z
G8	Fos-Choline®-Unisat-11-10	0.21	1%	Z
G9	Fos-Choline®-8	3.4	2%	Z
G10	Fosfen™-9	0.014	1%	Z
H6	Nopol-Fos™	1.4	2%	Z
H7	PMAL™-8	–	1%	Z
H8	PMAL™-C10	–	1%	Z
H9	n-Decyl-N.N-dimethylglycine	0.46	2%	Z
H10	n-Dodecyl-N.N-dimethylglycine	0.041	1%	Z
I7	n-Tetradecyl-N.N-dimethylamine-N-oxide	0.0075	1%	Z
I8	n-Dodecyl-N.N-dimethylamine-N-oxide	0.023	1%	Z
I9	Tripao	4.5	2%	Z
I10	n-Tetradecyl-N.N-dimethylamine-N-oxide	0.0075	1%	Z
J6	LAPAO	0.052	2%	Z
J7	PMAL™-C-12	–	2%	Z
J8	2-Carboxy-w-heptadecenamidopropyldimethylamine	–	1%	Z
J9	2-Carboxy-5-pentadecenamidopropyldimethylamine	–	1%	Z
J10	N.N dimethyl(3-carboxy-4-dodec-5-ene)aminopropylamine	0.0178	1%	Z
A6	Deoxycholic acid. sodium salt	0.24	1%	A
A7	Sodium cholate	0.41	2%	A
A8	Fosmea®-10	0.15	1%	A
A9	Sodium dodecanoyl sarcosine	0.42	2%	A
A10	n-Dodecyl-β-iminodipropionic acid (only used for MDR3)	N/A	1%	A

B6	Decyltrimethylammonium chloride	0.07	1%	C
B7	Dodecyltrimethylammonium chloride	0.0012	1%	C
B8	Hexadecyltrimethylammonium chloride	0.000102	1%	C
B9	Tetradecyltrimethylammonium chloride	0.0009	1%	C

Chapter 6

Transport of bile salts by the human bile salt export pump BSEP (ABCB11) expressed in yeast as an alternative to cell culture-based systems

Published in: *In Preparation*

Impact factor:

Own Proportion

 to this work: 90 %

Expression of BSEP in *P. pastoris*

Preparation of plasma membranes

Performing vesicular uptake assay

Writing of the manuscript

Transport of bile salts by the human bile salt export pump BSEP (ABCB11) expressed in yeast as an alternative to cell culture-based systems

Philipp Ellinger[1], Jan Stindt[2], Sander H. J. Smits[1], Ralf Kubitz[2] and Lutz Schmitt[1,#]

[1]Institute of Biochemistry, Heinrich Heine University, Düsseldorf, Germany
[2]Department of Gastroenterology, Hepatology and Infectiology, University Hospital Düsseldorf, Germany

[#] To whom correspondence should be addressed

Affiliation:

Institute of Biochemistry

Heinrich Heine University Düsseldorf

Universitätstr. 1

40225 Düsseldorf

Germany

Phone: +49-211-81-10773

Fax: +49-211-81-15310

E-Mail: Lutz.Schmitt@hhu.de

Abstract

The human bile salt export pump BSEP (ABCB11) is an ABC transporter, which is responsible for bile salt transport in the liver. Mutations within its gene can lead to different cholestatic diseases and BSEP might also be involved in cholestatic drug-induced liver injury (DILI). To characterize the impact of mutations or drugs as inhibitors on BSEP function, vesicular uptake assays based on inside-out membrane vesicles from human BSEP expressing insect cells are usually performed. Here, we provide the proof-of-principle that human BSEP expressed in the yeast *Pichia pastoris* is functionally active in a vesicular transport assay and displays similar Michaelis-Menten constants for the bile salts taurocholate and taurochenodeoxycholate as compared to the insect or human cell systems. Thus, the *P. pastoris* expression system has to be considered as an alternative to cell-based systems for characterizing vesicular uptake mediated by BSEP.

Keywords: BSEP, *Pichia pastoris*, vesicular uptake assay, Michaelis-Menten constant

Introduction

The ABC transporter bile salt export pump (BSEP, ABCB11) is the major bile salt transporter in vertebrates [1]. It is exclusively expressed in the liver and localized in the canalicular (apical) membrane of hepatocytes lining the canaliculus [2]. Here, BSEP is responsible for the bile salt dependent bile flow. It transports primary and secondary bile salts conjugated to taurine or glycine such as taurocholate (TC), taurochenodeoxycholate (TCDC), taurodeoxycholate (TDC) or glycocholate (GC) and glycochenodeoxycholate (GCDC) [3,4]. The bile salt transport across the canalicular membrane is ATP-dependent [2]. This catalyzes the uphill transport, because bile salt concentration within the canaliculus is approximately 1000 times higher than in the cell [5]. Together with the ABC transporters MDR3 (ABCB4) and ABCG5/G8, which transport phosphatidylcholine (PC) and cholesterol (CL), respectively, BSEP represents the bile forming machinery and bile salts are integrated into mixed micelles composed of PC and CL to dampen their harsh detergent effect [6,7]. Because of its strategic and crucial role in bile formation, BSEP is an important pharmaceutical target. Mutations within the *BSEP* gene can lead to severe cholestatic diseases in humans like progressive intrahepatic cholestasis type 2 (PFIC2) [8], benign recurrent intrahepatic cholestasis type 2 (BRIC2) [9], a milder form of cholestasis or intrahepatic cholestasis of pregnancy (ICP) [10]. Those diseases are manifested by a strongly reduced or even a complete abolishment of bile salt transport. In many cases, PFIC2 mutations lead to the absence of BSEP in the canalicular membrane and thus bile salt concentration rises within the cell and induces for example apoptosis [11,12]. Although different therapies exist, liver transplantation is the ultimate therapy for example for PFIC2 [13]. BSEP is also involved in drug-induced liver injury (DILI). Certain drugs inhibit BSEP *in vitro* and inhibition of BSEP is consistent with the disposition of various drugs to cause

cholestatic DILI in humans [14,15]. To assay BSEP for transport or inhibition under certain conditions, a reliable expression system for BSEP is necessary. Several expression systems for BSEP have been established so far. All of them are based on mammalian cell lines such as HEK293 cells, HepG2 cells, MDCK cells and LLC PK1 or insect cells such as *Sf*9 or *HighFive* cells and BSEP is functionally characterized without purification [3,4,16,17,18,19]. These systems are very convenient to address questions of localization, interactions or degradation of BSEP. Nonetheless, the yield of BSEP produced in those cell lines is not sufficient for purification of the transporter for functional *in vitro* assays and no reports have been published up to now.

Yeasts, like *Pichia pastoris* cells, have the advantage that they harbor the eukaryotic modification machinery, are inexpensive, require simple culture media, and exhibits well–studied genetics [20]. Thus, yeast might be another well-suited opportunity for the heterologous overexpression of human BSEP. Of course there are differences, for example the lipid composition of the plasma membrane is different compared to mammalian cells (they contain ergosterol instead of cholesterol) and the glycosylation pattern differs in respect to the human counterpart. Nevertheless, *P. pastoris* has been used to successfully express 25 of the 48 human ABC transporters [21]. Furthermore, the functional purification of human ABC transporters such as P-gp, MRP3 or ABCG5/G8 has been reported [21,22,23]. Recently, we could show the expression of human BSEP in the yeast *Pichia pastoris* and its subsequent purification for biochemical analysis in detergent solution [24,25].

In general, characterization of BSEP mostly utilized inside-out vesicles (IOVs) of insect cell membranes expressing BSEP to study its transport or inhibition potential without any subsequent purification [26,27]. Like *P. pastoris*, the glycosylation pattern also does not resemble the human glycosylation, but similar kinetic parameters have been determined compared to studies with IOVs from human origin like HEK293 cells

[1]. Thus, glycosylation of BSEP does not influence its transport kinetics, but can only have an influence on trafficking, as shown by mutational studies of N-linked glycosylation sites of BSEP in MDCK cells [28]. In addition to this well-established system, we demonstrate the applicability of IOVs of BSEP expressed in *P. pastoris* to perform transport assays as a proof-of-principle, showing similar kinetic parameters and make this system a new emerging competitor to the insect cell system.

Materials and Methods

Cloning of human BSEP for the *Pichia pastoris* expression system

Human BSEP (NCBI accession code: NM_003742.2) was cloned into the *P. pastoris* expression vector pSGP18 [21] as previously described using homologous recombination in *S. cerevisiae* [24,25].

Transformation and Expression of human BSEP in *Pichia pastoris* X33

The procedure is described in detail in Ellinger *et al*. [24]. In brief, 20 µg of *Pme*I (New England Biolabs) linearized pSGP18-hBSEP plasmid DNA was transformed into electrocompetent *P. pastoris* X33 (Life Technologies) by electroporation (1500 V, 5 ms). Clones were screened for expression in small scale analogous to Wang *et al*. [23]. Expressing clones were fermented in a 15 L table-top glass fermentor (Applikon Biotechnology) according to the Invitrogen *Pichia* fermentation guideline using the basal salt media. Typically a volume of 6 L media was inoculated with 1 L of an overnight culture grown in MGY (1.34% yeast nitrogen base, 1% glycerol and 4×10^{-5}% biotin) media. Aeration was kept above 20% O_2 saturation and the glycerol fed-batch was performed for 5 h feeding ~500 ml of 50% (v/v) glycerol. Protein expression was induced by addition of 3.6 ml/h/l (~1000 ml) methanol for 48 h. Cells were harvested by centrifugation (5,000xg, 10 min, 4°C), flash-frozen in liquid nitrogen and stored at -80°C until further use.

Preparation of yeast plasma membranes (PMs)

PMs of fermented *P. pastoris* cells were prepared as described for *S. cerevisiae* by Ernst *et al*. [29]. Briefly, approximately 30 g of BSEP expressing *P. pastoris* cells were resuspended in 50 mM Tris-acetate, pH 7.5 and 0.2 mM EDTA and disrupted in a bead beater (Biospec) with acid-washed glass beads (Roth). Unbroken cells and

cellular debris were removed by three centrifugation steps (twice 1,000xg, 5 min, 4 °C and 3,000xg, 5 min, 4°C). The supernatant was centrifuged for 45 min at 20,000xg and the membrane pellet was resuspended in 10 mM Tris-acetate, pH 7.5 and 0.2 mM EDTA. The suspension was adjusted to 5 mg/ml with the same buffer and mitochondrial membranes were precipitated by lowering the pH to 5.2 and removed by centrifugation (7,000xg, 5 min, 4°C). The supernatant was re-adjusted to pH 7.5 and PMs were harvested by ultracentrifugation for one hour at 100,000xg. After resuspension with a brush in buffer A (50 mM Tris-HCl, pH 8.0, 75 mM NaCl, 30% glycerol) protein concentration was determined via the Bradford assay (Pierce) and membranes were snap-frozen in liquid nitrogen and stored at -80°C. All steps were carried out at 4°C.

Immunoblotting analysis of BSEP

5 µg of PMs were separated on a 7% SDS-PAGE gel, blotted onto a nitrocellulose membrane and probed with the monoclonal BSEP F-6 antibody (Santa Cruz).

Electron Microscopy of inside-out vesicles

Highly enriched plasma membrane vesicles with a concentration of 1 mg/ml were diluted in transport buffer (50 mM Tris/HCl pH 7.5, 250 mM sucrose) to different ratios (1:100 – 1:10000). Ni 75 pioloform coated grids with a diameter of 3.05 mm were equilibrated with transport buffer and then 2 µl of vesicle suspension were spotted onto the grids. Excess liquid was removed, the grids were washed twice with Millipore water and for negative staining the grids were incubated in uranyl acetate. The dried vesicles were analyzed with an electron microscope (Zeiss) and photographed.

Vesicular transport assay of yeast PMs

P. pastoris PMs were thawed on ice, diluted fourfold in ice-cold transport buffer (50 mM Tris/HCl pH 7.5, 250 mM sucrose) and vesiculated by 10 passages through a 26 gauge needle to yield inside-out vesicles (IOVs). 20 µg of this vesicle preparation was assayed in a reaction volume of 20 µl with 0.1 mg/ml creatine kinase (Sigma-Aldrich), 10 mM creatine phosphate, and 10 mM $MgCl_2$ as an ATP-regenerating system. After addition of $[^3H]$-taurocholate (Perkin Elmer) or $[^3H]$-taurochenodeoxycholate (Hartmann Analytic), reactions were equilibrated on ice for 5 min. Bile salt transport was started by addition of ATP (or buffer for background) with a final concentration of 4 mM at 37°C and stopped by addition of 1 ml of ice-cold transport buffer containing 1 mM of unlabeled TC after 5 min or indicated time points. The stopped reaction was rapidly filtrated through a 0.2 µm cellulose membrane filter (type GSTF; Millipore) prewetted with transport buffer containing 1 mM of unlabeled TC. After filtration, membranes were immediately washed twice with 5 ml transport buffer and once with 5 ml of ice-cold TC-free transport buffer. The radioactivity retained on the filter discs was counted in a liquid scintillation counter (Packard Instruments) using Ultima Gold liquid scintillation cocktail (Perkin Elmer). The BSEP-dependent transport into IOVs was calculated by subtracting the amount of radioactivity in IOVs obtained in the presence of ATP from that obtained in the absence of ATP. The concentration-dependent transport of bile salts in BSEP IOVs was analyzed by incubation with increasing concentrations of bile salts as indicated in figure 4. All the experiments were performed in triplicate.

Results and Discussion

The ABC transporter BSEP is the major player in bile salt circulation in vertebrates. It has been assigned to cholestatic DILI and inheritable cholestatic liver diseases in humans. Most of its biochemical characterization has been performed in cell culture based expression systems, like HEK293 [16], HepG2 [17] and *Sf*9 [3,4] insect cells. Especially, the vesicular transport assay based on inside-out membrane vesicles (IOVs) of insect cells expressing BSEP has been used extensively to characterize the kinetics of transport or to identify BSEP inhibitors [1].

Cloning of the human *BSEP* cDNA is often hampered by its inherent instability in prokaryotic organisms used for cloning [25]. Therefore, the successful usage of other heterologous expression systems is limited. We developed a yeast-based cloning procedure and cloned the human *BSEP* cDNA for the yeast *P. pastoris* expression system and could also show the subsequent purification from *P. pastoris* cells [24]. Besides purification, we also evaluated the potential of *P. pastoris* IOVs containing BSEP for transport studies. Therefore, highly enriched plasma membranes of *P. pastoris* X33 cells either transformed with the empty plasmid (mock control) or the BSEP expression plasmid were prepared and expression of BSEP in plasma membranes could be detected using a monoclonal antibody (Figure 1). The mock sample did not show any immune-reactivity against the antibody.

To test, if highly enriched plasma membranes form vesicles in solution, we employed transmission electron microscopy to visualize the shape of the membranes by negative staining with uranyl acetate. As shown by Figure 2, small round-shaped particles could be observed with a diameter of roughly 20-30 nm. It was assumed that these particles represent vesicles of *P. pastoris* plasma membranes, because grids with no plasma membranes did not show this pattern (data not shown).

Next, we tested time-dependent transport into the lumen of the IOVs for BSEP and the mock sample (Figure 2). Hence, IOVs were incubated with 2 µM of [^{3}H]-TC and reactions were stopped with transport buffer at indicated time points and filtered through a nitrocellulose membrane. A BSEP-dependent TC transport could be observed with a linear rise up to 5 min and then turning into saturation. Mock IOVs also showed the same behavior, but the measured radioactivity remained approximately three fold lower over time. This background transport of the mock sample might be due to other ATP-driven transporters of e.g. the multidrug transporter family, which transport a huge variety of structurally unrelated substances. Furthermore, yeasts contain an ABC transporter for bile salt transport localized in the vacuole [30]. This was one of the rationales to prepare plasma membranes to prevent contamination of this transporter and therefore can be excluded as contaminating reason. Passive diffusion of bile salts should not appear, due to the negative charge of the taurine conjugate. It cannot be excluded that equilibration of TC or TCDC with membranes might be different between the mock and BSEP membrane preparation. Nonetheless, BSEP-containing IOVs showed a specific BSEP-dependent and ATP-dependent transport.

To test, if BSEP IOVs can be used for the determination of kinetic parameters and in subsequent experiments for inhibitor studies or mutational analysis of BSEP, kinetic analysis of radiolabelled TC and TCDC uptake was performed (figure 3A and B). The concentrations used for TC and TCDC ranges from 0 to 50 µM and 0 to 70 µM, respectively. For both bile salts, a concentration-dependent saturation curve could be observed and a Michaelis-Menten constant (K_M) for TC of 22.7 ± 6.2 µM could be determined. For TCDC the constant was 12.7 ± 1.8 µM. The measured K_M-values are in very good agreement with values obtained for human BSEP in cell culture-based expression systems [1]. The K_M-value for TC ranges between 4.0 and 20.0 µM for

BSEP expressed in IOVs in HEK293, *Sf*9 or *HighFive* insect cells [1]. For TCDC the K_M-value has been measured between 4.0 and 13.0 µM in different systems [1]. Although, the V_{max}-values might be lower than in cell-culture systems, this value is dependent on the number of transporter in the IOVs, which differs between different expression systems and membrane preparation procedures. Reproducibility of the obtained results differed slightly between different rounds of fermenter runs. *P. pastoris* was usually fermented in large volumes up to 7 liters to overexpress BSEP. This procedure is quite sensitive to different conditions such as feeding rate of carbon sources, oxygen supply etc. which might in turn influence the heterologous expression of BSEP. Furthermore, the way of preparing membranes also had an influence on BSEP activity. Cell disruption via pressure-based methods like the French press did not result in preparations compared to the quality of the sample derived from cell by glass beads in the transport assay. Most likely the different procedures affected the membrane integrity differently (data not shown). Nevertheless, when using plasma membranes prepared with the bead beater and resuspended with a brush rather with a potter-elvehjem homogenisator (potter) (Figure 5), results were comparable and reproducibility was enhanced.

In summary, IOVs containing BSEP prepared from *P. pastoris* cells are a valuable alternative to established systems for transport studies of BSEP. Thus, with this system it is possible to characterize clinically relevant mutants as well as inhibitors, which might cause DILI.

Acknowledgements

References

1. Stieger B (2011) The role of the sodium-taurocholate cotransporting polypeptide (NTCP) and of the bile salt export pump (BSEP) in physiology and pathophysiology of bile formation. Handbook of experimental pharmacology: 205-259.
2. Gerloff T, Stieger B, Hagenbuch B, Madon J, Landmann L, et al. (1998) The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver. The Journal of biological chemistry 273: 10046-10050.
3. Byrne JA, Strautnieks SS, Mieli-Vergani G, Higgins CF, Linton KJ, et al. (2002) The human bile salt export pump: characterization of substrate specificity and identification of inhibitors. Gastroenterology 123: 1649-1658.
4. Noe J, Stieger B, Meier PJ (2002) Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 123: 1659-1666.
5. Hofmann AF (1999) Bile Acids: The Good, the Bad, and the Ugly. News in physiological sciences : an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society 14: 24-29.
6. Trauner M, Boyer JL (2003) Bile salt transporters: molecular characterization, function, and regulation. Physiological reviews 83: 633-671.
7. Ellinger P, Kluth M, Przybylla S, Smits SH, Schmitt L Structure and function of hepatic ABC transporters. in: D Häussinger, V Keitel, R Kubitz (Eds), Hepatobiliary Transport in Health and Disease , de Gruyter, Berlin, 2012, pp 23–41.
8. Bull LN, Carlton VE, Stricker NL, Baharloo S, DeYoung JA, et al. (1997) Genetic and morphological findings in progressive familial intrahepatic cholestasis (Byler disease [PFIC-1] and Byler syndrome): evidence for heterogeneity. Hepatology 26: 155-164.
9. van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, et al. (2004) Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 127: 379-384.
10. Dixon PH, van Mil SW, Chambers J, Strautnieks S, Thompson RJ, et al. (2009) Contribution of variant alleles of ABCB11 to susceptibility to intrahepatic cholestasis of pregnancy. Gut 58: 537-544.
11. Keitel V, Burdelski M, Warskulat U, Kuhlkamp T, Keppler D, et al. (2005) Expression and localization of hepatobiliary transport proteins in progressive familial intrahepatic cholestasis. Hepatology 41: 1160-1172.
12. Strautnieks SS, Byrne JA, Pawlikowska L, Cebecauerova D, Rayner A, et al. (2008) Severe bile salt export pump deficiency: 82 different ABCB11 mutations in 109 families. Gastroenterology 134: 1203-1214.
13. Jacquemin E (2012) Progressive familial intrahepatic cholestasis. Clinics and research in hepatology and gastroenterology 36 Suppl 1: S26-35.
14. Dawson SE, Stahl S, Paul N, Barber J, Kenna JG (2011) In Vitro Inhibition of the Bile Salt Export Pump Correlates with Risk of Cholestatic Drug Induced Liver Injury in Man. Drug metabolism and disposition: the biological fate of chemicals.
15. Warner DJ, Chen H, Cantin LD, Kenna JG, Stahl S, et al. (2012) Mitigating the inhibition of human bile salt export pump by drugs: opportunities provided by physicochemical

property modulation, in silico modeling, and structural modification. Drug metabolism and disposition: the biological fate of chemicals 40: 2332-2341.

16. Hayashi H, Takada T, Suzuki H, Onuki R, Hofmann AF, et al. (2005) Transport by vesicles of glycine- and taurine-conjugated bile salts and taurolithocholate 3-sulfate: a comparison of human BSEP with rat Bsep. Biochimica et biophysica acta 1738: 54-62.

17. Kubitz R, Sutfels G, Kuhlkamp T, Kolling R, Haussinger D (2004) Trafficking of the bile salt export pump from the Golgi to the canalicular membrane is regulated by the p38 MAP kinase. Gastroenterology 126: 541-553.

18. Hayashi H, Takada T, Suzuki H, Akita H, Sugiyama Y (2005) Two common PFIC2 mutations are associated with the impaired membrane trafficking of BSEP/ABCB11. Hepatology 41: 916-924.

19. Mita S, Suzuki H, Akita H, Hayashi H, Onuki R, et al. (2006) Vectorial transport of unconjugated and conjugated bile salts by monolayers of LLC-PK1 cells doubly transfected with human NTCP and BSEP or with rat Ntcp and Bsep. American journal of physiology Gastrointestinal and liver physiology 290: G550-556.

20. Ramon A, Marin M (2011) Advances in the production of membrane proteins in Pichia pastoris. Biotechnology journal 6: 700-706.

21. Chloupkova M, Pickert A, Lee JY, Souza S, Trinh YT, et al. (2007) Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 46: 7992-8003.

22. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718-1722.

23. Wang Z, Stalcup LD, Harvey BJ, Weber J, Chloupkova M, et al. (2006) Purification and ATP hydrolysis of the putative cholesterol transporters ABCG5 and ABCG8. Biochemistry 45: 9929-9939.

24. Ellinger P, Kluth M, Stindt J, Smits SH, Schmitt L (2013) Detergent Screening and Purification of the Human Liver ABC Transporters BSEP (ABCB11) and MDR3 (ABCB4) expressed in the yeast Pichia pastoris. PloS one in press.

25. Stindt J, Ellinger P, Stross C, Keitel V, Haussinger D, et al. (2011) Heterologous overexpression and mutagenesis of the human bile salt export pump (ABCB11) using DREAM (Directed REcombination-Assisted Mutagenesis). PloS one 6: e20562.

26. Heredi-Szabo K, Kis E, Krajcsi P (2012) The vesicular transport assay: validated in vitro methods to study drug-mediated inhibition of canalicular efflux transporters ABCB11/BSEP and ABCC2/MRP2. Current protocols in toxicology / editorial board, Mahin D Maines Chapter 23: Unit 23 24.

27. van Staden CJ, Morgan RE, Ramachandran B, Chen Y, Lee PH, et al. (2012) Membrane vesicle ABC transporter assays for drug safety assessment. Current protocols in toxicology / editorial board, Mahin D Maines Chapter 23: Unit 23 25.

28. Mochizuki K, Kagawa T, Numari A, Harris MJ, Itoh J, et al. (2007) Two N-linked glycans are required to maintain the transport activity of the bile salt export pump (ABCB11) in MDCK II cells. American journal of physiology Gastrointestinal and liver physiology 292: G818-828.

29. Ernst R, Kueppers P, Klein CM, Schwarzmueller T, Kuchler K, et al. (2008) A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. Proceedings of the National Academy of Sciences of the United States of America 105: 5069-5074.

30. Ortiz DF, St Pierre MV, Abdulmessih A, Arias IM (1997) A yeast ATP-binding cassette-type protein mediating ATP-dependent bile acid transport. The Journal of biological chemistry 272: 15358-15365.

Figure Legends

Figure 1: Immunoblot analysis of purified plasma membranes (PMs) of *P. pastoris* X33 cells. X33 was either transformed with the empty pSGP18 plasmid (left lane, mock) or the expression plasmid with the *BSEP* cDNA (right lane, BSEP). 5 μg of membranes were separated on a 7% SDS-PAGE gel, blotted onto nitrocellulose and probed with a monoclonal BSEP antibody.

Figure 2: Electron microscopy picture of plasma membrane vesicles from *P. pastoris* expressing human BSEP. Round-shaped membrane vesicles are exemplified with arrows. The bar represents 100 nm.

Figure 3: Time dependent TC transport into inside-out vesicles (IOVs) prepared from mock cells (open squares) or BSEP expressing cells (black dots). IOVs were incubated with 2 μM [^{3}H]-TC in the presence of ATP (4 mM) or transport buffer, 10 mM MgCl$_2$ and an ATP regeneration system as described in Materials and Methods. Shown is the difference of the radioactivity obtained in the presence of ATP and buffer for each indicated time point ± SEM.

Figure 4: Kinetic analysis on BSEP-containing IOVs for [^{3}H]-TC and [^{3}H]-TCDC. The concentration ranges from 0 to 50 μM and 0 to 70 μM, respectively. Kinetic parameters for [^{3}H]-TC are K$_M$ of 22.7 ± 6.3 μM and a V$_{Max}$ of 58.8 ± 7.3 pmol / (mg x min). [^{3}H]-TCDC is transported with a K$_M$-value of 12.9 ± 1.8 μM and a V$_{Max}$ of 93.8 ± 4.0 pmol / (mg x min). Shown is the difference of the radioactivity obtained in the

presence of ATP and buffer for each indicated concentration ± SEM. Data were fitted

to Michaelis-Menten kinetics using the following equation: $v_0 = \frac{v_{max} \cdot S}{K_M + S}$.

Figure 5: Transport of [^{3}H]-TC by IOVs containing human BSEP prepared from the

same membrane preparation either resuspended wit a brush or a potter-elvehjem

homogenisator (potter). Transport was performed for 5 minutes and measured in

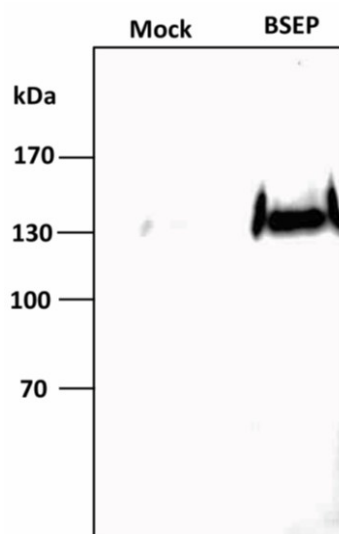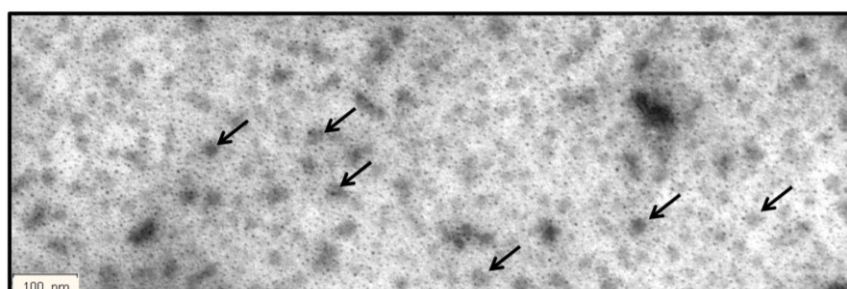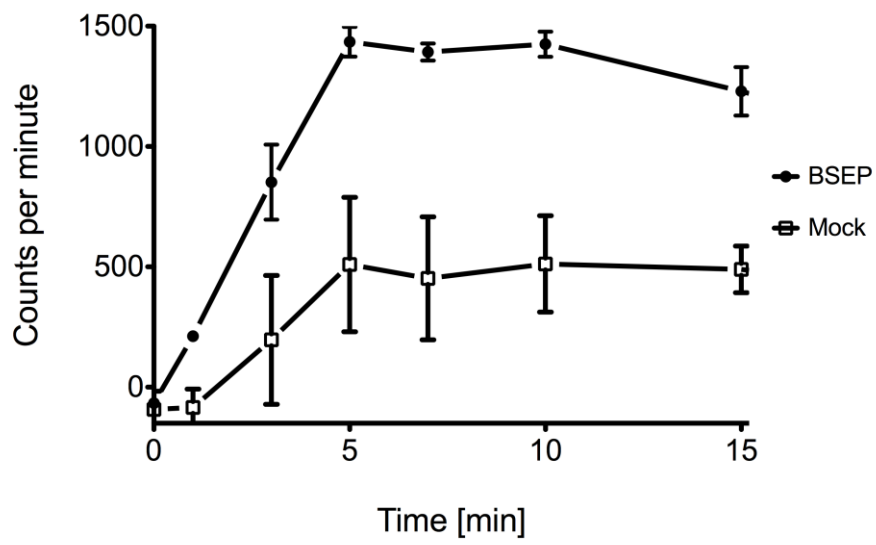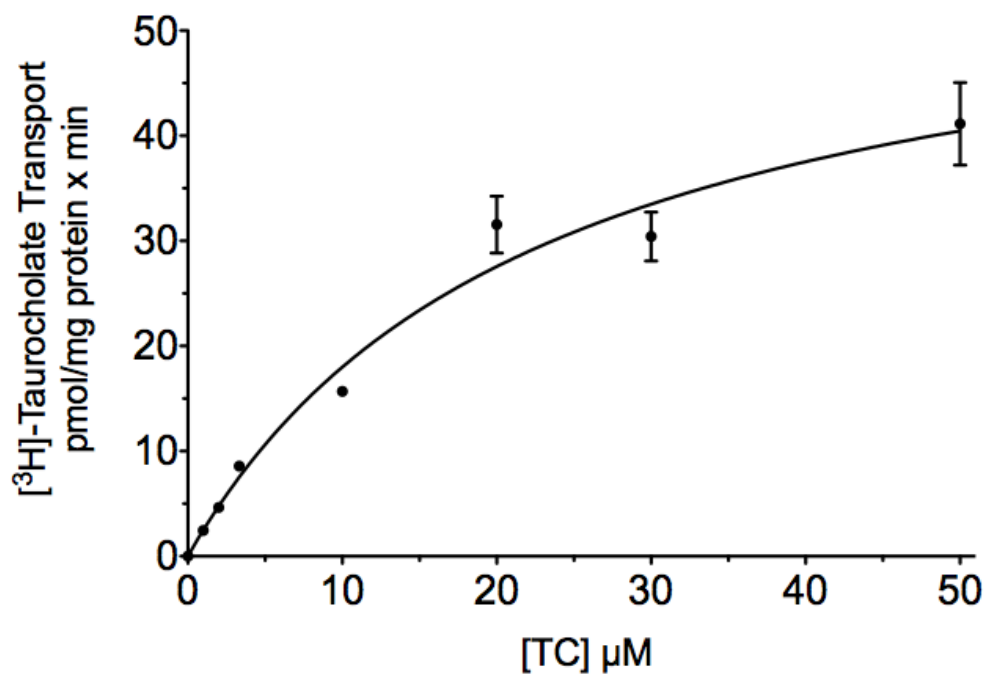triplicate (n=3, ± SEM).

Figures

Figure 1

Figure2

Figure 3

Figure 4

A

B

Figure 5

A novel mutation within a transmembrane helix of the bile salt export pump (BSEP, ABCB11) with delayed development of liver cirrhosis

A novel mutation within a transmembrane helix of the bile salt export pump (BSEP, ABCB11) with delayed development of liver cirrhosis
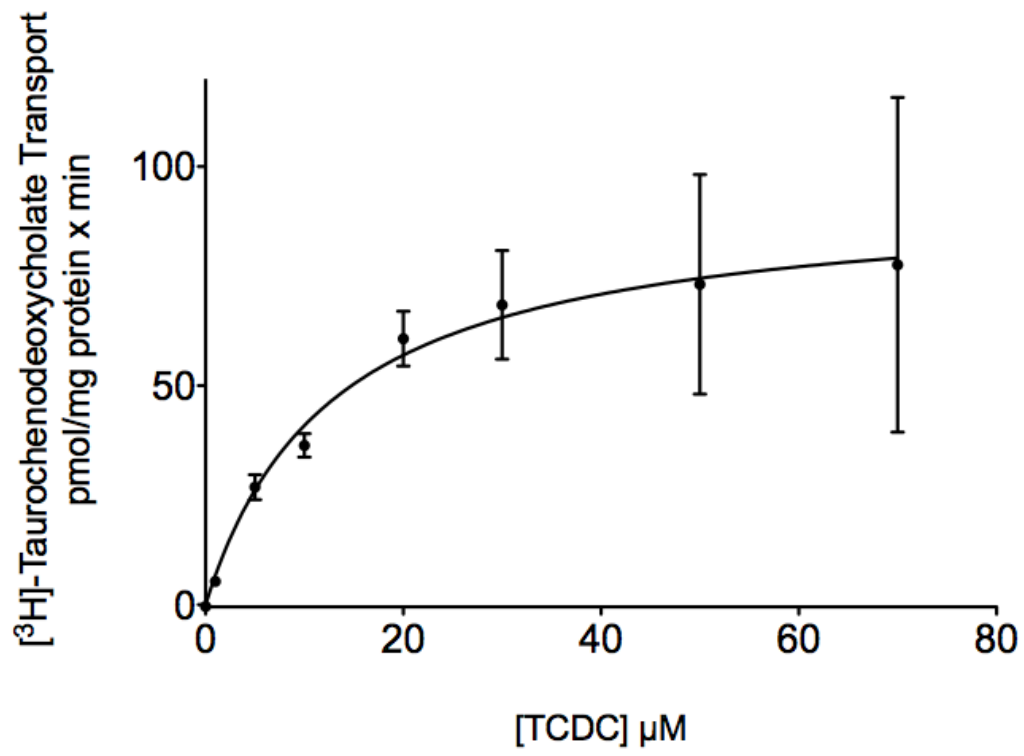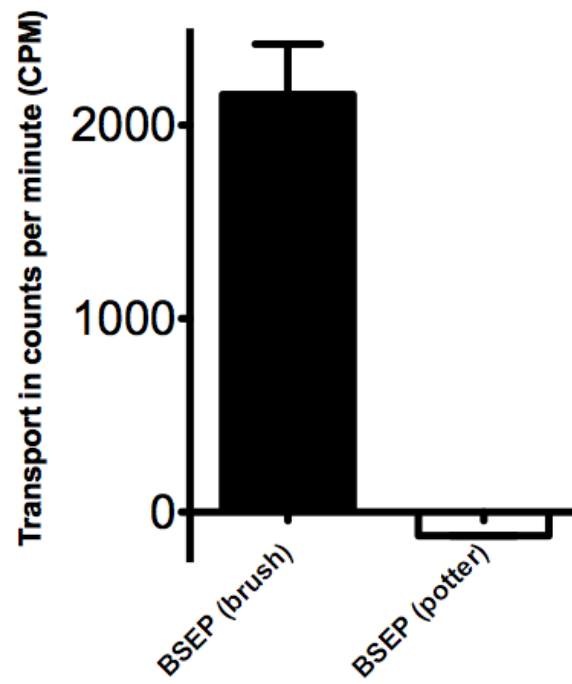
A novel mutation within a transmembrane helix of the bile salt export pump (BSEP, *ABCB11*) with delayed development of liver cirrhosis

Jan Stindt[1,2], Philipp Ellinger[2], Katrin Weissenberger[1], Carola Dröge[1], Diran Herebian[3], Ertan Mayatepek[3], Bernhard Homey[4], Stephan Braun[4], Jan Schulte am Esch[5], Michael Horacek[6], Ali Canbay[6], Lutz Schmitt[2], Dieter Häussinger[1] and Ralf Kubitz[1]

Short Title: First "BRIC-2" mutation in a transmembrane helix

Authors Affiliations

Medical Faculty of the Heinrich-Heine University, Düsseldorf, Germany

[1] Department of Gastroenterology, Hepatology and Infectiology, University Hospital, Düsseldorf

[3] Department of General Pediatrics and Neonatology, University Hospital, Düsseldorf

[4] Dermatological Clinic, University Hospital, Düsseldorf

[5] Department of General-, Visceral-, and Pediatric Surgery, University Hospital, Düsseldorf

[2] Institute of Biochemistry, Heinrich-Heine-University Düsseldorf

[6] Department of Gastroenterology and Hepatology, University of Essen Medical School, Essen, Germany

Authors Contributions

Corresponding Author

Ralf Kubitz, MD; kubitz@med.uni-duesseldorf.de
Department of Gastroenterology, Hepatology and Infectiology
Heinrich-Heine-University, Düsseldorf
Moorenstrasse 5
D-40225 Düsseldorf
Germany
Fax: (49)-211-8117517
Tel: (49)-211-8119648

Electronic word count

221 words (abstract)

5558 words (entire text, including title page, legends, and references)

Five Figures

Abbreviations

BRIC	benign recurrent intrahepatic cholestasis
BSEP	bile salt export pump (*ABCB11*)
CA	cholic acid
CDCA	chenodeoxycholic acid
DCA	deoxycholic acid
FIC1	familial intrahepatic cholestasis 1 gene product (*ATP8B1*)
LCA	lithocholic acid
MDR3	multidrug resistance protein 3 (*ABCB4*)
PFIC	progressive familial intrahepatic cholestasis
TC	taurocholate
TCDC	taurochenodeoxycholate
UDCA	ursodeoxycholic acid

Abstract

Background/Aims: The bile salt export pump (BSEP, *ABCB11*) is essential for bile salt secretion at the canalicular membrane of liver cells. Clinical phenotypes associated with BSEP mutations are commonly categorized as benign recurrent intrahepatic cholestasis (BRIC-2) or progressive familial intrahepatic cholestasis (PFIC-2).

Methods: The molecular basis of BSEP-associated liver disease in a sibling pair was characterized by immunostaining, gene sequencing, bile salt analysis, and recombinant expression in mammalian cells and yeast for localization and *in vitro* activity studies, respectively.

Results: BRIC-2 was considered in a brother and sister who both suffered from intermittent cholestasis since childhood. Gene sequencing of *ABCB11* identified the novel missense mutation p.G374S, which is localized in the putative sixth transmembrane helix of BSEP. Liver fibrosis was present in the brother at the age of 18 with progression to liver cirrhosis within 3 years. Immunofluorescence of liver tissue showed clear canalicular BSEP expression; however, biliary concentration of bile salts was drastically reduced. In line with these *in vivo* findings, HEK293 cells showed regular membrane targeting of human BSEP^{G374S}, whereas *in vitro* transport measurements revealed a strongly reduced transport activity.

Conclusions: The novel mutation p.G374S impairs transport function without disabling membrane localization of BSEP. While all other known BSEP mutations within transmembrane helices are associated with PFIC-2, the new p.G374S mutation causes a transitional phenotype between BRIC-2- and PFIC-2.

Keywords

Bile salt export pump; BSEP; *ABCB11*; Progressive familial intrahepatic cholestasis;

PFIC; benign recurrent intrahepatic cholestasis; BRIC; ABC transporter

Introduction

BSEP is the major canalicular bile salt transporter in humans and is responsible for the bile salt-dependent bile flow. It mainly transports monovalent bile salts including taurine and glycine conjugates of primary bile acids (cholic and chenodeoxycholic acid) as well as secondary bile acids (deoxycholic acid) and ursodeoxycholic acid [1]. "Severe" mutations of BSEP cause progressive familial intrahepatic cholestasis (PFIC) type 2 [2]. PFIC-2 is similar in presentation to PFIC-1, which is caused by mutations of the *ATP8B1* gene [3]. *ATP8B1* encodes the "familial intrahepatic cholestasis 1" protein (FIC1), a P-type ATPase involved in the flipping of aminophospholipids from the outer to the inner leaflet of the canalicular membrane [4]. Some mutations of *ATP8B1* and *ABCB11* cause milder forms of cholestatic liver diseases, which are termed benign recurrent intrahepatic cholestasis (BRIC) type 1 and 2 [5-7], respectively. Here we report on a novel BSEP mutation identified in a sibling pair that is initially associated with a typical BRIC-2 phenotype with intermittent cholestasis but eventually progresses towards liver cirrhosis. We attribute its severity to its localization close to the putative translocation pore within one of the transmembrane α-helices, which commonly harbour PFIC-2 mutations.

Materials and methods

Analysis of serum bile salts

Bile salt analysis was performed by tandem mass spectroscopy. The system consisted of an HPLC Waters Alliance 2795 separation module (Waters, Milford, UK) coupled to a Quattro Micro triple quadrupole mass spectrometer (Micro Mass, Manchester, UK). Electrospray ionization was performed in the negative ionization mode. Chromatographic separation was performed on an analytical HPLC Phenomenex Luna C18 column coupled to guard column Phenomenex Gemini C18. The mobile phase consisted of water containing 0.01 % formic acid and 5 mM ammonium acetate (Eluent A) and methanol (Eluent B). Sample elution was run isocratically in a ratio of 18:82 (A:B, v/v) over 15 min at a flow rate of 0.2 ml/min. Unconjugated bile acids were detected unfragmented. Taurine- and glycine-conjugated bile salts were analyzed using their specific fragment ions at mass-to-charge (m/z) 80 and 74. Analytes were detected in the multiple reaction mode. Quantification analysis was performed using standard calibration curves in five different concentrations (0.1, 0.5, 1.0, 2.5, 5.0 μg/ml) including the deuterated internal standard (d4-CA). Concentrations were calculated using the QuanLynx (MassLynx 4.1, Waters) software. A solid phase extraction (Chromabond HR-X cartridge, Macherey-Nagel, Düren, Germany) procedure was applied for serum samples prior to HPLC-MS/MS analysis. Normal bile samples were collected during surgery from patients who underwent partial liver resection for hepatic metastasis due to colorectal or breast cancer and whose livers were otherwise healthy. Their bile ducts were canulated before routine removal of the gall bladder.

Immunofluorescence of liver tissue and transfected cells

Immunofluorescence staining and microscopy were performed as described recently [7, 8] and by the use of the transporter specific antibodies K24 for BSEP [9], P3II26 for MDR3 (Thermo Scientific, Schwerte, Germany) and A276 for the Na$^+$/K$^+$-ATPase (clone M7-PB-E9, Sigma Aldrich, Munich, Germany).

Cloning and mutagenesis of human BSEP

For large-scale expression in the methylotrophic yeast *Pichia pastoris*, the BSEP coding sequence was PCR-amplified from the mammalian expression vector pEYFP-N1-hBSEP [10] using primers BSEP-HR-PP-S1 (5´-ATCAAAAAACAACTAATTATTCGAACGAGGTAAAAGAATGTC TGACTCAGTAATTCTTCGAAGTATA-3´) and BSEP-HR-PP-S2 (5´-ACGTTTGGACCTTGGAAAAG ACTTCTAAGGAGTTGGAGGCACTGATGGGGGATCCAGTGGTGACTAGTTT-3´).

The amplified coding sequence was cloned into a yeast-compatible derivative of the custom *P. pastoris* expression vector pSGP18 [11] (kindly provided by M. Dumont, Rochester) via homologous recombination as described in Stindt *et al.* [12]. The integrity of the coding sequence on the resulting plasmid pSGP18-2µ-BSEP^{WT} was verified by DNA sequencing of the entire gene. In addition to the removal of the cryptic prokaryotic promoter motif (a TATA box motif was removed by the silent modifications c.78T>C and c.81T>C), four silent nucleotide exchanges were found to be present with respect to the BSEP reference sequence NM_003742.2, with the first nucleotide of the start codon being numbered as 1: c.2064G>A, c.2559C>T, c.3084A>G, and c.3390T>C as described [10].

For transfection experiments, a yeast-compatible version (pEYFP-N1-OriLeu-BSEP) of the mammalian expression vector pEYFP-N1-hBSEP was generated as described

recently [12]. The mutation c.1120G>A (p.G374S) was introduced into the two plasmids pSGP18-2µ-BSEP and pEYFP-N1-OriLeu-BSEP by DREAM mutagenesis [12]. The mutagenesis primers used were BSEP-G374S-HR-S1 (5´-CTTTAAATCTTAGCAATGCCTCTCCTTGTTTGGAAGCCTTTGCAACT-3´) and BSEP-G374S-HR-S2 (5´-AGGCATTGCTAAGATTTAAAGCTCCTACTATGACAC-TGAGGAAAATCTG-3´). The plasmid pEYFP-N1-OriLeu-BSEP containing the p.G374S variant was found to have one additional silent nucleotide exchange with respect to the parent construct (c.1246G>A).

Expression of human BSEP in *Pichia pastoris*

Electrocompetent *P. pastoris* strain X33 cells (Invitrogen, Karlsruhe, Germany) were transformed with the linearized pSGP18-2µ-BSEP^WT, pSGP18-2µ-BSEP^G374S, or the empty pSGP18-2µ construct according to the manufacturer´s instructions. Positive clones were selected on YPDS plates containing 100 µg/ml Zeocin. Small-scale expression screens of the obtained *P. pastoris* clones were performed as described by Lerner-Marmorosh *et al*. [13]. For large-scale expression, BSEP-expressing clones were fermented in a 15 litre table-top glass fermentor (Applikon Biotechnology, Schiedam, Netherlands) according to the Invitrogen *Pichia* fermentation guidelines. Protein expression was induced by constant methanol feeding for 48 h. Cells were then harvested by centrifugation at 4°C for 10 min at 5,000xg, snap-frozen in liquid nitrogen and stored at -80°C until further use.

Analysis of BSEP^G374S expression in mammalian cells

Wildtype BSEP-EYFP as well as BSEP^G374S-EYFP were transiently transfected into HEK293 (human embryonic kidney) cells by the use of X-treme Gene HP (Roche Applied Science, Mannheim, Germany) according to the manufacturer's guidelines.

After transfection cells were grown for 48 h, fixed with methanol (-20°C, 30 sec) and were immunostained for the Na^+/K^+-ATPase. The secondary antibody was goat anti-mouse-Cy3 (1:500; Dianova, Hamburg, Germany). Cells were visualized with a LSM 510 Meta confocal laser scanning microscope (Zeiss, Jena, Germany). The excitation wavelength was 488 nm for the enhanced yellow fluorescent protein (EYFP) and 543 nm for Cy3. Emission was detected by 505-530 nm (green) and 560-615 nm band pass filters (red).

Plasma membrane preparation from *P. pastoris* yeast cells

Plasma membranes of fermented *P. pastoris* cells were prepared as described for *S. cerevisiae* by Ernst *et al.*[14]. All steps were carried out at 4°C and in the presence of protease inhibitors (Complete[TM] without EDTA, Roche Applied Science, Mannheim, Germany). Briefly, approximately 30 g of respective *Pichia* cells were resuspended in 50 mM Tris-acetate, pH 7.5 and 0.2 mM EDTA and disrupted in a bead beater with acid-washed glass beads. Unbroken cells and cellular debris were removed by two centrifugation steps at 1,000xg and one step at 3,000xg for 5 min each. The supernatant was centrifuged for 45 min at 20,000xg, while the membrane pellet was resuspended in 10 mM Tris-acetate, pH 7.5 and 0.2 mM EDTA. After determination of protein concentration with the Coomassie Plus Protein Assay reagent (Thermo Scientific), the suspension was adjusted to 5 mg/ml with the same buffer and mitochondrial membranes were precipitated by lowering the pH to 5.2 and removed by centrifugation at 7,000xg for 5 min. The supernatant was re-adjusted to pH 7.5 and plasma membranes were harvested by ultracentrifugation for one hour at 140,000xg. After resuspension in buffer A (50 mM Tris-HCl, pH 8.0, 75 mM NaCl, 30 % glycerol) protein concentration was determined and membranes were snap-frozen in liquid nitrogen and stored at -80°C.

Western Blotting Analysis of BSEP

5 µg of *P. pastoris* plasma membrane proteins were separated on a 7 % SDS polyacrylamide gel and transferred to a nitrocellulose membrane (PALL, Dreieich, Germany). For comparative expression analysis of wildtype and mutant BSEP, HEK293 cells were transiently transfected with 10 µg of the appropriate BSEP expression plasmids. After 48 hours they were lysed with 1 % SDS in PBS (4°C) and were incubated for 10 min at RT. In order to reduce viscosity 0.5 µl of Benzonase (Merck, Darmstadt, Germany) were added per 100 µL of lysate. 10 µg of these lysates were used for SDS-PAGE and Western blotting. Membranes were probed with the monoclonal BSEP antibody F6 (Santa Cruz Biotechnology, Heidelberg, Germany).

[^{3}H]-taurocholate transport assays

Transport activity of human BSEP was measured as described [15]. *Pichia* plasma membranes were thawed on ice, diluted fourfold in ice-cold buffer (50 mM Tris/HCl pH 7.5, 250 mM sucrose) and vesiculated by 20 passages through a 26 gauge needle. 20 µg of this vesicle preparation were assayed in a reaction volume of 20 µl with 0.1 mg/ml rabbit muscle creatine kinase (Sigma-Aldrich, Munich, Germany), 10 mM creatine phosphate, and 10 mM $MgCl_2$ as an ATP-regenerating system. After addition of [^{3}H]-taurocholate (Perkin Elmer, Rodgau, Germany) or [^{3}H]-taurochenodeoxycholate (Hartmann Analytic, Braunschweig, Germany) to a final concentration of 2 µM, reactions were allowed to equilibrate on ice for 5 min. Bile salt transport was then initiated by addition of ATP to a final concentration of 4 mM. Control reactions contained buffer instead of ATP. The reactions were then immediately shifted to 37°C for 5 min. Bile salt transport was stopped by addition of 1 ml of ice-cold buffer containing 1 mM of unlabelled taurocholate (TC), followed by

rapid filtration through 0.2 µm cellulose membranes (type GSTF; Millipore, Schwalbach, Germany) prewetted with the same buffer. After filtration, the membranes were immediately washed twice with 5 ml of the same, ice-cold buffer and once with 5 ml of ice-cold TC-free buffer. The radioactivity retained on the filter discs was counted in a liquid scintillation counter (Packard Instruments, Frankfurt, Germany) using Ultima Gold liquid scintillation cocktail (Canberra Packard, Frankfurt, Germany).

Case description and results

A 22-year-old woman (=sister) of Moroccan origin presented with intermittent jaundice for several years and recurrent pruritus of increasing intensity for five months. Her body mass index was 12.5 kg/m^2. Erythematous macules on the back and abdomen with discrete desquamation accompanied by generalized xerosis cutis were observed. Alanine aminotransferase (ALT) was mildly elevated with 64 U/l (normal: <45 U/l), alkaline phosphatase (AP) was 132 U/l (<108 U/l), while aspartate aminotransferase (AST) and gamma-glutamyltransferase (γGT) were not increased. Due to the history of recurrent jaundice, a hepatic cause of the pruritus was taken into account. While anti-nuclear antibodies (ANA) were detected at a titer of 1:80, all other parameters related to liver diseases (anti-mitochondrial antibodies, anti-smooth muscle antigen antibodies, liver kidney microsomal antibodies, anti-neutrophil cytoplasmatic antibodies, coeruloplasmin, α1-antitrypsin, ferritin, serology for hepatitis B and C and phorphyrins) were within their normal ranges.

On abdominal ultrasound the liver appeared normal. A single bile stone was detected. Liver stiffness was normal at 6.2 kPa as determined by Fibroscan^{TM} [16]. To control pruritus, the patient was topically treated with polidocanol-containing emollients and received narrow band ultraviolet B (UVB, 311 nm) therapy up to a

dose of 0.26 J/cm^2 with limited improvement of symptoms. Finally, the patient received ursodesoxycholic acid (UDCA, 1000 mg/d) as well as cholestyramin (12 g/d). Relief of pruritus was only moderate at the beginning but improved consecutively due to continued intake of UDCA.

Her brother experienced intermittent jaundice and pruritus since the first year of life and presented with increased serum bile salt levels. Increased levels of aminotransferases completely normalized between the pruritic episodes whereas γGT levels were normal at all times, leading to the diagnosis of benign recurrent intrahepatic cholestasis (BRIC). Symptoms were moderately controlled by UDCA, which was taken sporadically during longer periods. He presented again at the age of 18 due to progressive jaundice. Bilirubin reached a level of 10.6 mg/dl (normal: <1.1 mg/dl). Liver stiffness was 11.6 kPa (Fibroscan^{TM}, normal range <7.5 kPa) suggesting the presence of fibrosis. In line with this, a liver biopsy showed portal fibrosis stage 1, ductopenia and cytokeratin 7 positive metaplasia of hepatocytes.

Three years later, symptoms deteriorated, liver stiffness was 25.7 kPa and incomplete liver cirrhosis, signs of chronic cholestasis, ductular proliferation and hepatocellular injury were detected in a subsequent liver biopsy. Four months later, endoscopic retrograde cholangioscopy was performed due to progressive jaundice. Bile and serum were collected for analysis of bile salt profiles. At that point, liver stiffness was 39.1 kPa, indicating advanced liver cirrhosis [16, 17].

Gene sequencing of *ABCB11* reveals a novel BSEP mutation

Based on the history of recurrent pruritus with intermittent jaundice, the positive family history and consanguinity of the parents, an inherited form of cholestasis was considered. In view of low γGT levels despite high serum bile salt concentrations

mutations of the familial intrahepatic cholestasis 1 (FIC1) gene product (gene symbol: *ATP8B1*) or of the bile salt export pump (BSEP/*ABCB11*) were the most likely cause of cholestasis in both patients. Because bile stones are much more common in patients with BSEP mutations [6], the entire coding sequence and flanking intronic regions of *ABCB11* were sequenced (informed consent was obtained from all family members). In comparison to the reference sequence NM_003742.2 nine homozygous genetic variants were detected in the siblings. Six intronic variants (c.77-17delA, c.99-18T>C, c.477+16G>A, c.909-17G>A, c.909-15A>G, c.3766-34G>A) were all located at longer distances from splice sites. Furthermore, two synonymous exonic variants (c.402C>T; p.I134I and c.957A>G; p.G319G) were found. Most importantly, a homozygous genetic variant was detected at position 1120 at the level of coding DNA. Here, a guanosine was replaced by an adenine (c.1120G>A), leading to an amino acid substitution from glycine to serine (p.Gly374Ser) at the protein level (Fig. 1). The parents, both of whom were symptom-free were heterozygous for this and all other variants. Likewise, the younger, healthy sister was heterozygous for p.G374S. In a control cohort of 6 Moroccans and of 148 unrelated persons p.G374S was not detected.

The p.G374S mutation does not impair BSEP trafficking to the plasma membrane

In order to investigate the pathophysiological relevance of this newly discovered mutation, a liver biopsy (taken at the age of 21) of the brother was stained with transporter-specific antibodies. The phospholipid floppase MDR3 (*ABCB4*) and the bile salt export pump BSEP were both detected at the canalicular membrane of hepatocytes as in healthy livers (Fig. 2A). In order to analyze intracellular trafficking of mutated BSEP the mutation p.G374S was introduced into a plasmid encoding

BSEP fused to the enhanced yellow fluorescent protein (EYFP) at the C-terminus. After transfection into HEK293 cells both the wildtype and the mutant protein were targeted to the plasma membrane (Fig. 2B) as compared to the Na^+/K^+-ATPase as a plasma membrane marker. Western blot analysis showed an equal expression level of wildtype and mutated BSEP in HEK293 cells (Fig. 2C). In line with the finding from the liver biopsy, these results clearly demonstrate that the amino acid exchange in BSEP[G374S] has no major impact on intracellular trafficking of BSEP[G374S] *in vivo* and *in vitro*.

Bile salt analysis confirms disturbed bile salt secretion

Bile acids (CA, CDCA, DCA, UDCA and LCA) and their taurine and glycine conjugates were analyzed from sera of the sister, her brother and 40 controls as well as from the bile of the brother and 7 persons with normal liver function by HPLC-MS/MS (high performance liquid chromatography coupled to tandem mass spectrometry). Although bilirubin concentration in the sister's serum was only 0.66 mg/dl (normal <1 mg/dl), serum bile salts were significantly elevated to 28.2 µM and in the brother's serum to 47.8 µM (normal: 1.4 ± 1.2 µM, n=40). Relative amounts of derivatives of cholic (CA) and chenodeoxycholic acid (CDCA) were within normal ranges as compared to controls (Fig. 3). Deoxycholate, which is derived from CA by bacterial dehydroxylation within the large bowel, was slightly elevated in the sister's serum (0.82 µM; normal 0.27 ± 0.20 µM) but absent in the brother's serum. Absolute and relative amounts of unconjugated bile acids were decreased. Most notably, total bile salt concentration in the bile of the brother was as low as 0.061 mM as compared to 35.1 ± 18.0 mM in patients with healthy livers (n=7). Thus, the brother's liver was unable to generate a sufficient bile salt gradient between serum and bile while showing a normal canalicular pattern of BSEP expression.

Bile salt transport by BSEP[G374S] is strongly impaired

According to a homology model of BSEP the mutation is located in the interior of the first transmembrane domain near the putative translocation channel (Fig. 4A-C). In order to analyse the functional consequence of this amino acid substitution, bile salt transport was measured in plasma membrane vesicles from *Pichia pastoris* expressing either BSEP[WT] or BSEP[G374S] (Fig. 4D). Vesicles from *P. pastoris* cells transformed with the empty vector (=control) showed unspecific binding of [3H]-TC or [3H]-TCDC; however, this was not significantly increased upon addition of ATP, excluding energy-dependent bile salt transport in these vesicles. When ATP was added to vesicles with BSEP[WT], [3H]-TC transport increased 1.65-fold (n=3) compared to TC-transport in the absence of ATP. Likewise, transport of [3H]-TCDC increased 1.55-fold (n=3) upon addition of ATP. [3H]-TC as well as [3H]-TCDC transport was strongly reduced in BSEP[G374S]-containing vesicles (Fig. 4D). In summary, the data indicate that the amino acid substitution p.G374S severely impairs bile salt transport by BSEP.

G374S is the first BRIC-associated mutation in a transmembrane helix of BSEP

In order to relate the p.G374S mutation to its localization within a transmembrane helix, all available missense mutations of BSEP were searched from databases and literature as reviewed recently [18]. 112 missense mutations related to PFIC-2 and 21 mutations associated with BRIC-2 were found. The HMMTOP 2.0 topology tool [19] predicts 12 putative transmembrane helices (TMH) for human BSEP. Interestingly, none of the previously identified 21 BRIC-2 mutations are localized within a TMH in this model, whereas 20 percent (i.e., 22 of 112) of the PFIC-2 mutations affect the TMHs of BSEP (Fig. 5). Accordingly, p.G374S is the first BRIC-2-associated mutation

that influences a transmembrane helix, a domain otherwise exclusively affected by PFIC-2 mutations.

Discussion

In this study we describe the molecular basis of the novel BSEP mutation p.G374S, which was found in a pair of siblings. The initial clinical presentation of the siblings was consistent with a BRIC-2 phenotype. In particular, the complete remission of symptoms and laboratory abnormalities during childhood is in line with the definition of BRIC-2. Recently, "BSEP-disease" has been recognized as a continuum of cholestasis of varying severity [6]. Mutations such as p.E297G or p.A570T have been associated with an intermediate phenotype between BRIC-2 and PFIC-2. Furthermore, a patient who primarily presented with a BRIC-2 phenotype developed liver cirrhosis at the age of 8 years, which was also suggested to represent a phenotypic transition between BRIC-2 and PFIC-2 [20]. In contrast to these reports, in our patient liver cirrhosis was diagnosed much later, both by transient elastography and liver histology. At that time the ability of the liver to accumulate serum bile salts into bile was severely impaired as the ratio of bile salt concentrations between bile and serum was less than 10 (µM/µM) as compared to more than 5000 (µM/µM) in healthy individuals (Herebian, Mayatepek, Kubitz and Häussinger, unpublished).

Because BSEP targeting was apparently normal, reduction of transporter activity was likely to explain the phenotypic presentation of the patients. While in other studies transport activity of BSEP was analyzed in membrane vesicles from insect cells [9, 21], we established heterologous BSEP expression in *P. pastoris* and performed transport assays using highly enriched plasma membrane vesicles. Transport of both primary bile salts (taurine conjugates of CA and CDCA) was reduced by the same order of

magnitude *in vitro* (Fig. 4D), explaining their reduced concentration in bile despite an unchanged ratio of TC to TCDC in both serum and bile *in vivo*. Although this *in vitro* transport assay may not exactly reflect the residual activity of the mutated transporter *in vivo* it strongly suggests a drastic reduction of BSEP[G374S] activity. The remaining BSEP transport activity may account for the stable BRIC-2 condition of the sister who closely followed the UDCA treatment.

A frequent finding in BSEP-disease is the reduction of BSEP expression at the canalicular membrane [8, 22], which has been estimated to occur in more than 90 % of PFIC-2 patients [23]. In our male patient, expression of BSEP[G374S] at the canalicular membrane was apparently unchanged as shown in his liver biopsy (Fig. 2A). This *in vivo* finding was supported *in vitro* by a comparable membrane localization of BSEP[WT] or BSEP[G374S] in transfected HEK293 cells (Fig. 2B). In contrast, many other disease-related mutations such as p.G238V, p.D482G, p.G982R, p.R1153C, and p.R1268Q impair membrane targeting of BSEP in transfected cells [10, 24], and it was suggested that the degree of membrane expression inversely correlates with the severity of cholestasis (PFIC-2 versus BRIC-2 versus ICP) [25]. However, the p.G374S mutation does not fit into this scheme. Systematic analysis revealed that all other missense mutations of BSEP within putative transmembrane α-helices (TMH) are linked to PFIC-2, the more severe form of BSEP-disease. Transmembrane helices may be particularly sensitive to missense mutations because substrate recognition, binding and translocation occur within these protein domains. p.G374S is the first missense mutation within a TMH that initially presents as "benign recurrent cholestasis" before ultimately leading to development of PFIC-2-like disease if not treated properly. Stable, BRIC-like liver disease in the sister but development of cirrhosis in the brother who took UDCA discontinuously indicates that p.G374S is a more severe BRIC-2 mutation which predisposes to a PFIC-2 phenotype.

In summary, p.G374S is, to the best of our knowledge, the first mutation associated with a BRIC-2 phenotype, which affects a transmembrane helix and strongly reduces transport without influencing canalicular expression of BSEP. The severity of this mutation may be due to its localisation in a transmembrane helix of BSEP.

Figure Legends

Figure 1: Genetic analysis of *ABCB11*

Gene sequencing of BSEP (*ABCB11*) revealed a homozygous nucleotide exchange at position 1120 of the coding sequence from guanine to adenine (c.1120G>A) in the two patients (sister and brother), resulting in an amino acid change from glycine to serine at position 374 of the polypeptide sequence of BSEP (p.G374S). Both parents and the younger sister were heterozygous. The genotypes correlate with the phenotypes: heterozygous, phenotypically healthy family members are represented by half-filled symbols.

Figure 2: BSEP[G374S] expression in the patient's liver and in cell culture

(A) Liver biopsy of the brother (21 years) and of a normal human liver (control) were immunostained for BSEP (green) and the phospholipid floppase MDR3 (red). Both proteins are localized at the canalicular membrane. Some canaliculi are considerably enlarged in the liver of the patient (arrows, affected brother; white bars = 20 μm). **(B)** Human BSEP was fused to yellow fluorescent protein (green). The mutation p.G374S was introduced by DREAM mutagenesis [12]. Transfection of HEK293 cells showed targeting of BSEP[WT] as well as of BSEP[G374S] to the plasma membrane. Na^+/K^+-ATPase as a plasma membrane marker is shown in red (white bars = 10 μm). **(C)**

Wildtype and mutated BSEP are equally expressed in HEK293 cells as indicated by densitometric analysis of three independent lysates of transfected HEK293 cells (for details see materials and methods). The Ponceau stain shows equal sample loading.

Figure 3: Serum bile salt profiles of both patients

(A, B) The sister and brother had significantly elevated serum bile salt concentrations with absent deoxycholate (DCA)-derivatives in the brother's serum. **(C)** The relative amounts of the primary bile salts cholic acid (CA) and chenodeoxycholic acid (CDCA) were within the normal ranges. **(B, D)** Unconjugated bile acids were (almost) absent in their sera. Individual bile salt/acid concentrations were compared to measurements of 40 control persons (dark grey bars). Mean values are represented by bars; full ranges (minima til maxima) are represented by vertical lines.

Figure 4: Strongly reduced activity of BSEP^{G374S} and position of G374

(A-C) Based on the structure of the MDR transporter Sav1866 (pdb entry: 2HYD) [26] a homology model of BSEP was constructed. The position of G374 in the BSEP model is shown in red. **(A)** Side view (arrow indicates G374 position). G374 lies in the transmembrane part of the predicted sixth helix of the first transmembrane domain. **(B)** Top view on optical section (transporter channel in green). **(C)** Side view on optical section. **(D)** Transport of [^{3}H]-labeled bile acids into yeast plasma membrane (PM) vesicles by BSEP^{G374S} is strongly reduced (mean and SEM, n=3) as compared to BSEP^{WT}. Ctrl: PM vesicles without BSEP. Right inset: Western blot analysis of equal total PM protein amounts.

Figure 5: Localization of all known BRIC- and PFIC-associated BSEP missense mutations

(A, B) All known missense mutations of BSEP were grouped according to their associated clinical phenotypes (BRIC-2 or PFIC-2) and to their localization in relation to the putative transmembrane helices (TMHs, grey boxes in **A**). All mutations within TMHs are related to PFIC-2 except p.G374S (red), which is the only BRIC-2 mutation so far localized in a TMH. **(B)** All known mutations within the predicted transmembrane segments are shown in the model of the BSEP transmembrane domain pair. The parts of the α-helices predicted to be membrane-integral are shown in dark grey. Amino acid positions of PFIC-2 mutations are coloured in green.

Reference List

1. Stieger B. Recent insights into the function and regulation of the bile salt export pump (ABCB11). Curr Opin Lipidol 2009;20:176-181.
2. Strautnieks SS, Bull LN, Knisely AS, Kocoshis SA, Dahl N, Arnell H, Sokal E, Dahan K, Childs S, Ling V, Tanner MS, Kagalwalla AF, Nemeth A, Pawlowska J, Baker A, Mieli-Vergani G, Freimer NB, Gardiner RM, Thompson RJ. A gene encoding a liver-specific ABC transporter is mutated in progressive familial intrahepatic cholestasis. Nat Genet 1998;20:233-238.
3. Bull LN, van Eijk MJ, Pawlikowska L, DeYoung JA, Juijn JA, Liao M, Klomp LW, Lomri N, Berger R, Scharschmidt BF, Knisely AS, Houwen RH, Freimer NB. A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis. Nat Genet 1998;18:219-224.
4. Ujhazy P, Ortiz D, Misra S, Li S, Moseley J, Jones H, Arias IM. Familial intrahepatic cholestasis 1: studies of localization and function. Hepatology 2001;34:768-775.
5. Carlton VE, Knisely AS, Freimer NB. Mapping of a locus for progressive familial intrahepatic cholestasis (Byler disease) to 18q21-q22, the benign recurrent intrahepatic cholestasis region. Hum Mol Genet 1995;4:1049-1053.
6. van Mil SW, Van Der Woerd WL, Van Der BG, Sturm E, Jansen PL, Bull LN, Van DB, I, Berger R, Houwen RH, Klomp LW. Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 2004;127:379-384.
7. Kubitz R, Keitel V, Scheuring S, Köhrer K, Häussinger D. Benign Recurrent Intrahepatic Cholestasis Associated With Mutations of the Bile Salt Export Pump. J Clin Gastroenterol 2006;40:171-175.
8. Keitel V, Burdelski M, Warskulat U, Kühlkamp T, Keppler D, Häussinger D, Kubitz R. Expression and localization of hepatobiliary transport proteins in Progressive Familial Intrahepatic Cholestasis. Hepatology 2005;41:1160-1172.
9. Noe J, Stieger B, Meier PJ. Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 2002;123:1659-1666.
10. Keitel V, Burdelski M, Vojnisek Z, Schmitt L, Häussinger D, Kubitz R. De novo bile salt transporter antibodies as a possible cause of recurrent graft failure after liver transplantation: a novel mechanism of cholestasis. Hepatology 2009;50:510-517.

11. Chloupkova M, Pickert A, Lee JY, Souza S, Trinh YT, Connelly SM, Dumont ME, Dean M, Urbatsch IL. Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 2007;46:7992-8003.
12. Stindt J, Ellinger P, Stross C, Keitel V, Haussinger D, Smits SH, Kubitz R, Schmitt L. Heterologous overexpression and mutagenesis of the human bile salt export pump (ABCB11) using DREAM (Directed REcombination-Assisted Mutagenesis). PLoS One 2011;6:e20562.
13. Lerner-Marmarosh N, Gimi K, Urbatsch IL, Gros P, Senior AE. Large scale purification of detergent-soluble P-glycoprotein from Pichia pastoris cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins. J Biol Chem 1999;274:34711-34718.
14. Ernst R, Kueppers P, Klein CM, Schwarzmueller T, Kuchler K, Schmitt L. A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. Proc Natl Acad Sci U S A 2008;105:5069-5074.
15. Gerloff T, Stieger B, Hagenbuch B, Madon J, Landmann L, Roth J, Hofmann AF, Meier PJ. The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver. J Biol Chem 1998;273:10046-10050.
16. Erhardt A, Lörke J, Vogt C, Poremba C, Willers R, Sagir A, Häussinger D. [Transient elastography for diagnosing liver cirrhosis]. Dtsch Med Wochenschr 2006;131:2765-2769.
17. Castera L, Vergniol J, Foucher J, Le BB, Chanteloup E, Haaser M, Darriet M, Couzigou P, De L, V. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 2005;128:343-350.
18. Kubitz R, Dröge C, Stindt J, Weissenberger K, Häussinger D. The bile salt export pump (BSEP) in health and disease. Clin Res Hepatol Gastroenterol 2012;36:536-553.
19. Tusnady GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol 1998;283:489-506.
20. Lam CW, Cheung KM, Tsui MS, Yan MS, Lee CY, Tong SF. A patient with novel ABCB11 gene mutations with phenotypic transition between BRIC2 and PFIC2. J Hepatol 2006;44:240-242.
21. Noe J, Hagenbuch B, Meier PJ, St-Pierre MV. Characterization of the mouse bile salt export pump overexpressed in the baculovirus system. Hepatology 2001;33:1223-1231.
22. Jansen PL, Strautnieks SS, Jacquemin E, Hadchouel M, Sokal EM, Hooiveld GJ, Koning JH, Jager-Krikken A, Kuipers F, Stellaard F, Bijleveld CM, Gouw A, Van Goor H, Thompson RJ, Muller M. Hepatocanalicular bile salt export pump deficiency in patients with progressive familial intrahepatic cholestasis. Gastroenterology 1999;117:1370-1379.
23. Strautnieks SS, Byrne JA, Pawlikowska L, Cebecauerova D, Rayner A, Dutton L, Meier Y, Antoniou A, Stieger B, Arnell H, Ozcay F, Al-Hussaini HF, Bassas AF, Verkade HJ, Fischler B, Nemeth A, Kotalova R, Shneider BL, Cielecka-Kuszyk J, McClean P, Whitington PF, Sokal E, Jirsa M, Wali SH, Jankowska I, Pawlowska J, Mieli-Vergani G, Knisely AS, Bull LN, Thompson RJ. Severe bile salt export pump deficiency: 82 different ABCB11 mutations in 109 families. Gastroenterology 2008;134:1203-1214.
24. Wang L, Dong H, Soroka CJ, Wei N, Boyer JL, Hochstrasser M. Degradation of the bile salt export pump at endoplasmic reticulum in progressive familial intrahepatic cholestasis type II. Hepatology 2008;48:1558-1569.
25. Lam P, Pearson CL, Soroka CJ, Xu S, Mennone A, Boyer JL. Levels of plasma membrane expression in progressive and benign mutations of the bile salt export pump (Bsep/Abcb11) correlate with severity of cholestatic diseases. Am J Physiol Cell Physiol 2007;293:C1709-C1716.
26. Dawson RJ, Locher KP. Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. FEBS Lett 2007;581:935-938.

Figures

Figure 1: Genetic analysis of ABCB11

Gene sequencing of BSEP (ABCB11) revealed a homozygous nucleotide exchange at position 1120 of the coding sequence from guanine to adenine (c.1120G>A) in the two patients (sister and brother), resulting in an amino acid change from glycine to serine at position 374 of the polypeptide sequence of BSEP (p.G374S). Both parents and the younger sister were heterozygous. The genotypes correlate with the phenotypes: heterozygous, phenotypically healthy family members are represented by half-filled symbols.

Figure 2: BSEPG374S expression in the patient's liver and in cell culture

(A) Liver biopsy of the brother (21 years) and of a normal human liver (control) were immunostained for BSEP (green) and the phospholipid floppase MDR3 (red). Both proteins are localized at the canalicular membrane. Some canaliculi are considerably enlarged in the liver of the patient (arrows, affected brother; white bars = 20 μm). **(B)** Human BSEP was fused to yellow fluorescent protein (green). The mutation p.G374S was introduced by DREAM mutagenesis 12. Transfection of HEK293 cells showed targeting of BSEPWT as well as of BSEPG374S to the plasma membrane. Na+/K+-ATPase as a plasma membrane marker is shown in red (white bars = 10 μm). **(C)** Wildtype and mutated BSEP are equally expressed in HEK293 cells as indicated by densitometric analysis of three independent lysates of transfected HEK293 cells (for details see materials and methods). The Ponceau stain shows equal sample loading.

178

Figure 3: Serum bile salt profiles of both patients
(A, B) The sister and brother had significantly elevated serum bile salt concentrations with absent deoxycholate (DCA)-derivatives in the brother's serum. **(C)** The relative amounts of the primary bile salts cholic acid (CA) and chenodeoxycholic acid (CDCA) were within the normal ranges. **(B, D)** Unconjugated bile acids were (almost) absent in their sera. Individual bile salt/acid concentrations were compared to measurements of 40 control persons (dark grey bars). Mean values are represented by bars; full ranges (minima til maxima) are represented by vertical lines.

Figure 4: Strongly reduced activity of BSEPG374S and position of G374
(A-C) Based on the structure of the MDR transporter Sav1866 (pdb entry: 2HYD) 26 a homology model of BSEP was constructed. The position of G374 in the BSEP model is shown in red. **(A)** Side view (arrow indicates G374 position). G374 lies in the transmembrane part of the predicted sixth helix of the first transmembrane domain. **(B)** Top view on optical section (transporter channel in green). **(C)** Side view on optical section. **(D)** Transport of 3[H]-labeled bile acids into yeast plasma membrane (PM) vesicles by BSEPG374S is strongly reduced (mean and SEM, n=3) as compared to BSEPWT. Ctrl: PM vesicles without BSEP. Right inset: Western blot analysis of equal total PM protein amounts.

A

TMH/loops	BRIC-2 (n=22)	PFIC-2 (n=112)
Inside		M1V, G19R, L50S
59-83		M62K, C68Y
Outside	E135K, E137K	C107R, I112T, W114R
145-168		Y157C, A167T, A167V
Inside	E186G, L198P	I182K, M183T, M183V, G188W
217-236		M217R, R223C, S226L
Outside		G238V
241-264		T242I, A257G
Inside	E297G	V284L, E297G, R303G, R303K, Q312H, R313S
319-341		G327E, W330R, C336S
Outside		W342G
362-385	G374S	A382G
Inside	A390P, R432T, I498T, A570T, T586I, G648V, T655I	R387H, A390P, G410D, L413W, I420T, D440E, G455E, K461E, T463I, Q466K, R470Q, Y472C, V481E, D482G, R487H, R487P, N490D, I498T, G499E, I512T, N515T, R517H, F540L, I541L, I541T, F548Y, D549V, G556R, G562D, A570T, L581F, A588V, S593R, I627T, E636G, R698C, S699P, E709K
754-777		G758R, G766R
Outside		
798-821		Y818F
Inside		R832C, R832H, T859R, A865V, Q869P
876-895		G877R
Outside		
900-917		S901R
Inside	T923P, A926P, R948C	R948C, N979D
980-999		G982R
Outside	G1004D	G1004D
1020-1042		T1029K, G1032R
Inside	R1050C, G1116R, R1128H, L1197G, R1231Q	A1044P, L1055P, C1083Y, A1110E, S1114R, G1116E, G1116F, G1116R, S1120N, R1128C, S1144R, R1153C, R1153H, S1154P, N1173D, T1210P, N1211D, V1212F, R1231Q, R1231W, L1242I, D1243G, R1268Q, A1283V, G1292V, G1298R

B

Figure 5: Localization of all known BRIC- and PFIC-associated BSEP missense mutations
(A, B) All known missense mutations of BSEP were grouped according to their associated clinical phenotypes (BRIC-2 or PFIC-2) and to their localization in relation to the putative transmembrane helices (TMHs). All mutations within TMHs are related to PFIC-2 except G374S (red), which is the only BRIC-2 mutation so far localized in a TMH. **(B)** All known mutations within the predicted transmembrane segments are shown in the model of the BSEP transmembrane domain pair. The parts of the α-helices predicted to be membrane-integral are shown in dark grey. Amino acid positions of PFIC-2 mutations are coloured in green.

3 General Discussion

This doctoral thesis deals with the structural and functional investigation of two different proteins with a common ground. They contain a structural motif known as the P-loop. The P-loop, a highly conserved structural feature found in many proteins across all kingdoms of life, is responsible for the binding of nucleotides like ATP or GTP. The first two chapters are concerned with the CRISPR-associated protein Csn2 from *Streptococcus agalactiae* involved in mediating adaptive immunity within the recently discovered CRISPR/Cas-system in prokaryotes. The three-dimensional structure of Csn2 was solved by using X-ray crystallography to derive a function from its structure. Furthermore, the binding properties of Csn2 to nucleic acids were investigated. Those combined studies were used to define a role of Csn2 in spacer integration in the CRISPR/Cas-system.

The last five chapters deal with the human, P-loop containing ABC transporter bile salt export pump (BSEP). BSEP is the main bile salt transporter in vertebrates and the bottleneck in the formation of bile. The aim of this thesis was to find a suitable overexpression system for BSEP for investigations of its isolated form and furthermore to establish a system for the characterization of clinically relevant BSEP mutants leading to cholestatic diseases. For both aims, the *BSEP* cDNA, which appeared intrinsically unstable in prokaryotic cloning hosts, was first cloned for a yeast-based heterologous overexpression system using a new cloning and mutagenesis procedure. For the first time, BSEP could be expressed in *P. pastoris* and subsequently purified and characterized to some extent in detergent solution. To characterize clinically relevant mutations, which were found in cholestatic patients, a vesicular transport assay based on BSEP expressing yeast cells was established. With this system a first mutant could be investigated and characterized.

3.1 CRISPR/Cas mediated immunity

The CRISPR/Cas-system has been discovered to confer immunity against foreign genetic elements like plasmids or phages in prokaryotes [23]. Roughly 40 % of all sequenced bacteria and almost 90 % of all archaea are equipped with this inheritable, adaptive immune system [167]. It consists of a CRISPR array, where pieces of foreign DNA get integrated by Cas proteins, which furthermore provide immunity [25,26]. It acts in three stages, namely I) cleavage of invading foreign nucleic acids and integration into the CRISPR array, II) expression of the CRISPR array and maturation of crRNAs and III) targeting foreign nucleic acids with the crRNA as probe and destruction by cleavage (see Figure 3 in the introduction).
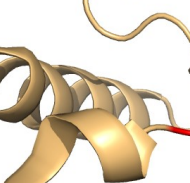
3.1.1 Spacer integration into CRISPR/Cas arrays

The most important step in the CRISPR/Cas mechanism is the acquisition of novel protospacers, nucleic acids from foreign origin, which are later used as probe to detect and destroy nucleic acids of this particular origin. Little is known about this step and many aspects still remain unclear. CRISPR/Cas-systems contain a diverse set of Cas proteins and they have been classified according to their composition [29,30]. A diagnostic feature of all CRISPR/Cas-systems is the presence of two Cas proteins known as Cas1 and Cas2. These two proteins are present in all three CRISPR/Cas types and their subtypes (see Figure 2 in the introduction). Because of this finding it is assumed that both proteins are likely to be involved in the spacer integration step. Studies have proven that the overexpression of Cas1 and Cas2 in *E. coli* mediated the incorporation of new spacers from the expression plasmid when both proteins were present whereas the overexpression of only one of the two did not lead to spacer integration [35]. Furthermore, mutations within the Cas1 protein abolished spacer integration [35]. This process is does not depend on other Cas proteins like the Cascade complex or other nucleases [35]. The exact mechanism of Cas1 or Cas2 is still unknown. Up to now, crystal structures and biochemical characterizations of Cas1 from *Pseudomonas aeruginosa*, *E. coli* and *Sulfolobus solfataricus* are available [32,33,160]. Cas1 from *S. solfataricus* binds to dsDNA, ssDNA and hybrids of RNA, and DNA with high affinity, but without any sequence specificity [160]. The Cas1 protein from *P. aeruginosa* contains a stirrup-like structure

and cleaves dsDNA in ~80 nt long pieces. Those pieces are much longer than the spacer sequence of 33 nt integrated into the CRISPR array, and no sequence specificity for cleavage could be observed, tempting to speculate that other proteins are necessary for further processing [33]. Interestingly, the Cas1 protein from *E. coli* is able to cleave ssRNA, dsDNA as well as ssDNA and also to resolve Holliday junctions, DNA structures which occur during homologous recombination, e.g. in DNA repair [32]. Furthermore, Cas1 was found to co-localize with DNA repair enzymes like RecB, RecC or RuvB [32]. This indicates that Cas1 probably plays a role in other mechanisms like DNA repair and DNA repair proteins are also involved in spacer integration. Cas2, which is always found together with Cas1 in the *cas* operons, has also been studied extensively in a structural and biochemical context [34,161,162]. Cas2 from different organisms showed diverse properties. This includes cleavage of ssRNA and ssDNA of Cas2 from *S. solfataricus* [34], no *in vitro* activity of Cas2 from *Desulfovibrio vulgaris* [161] and a magnesium-dependent dsDNA cleavage of Cas2 from *Bacillus halodurans* [162]. However, Cas2 has not yet been assigned a specific role until now. The missing sequence specificity for Cas1 and Cas2 is in line with the selection of protospacers. Protospacers contain random sequences of the foreign genetic element from the sense or antisense strands [23]; the only requirement is the PAM motif. Spacers are selected on the basis of a short sequence adjacent to the protospacer and therefore is called protospacer-adjacent motif (PAM) [31]. It could be shown in *Streptococcus thermophilus* that all new spacers selected were based on this PAM sequence and integrated proximal to the leader sequence [23]. The fact that this motif does not enhance the affinity of Cas1 or Cas2 proteins implies that additional proteins or recognition patterns are responsible for cleavage or processing of the protospacer [33]. Besides the ubiquitous occurrence of Cas1 and Cas2 a role for subtype specific Cas proteins has been assigned to spacer integration [23,163]. Fusion proteins of Cas1 and Cas4 as well as Cas3 and Cas2 are found and are likely to be involved in spacer integration [164].

3.1.2 The P-loop protein Csn2 and the link to the spacer integration mechanism

One of the first Cas proteins thought to be involved in spacer integration was Csn2 from the type IIA CRISPR/Cas-system. It was shown that a knock-out of the *csn2* gene still conferred immunity with the existing spacers, but that no acquisition of new

protospacer could be detected in *S. thermophilus* [23,36]. Therefore, Csn2 is a hallmark protein for spacer integration and unraveling its function might explain spacer integration and this function could be transferred to the other CRISPR/Cas-systems. In this thesis, the X-ray structure of Csn2 from *S. agalactiae* was solved at a resolution of 2.0 Å. It possesses a ring-like, toroidal structure built up of four protomers with a central hole in the middle with a diameter of roughly 30 Å. Furthermore, calcium ions were found in the electron density. The structure of Csn2 was aligned against structures deposited in the protein databank (PDB) and showed that Csn2 exhibits an ABC-like head domain reminiscent of the NBDs found in ABC transporters or T7 bacteriophage primase/helicase (Figure 14A). This domain belongs to group II of P-loop containing proteins and is referred to as "ASCE" type.

A

B

Figure 14: The Csn2 ABC domain harboring the P-loop.
A) The ABC domain (head domain) of Csn2 with the P-loop depicted in blue and the Walker B in red (picture taken from chapter 1). **B)** Section of an amino acid sequence alignment between Csn2 and P-loop containing proteins with a similar structural fold as identified by a structural alignment using the DALI server [165]. The P-loop (Walker A) is depicted in blue, the Walker B motif in red and the consensus sequence is shown under the respective box.

A helicase activity of Csn2 would fit into a general model of spacer integration because helicases usually catalyze the separation of DNA double-strands using the energy of ATP hydrolysis in processes like recombination or transcription [166]. As mentioned above, Cas1 co-localizes with helicases like RecB and D in *E. coli* [32], so Csn2 could adopt such a function in *Streptococcus spec*. Nonetheless, on the sequence level the P-loop is completely degenerated and the Walker B does not contain an aspartate or glutamate, either (Figure 14B). Csn2 was assayed for nucleotide hydrolysis and binding, but no activity could be detected (Figure 15).
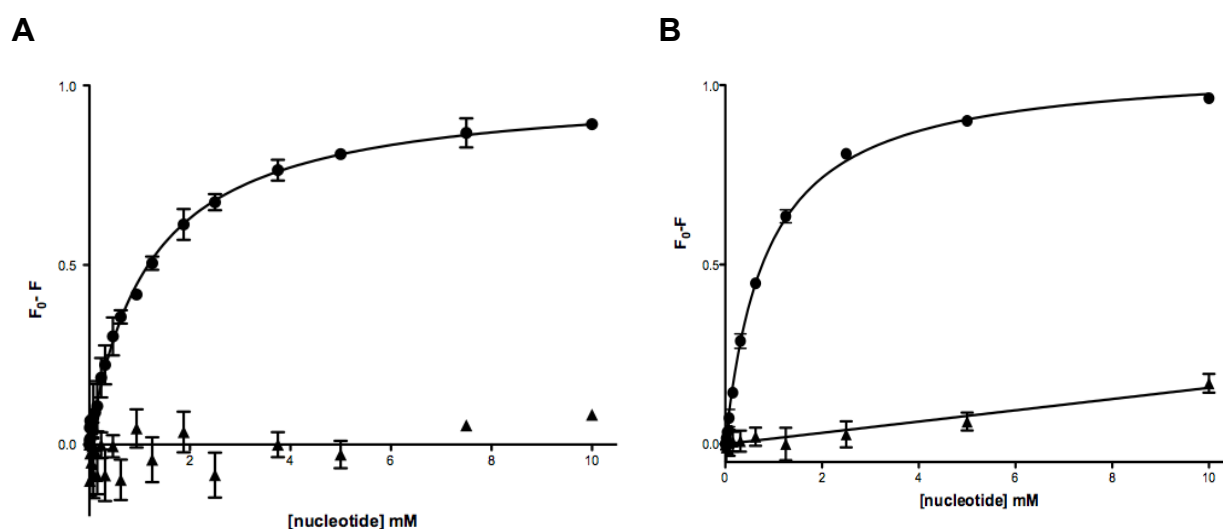
A   **B**

Figure 15: Binding curves for GTP and ATP to **A)** Csn2 Y29W and **B)** N-Acetyl-L-tryptophanamide (NATA), a tryptophane blocked at its ends, determined using intrinsic tryptophane fluorescence. (N=3, mean ± SD).black circle = GTP, black triangle = ATP

For binding experiments a Csn2 mutant, Csn2 Y29W, was generated to use intrinsic tryptophane fluorescence to monitor binding of nucleotides. The mutated tyrosine 29 (Y29) is part of the P-loop (see alignment Figure 14B), which usually participates in nucleotide binding [8]. As a control experiment binding of nucleotides (GTP and ATP) to a tryptophane analogue, N-Acetyl-L-tryptophanamide (NATA) was also determined. In both cases, the dissociation constant (K_d) for GTP was similar (1.16 ± 0.062 mM vs. 0.85 ± 0.029 mM, respectively) which leads to the conclusion that the measured effect was due to effects like inner filter effects of high concentrations of nucleotide rather than binding of nucleotides to Csn2. For ATP, no K_d could be determined. This finding was also observed for Csn2 from *Enterococcus faecalis*, whose structure was also determined and possesses a similar shape and oligomeric state [38]. Thus the possibility of helicase action could be excluded. Therefore, the

Csn2 protein is a nice example of structural conservation of the P-loop, but without any enzymatic function. Strikingly, Csn2 exhibits a conserved DxD motif as well as conserved arginine residues in the head domain. Those residues are also conserved in RNA polymerases and primases [167]. One option for spacer integration could be the detection of the PAM motif by Csn2 in foreign nucleic acids and the subsequent generation of a primer for transcription of the protospacer. The protospacer thus has to be converted back from RNA to DNA and then could be integrated by the help of Cas1 or Cas2. Even Cas1 fusion proteins with reverse transcriptase domains that underline this idea exist [168]. But *in vitro* experiments could not detect a polymerase function of Csn2. The only function which could be assigned to Csn2 was binding to dsDNA in a calcium-dependent manner. Until now, this behavior has already been shown for Csn2 proteins from other organisms, like *E. faecalis* and *S. pyogenes* [38,69]. The Csn2-like protein from *S. thermophilus* binds linear dsDNA without calcium. Mutational analysis revealed that the dsDNA was not bound by the head region close to the P-loop or the conserved DxD motif, but probably around the hole in the middle. The cavity is lined by a set of conserved lysine residues and molecular dynamics studies revealed two possible binding modes. In one of the binding modes Csn2 sits on the DNA, whereas in the other DNA gets threaded into the hole. That is why DNA can only lace at its end through Csn2, because Csn2 is a very stable tetramer under certain conditions and may not assemble around the DNA like other DNA binding proteins (e.g clamp loader PCNA [169]). Thus, in chapter 1, the Csn2 structure and its binding to dsDNA could be revealed, but no distinct function could be derived from its ABC domain or overall structure, except binding to dsDNA in a sequence unspecific manner. In Chapter 2, the Csn2 DNA binding properties were investigated in more detail using various biochemical techniques. It could be shown that Csn2 just binds to linear dsDNA. Free DNA ends are a requirement for the binding of Csn2 because neither supercoiled or relaxed circular DNA nor DNA with blocked ends could be bound by Csn2. Atomic force microscopy also showed a cooperative binding found in other DNA end-binding proteins (e.g. RdgC [170]), which lead to a "pearls-on-a-string" structure. For binding to dsDNA, Csn2 has to bind at the very end and the DNA has to be threaded through the center.

3.1.3 Non-homologous end-joining as mechanism for spacer integration

As Csn2 seems to have no enzymatic activity (no helicase, nuclease or polymerase activity), Csn2 might "just" be a DNA end-binding protein. The behavior described in Chapters 1 and 2 is reminiscent of the Ku protein, which is involved in non-homologous end-joining (NHEJ), a DNA double-strand break (DSB) repair mechanism [171,172]. Ku is a conserved protein found in eukaryotes like humans or yeast as well as in prokaryotes. Depending on the organism, it forms hetero- or homodimeric complexes and like Csn2 binds DNA DSBs with high affinity (K_d ~30-60 nM), but most importantly in a sequence-independent manner. Thereby, the free DNA ends are protected against nucleolytic degradation [171,172]. Furthermore, it has the ability to be loaded onto the DNA in a higher quantity which then gives the above-mentioned "pearls-on-a-string" structure [173]. Ku serves as a platform for the recruitment of other proteins like LigD in bacteria, a protein consisting of a polymerase, a nuclease and a ligase domain [174]. Therefore, eukaryotic Ku proteins exhibit additional C- and N-terminal extensions (van Willebrand factor A and SAP domains) which the bacterial counterparts do not have [175]. The Csn2-like protein from *S. thermophilus* also contains a C-terminal extension, which might also participate in protein-protein interactions and recruit proteins involved in DNA repair [70]. NHEJ in bacteria just depends on Ku and LigD. DSBs are either repaired using microhomology between complementary 3´ overhangs of DNA breaks or are directly joined [171,172]. In eukaryotes, both free DNA ends are occupied by the Ku heterodimer, tethered together by DNA-dependent protein kinases (DNA-PKs), and repair enzymes are recruited [171,172]. This mechanism could likely be involved in spacer integration in the CRISPR/Cas-system. Cas1 or Cas2 or another nuclease could introduce a DSB into the first repeat adjacent to the leader sequence, while Cas1 or Cas2 could also be involved in protospacer selection depending on the PAM and its nucleolytic digestion from the foreign nucleic acid. Csn2 then binds to the free DNA ends and recruits the protospacer, which is bound by a Cas protein as well as a LigD-like host protein to join the protospacer between the opened repeat. The polymerase activity of the LigD-like protein fills gaps and the spacer is inserted and the repeat sequence duplicated (see model chapter 2). Thus, this work sheds a first light onto spacer integration in the CRISPR/Cas-system.

3.2 The bile forming ABC transporter BSEP

The formation of bile is a crucial aspect in all vertebrates. Bile promotes the digestion of fat as well as the uptake of lipophilic substances like vitamins. Furthermore, bile circulation presents a pathway to get rid of endo- or exogenous xenobiotics [130].

For bile formation, three ABC transporters are essential: BSEP, which transports bile salts, MDR3, a transporter responsible for flopping the phospholipid phosphatidylcholine from the inner to the outer leaflet of canalicular membranes, and ABCG5/G8, a heterodimeric ABC transporter involved in cholesterol transport [128]. Together, those three components constitute bile.

Because BSEP is the limiting factor in this formation and because mutations within its gene can lead to severe cholestatic diseases, an understanding of its biochemical and mechanistical properties and the analysis of structure-function relationship of mutant BSEP variants would greatly contribute to improve potential drug design and the treatment of patients. Thus, the first objective was to find a suitable overexpression system to obtain BSEP in high yields for purification and its subsequent analysis in its isolated state.

3.2.1 Heterologous overexpression systems for eukaryotic membrane proteins

In order to study a protein in its isolated form, expression and purification are necessary steps towards an understanding of its characteristics. To date, the overexpression system most widely used for homo- or heterologous recombinant proteins is based on the Gram-negative bacterium *E. coli* [176]. It provides a well-known molecular biology and promotor-polymerase systems are available to achieve high yields of homo- or heterologously expressed proteins. Also, the Cas protein Csn2 could be expressed very easily to high yields in *E. coli* (see chapter 1). Although this might be true for a large number of proteins, the overexpression, especially of eukaryotic membrane proteins in *E. coli*, is a major obstacle [177]. Prokaryotic hosts might have problems with the transcription or translation of eukaryotic genes due to codon bias and their lack of the eukaryotic protein processing machinery for trafficking, glycosylation or membrane insertion [178]. Another very important factor is the non-native lipid environment, because it is known that specific protein-lipid interactions are crucial for folding, stability or activity of

membrane proteins [179]. The human ABC transporters P-gp and ABCG2 (BCRP) have also been expressed in *E. coli*, but it was shown that the topology of P-gp expressed in cell-culture systems was different [180,181,182]. ABCG2 showed transport activity in living bacteria, but was inactive in inverted membrane vesicles and could only be solubilized with SDS [182]. A codon-optimized gene of BSEP for *E. coli* was also tested during this thesis, but no expression could be detected. Thus, bacterial systems like *E. coli* might not be the first choice for the expression of eukaryotic ABC transporters [183]. Nonetheless, a number of eukaryotic integral membrane proteins (IMPs) have been produced in a functional form in *E. coli* [184]. The Gram-positive bacterium *Lactococcus lactis* was successfully used to overexpress eukaryotic membrane proteins, but only few reports were published [185]. For this reason, prokaryotic homologues are often expressed and purified, because they are easier to handle and results can be transferred to eukaryotic membrane proteins [178].

BSEP is well-known to be expressed in eukaryotic cell culture systems. Compared to bacterial hosts the advantage is obvious. Cell culture systems harbor all the native systems necessary for translation, targeting, folding and insertion into the membrane and post-translational modifications yielding functional protein. Many cell types have been used to express BSEP. Examples are HEK293, HepG2, LLC PK1, MDCK or *Sf*9 cells [186,187,188,189]. These cells resemble the natural canalicular membrane as much as possible. They contain the mammalian lipid environment and glycosylation pattern and BSEP is localized in the plasma membrane of those cells. Cell culture systems have been used for a wide range of human ABC transporters, primarily to study questions on cellular processes like trafficking, localization or transport. Furthermore, active purification of human ABC transporters from HEK293 or *Sf*9 cells has also been reported for transporters like ABCG2, P-gp or MRP2 [190,191,192]. Nonetheless, for structural and mechanistical *in vitro* approaches larger amounts of purified protein are needed which makes the use of cell culture systems extremely cost intensive and therefore not feasible. In general yeast-based systems have been found to be suitbale for the heterologous overexpression of eukaryotic membrane proteins [193]. In general, yeast systems are inexpensive, only need simple culture requirements, have well established genetics and contain the eukaryotic protein processing machinery.

A prominent system is based on the yeast *S. cerevisiae*. It cannot only be used as a cloning organism, but also as a host for the heterologous overexpression of eukaryotic membrane proteins, like rabbit Ca^{2+}-ATPase or human AQP0 aquaporin [194,195]. Furthermore, human ABC transporters have already been overexpressed in this host. Human P-gp and MRP1 are two examples for functional expression [196,197]. In Chapter 4, BSEP was also expressed in *S. cerevisae*, but could only be detected on immunoblot levels. Even a variation of the tag position had no influence on the expression level. There can be several reasons for low expression, like codon bias or a difference in glycosylation, which might be important for trafficking or stability. *S. cerevisiae* is known to hyperglycosylate proteins (~40-150 mannose residues) [178]. This does not seem to be the case for BSEP, as it migrates at ~130 kDa in the SDS-PAGE in contrast to fully glycosylated BSEP with a size of ~160-180 kDa expressed in cell culture [151]. Proteolytic degradation of recombinant proteins, also often observed in yeast could not be detected in immunoblotting. Because of this low abundance of BSEP, the expression system was changed to the methylotrophic yeast *P. pastoris*. This yeast can be grown to very high cell densities during fermentation yielding kilogram quantities of wet cell weight and has been used as host for the expression of different eukaryotic membrane proteins which ended up in structure determination [178]. One study reported the heterologous overexpression of 25 of the 48 human ABC transporters [198]. Remarkably, the human liver ABC transporter MRP2 could not be expressed, but three other liver ABC transporters (P-gp, ABCG5/G8 and MRP3) were purified to homogeneity and yielded roughly 1-6 mg of protein [105,198,199,200,201]. Furthermore, the crystal structure of mouse P-gp was solved based on the expression in *P. pastoris*, thus making this expression system extremely suitable for human ABC transporters [105]. The bile forming ABC transporters BSEP and MDR3 were not included in this study, likely due to their instable cDNAs during cloning. Therefore, in this thesis, the *BSEP* cDNA was cloned for the *P. pastoris* expression system with the cloning procedure established in chapter 4 and a substantial overexpression, as shown in chapters 4 and 5 could be achieved. Again, no proteolytic degradation in *P. pastoris* could be observed. Hence, in this thesis the expression of the important human ABC transporter could be established yielding ~1 mg out of 100 g of cells.

3.2.2 A generalized cloning and mutagenesis approach for the instable *BSEP* cDNA

A big drawback in cloning the *BSEP* cDNA is its instability during cloning in the common host *E. coli*, used in classical cloning approaches based on restriction and ligation enzymes [151,152]. *E. coli* cells transformed with the *BSEP* cDNA are mostly not viable and clones are rarely observed. Moreover, the observed clones then often contain random mutations or deletions of the sequence. Furthermore, the *BSEP* cDNA contains a cryptic prokaryotic promotor, but even silencing this sequence did not abolish the instability [151,152]. As a consequence, cloning of the *BSEP* cDNA is a tremendous effort and only limited information of BSEP from heterologous expression systems is available. There are different reasons, why DNA sequences can be instable in bacterial hosts. On the sequence level, secondary structure elements like z-DNA or direct repeats can be responsible [202]. On the other hand, the gene product can be toxic for the host, either interfering with cellular processes or overloading the translation and degradation machinery, especially observed for eukaryotic proteins, which need additional processing not available in prokaryotic hosts [203]. In particular, eukaryotic membrane proteins in particular show this effect. For example the b and c subunits of the F-ATPase or the oxoglutarate-malate carrier protein from mitochondrial membranes were hard to express in *E. coli* [203]. Another example of failure in cloning is the fungal high-affinity calcium channel CCH1 [204]. *E. coli* clones did not appear and remarkably, the deletion of the start codon ATG reversed those toxic, instable effects, although no *E. coli* promotor for protein expression was upstream of the gene sequence. This is also the case for the *BSEP* cDNA, as the attempt to clone it in *E. coli* failed. To obtain yeast expression constructs, BSEP was cloned using homologous recombination (HR) in the yeast *S. cerevisiae* itself. HR is an efficient way for general cloning and to prevent instable side effects in prokaryotes [205]. The size of inserted fragments does not matter and cloning "artefacts" like linker regions often occurring in classical cloning can be prevented. Required are homologous ends with a length of 20 to 50 bp between the expression plasmid backbone and the gene-of-interest, which can easily be introduced by PCR. To generalize this cloning procedure for BSEP for other expression systems, an origin of replication for *S. cerevisae* was introduced into the backbone of expression vectors for other systems like the *P. pastoris* overexpression

system. In that way, this cloning procedure opens the way to clone each instable gene-of-interest for any expression system, just by using *S. cerevisae* and its origin of replication.

Mutational studies on proteins are important to understand their structure-function relationships or protein-protein interactions. Usually, mutations are inserted via a site-directed mutagenesis procedure [206]. Thereby, the primers containing the mutation are completely complementary and the mutagenesis PCR is a non-exponential amplification of the template. The product of the first cycle cannot be used as template for the second cycle. This prevents errors occurring during the PCR reaction. Furthermore, all those techniques all rely on the use of *E. coli*. The final products all contain nicks in the mutated plasmid, which are ligated in *E. coli*. Because of the inherent instability of the *BSEP* cDNA obtaining mutated BSEP variants is very hard to achieve. Therefore, in this thesis a complete new mutagenesis method, just relying on yeast, was developed (see Figure 16). By changing the primer design from complete to partial overlap, the PCR reaction becomes exponential with the rationale to outnumber the template. Usually, the template, which of course does not carry the mutation, is digested by a restriction enzyme recognizing methylated DNA originated from *E. coli*. This is not the case for yeast and thus the template has to be outnumbered during PCR. The primer design furthermore, introduces homologous ends to the mutated PCR product that directly can be transformed into yeast and is recombined to the complete vector carrying the mutation. This method provides an easy and rapid way to mutate instable cDNAs with a minimum of effort. Here, it was used to introduce clinically relevant mutations into the *BSEP* cDNA for further biochemical studies (see below).
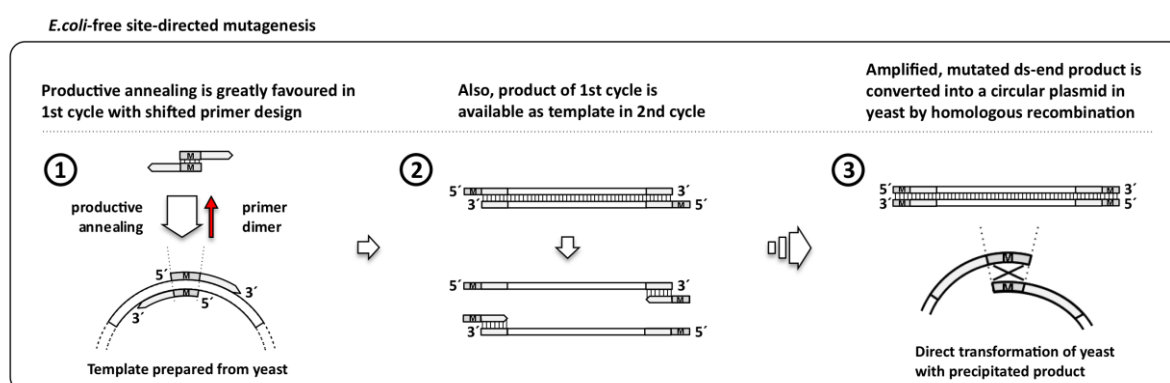

Figure 16: Scheme of the DREAM mutagenesis just relying on *S. cerevisiae*.
M = mutation within the primer.

3.2.3 Purification of human BSEP – the need for detergent screening

Studying a protein in its isolated form has the big advantage of avoiding disturbing side effects. Therefore, the subsequent purification after its successful overexpression has to be performed. Protein purification technologies have made big advances and many strategies are available now. The most used technique is the affinity purification relying on affinity tags genetically fused to the protein to be purified [207]. The most prominent one is the histidine tag with varying length. After disrupting the biological material the protein of interest can easily be purified in most cases. This might also apply to membrane proteins, but a big obstacle is the environment the protein is located in: a biological membrane. Because lipids interfere with most downstream applications, the membrane protein has to be solubilized out of the membrane first [208]. This is achieved by the use of detergents. Detergents are amphipathic molecules with a polar head group and a hydrophobic tail. They can be subdivided into four main groups based on their charge: anionic, cationic, zwitterionic and neutral [208]. A membrane protein in solution would directly aggregate due its nature. The transmembrane domains are hydrophobic and line the lipids in the membrane; thereby the entropic effect favors a clustering of membrane proteins in solution. To avoid this and to disrupt the membrane structure, detergents are used to solubilize membranes. The detergent destabilizes the membrane structure and the hydrophobic tail lines the TMD of the membrane protein. Detergents have to be used above their critical micellar concentration (cmc), so that they form micelles, where the membrane protein gets inserted in. Finding the detergent which, on the one hand extracts the membrane protein efficiently and, on the other hand keeps the membrane protein in a functional, folded state is an empirical process like the overexpression [208]. Usually detergents are chosen that have been proven to be well suited in many cases, like β-DDM, a neutral maltoside detergent [208]. For BSEP expressed in *P. pastoris*, a high-throughput detergent screening was employed to directly test over 100 different detergents at once using a dot blot technique. This technique was also used before for other eukaryotic membrane proteins [209]. In chapter 5, this technique was employed for BSEP containing two affinity tags for purification consisting of a calmodulin-binding peptide (CBP) tag and a his_{6x}-tag as well as for a BSEP-GFP fusion protein. Remarkably, only the zwitterionic, harsh detergents of the Fos-Choline series, which are lipid-like

detergents, were able to solubilize BSEP in an efficient way. This result is in contrast to other used detergents used for liver ABC transporters expressed heterologously in *P. pastoris*. ABCC3 was solubilized in β-DDM like ABCG5/G8 [198,200]. P-gp was solubilized in various detergents from *P. pastoris* membranes including β-DM, β-DDM, Lyso-PC, deoxycholic acid or Triton-X100 [105,210,211,212,213]. Despite the high degree of sequence identity between P-gp and BSEP (~50 %), it could not be solubilized with Triton-X100, which was used to solubilize mouse P-gp [105]. Also ABCG2 was solubilized in β-DDM, but could only be solubilized in Fos-Choline-16 when expressed in *High Five* insect cells [192,214]. In some cases, proteins could only be solubilized in Fos-Cholines and resulted in non-native or inactive proteins [215]. To gain insight into the solubilized state of BSEP, fluorescence-detection size-exclusion chromatography (FSEC) was employed. This technique has the advantage to monitor the monodispersity of membrane proteins in detergent solution without the need for purification [216]. Only a fluorescence tag, like GFP, has to be fused to the target protein. Thereby, only the protein of interest is monitored by fluorescence detection. As shown in chapter 5, BSEP behaved non-monodispersely in the Fos-Choline series, but showed nice symmetrical peaks in neutral, maltoside detergents like β-DDM, β-DM or Cymal5. As it is a common procedure to change the detergent after solubilization, this was done for BSEP as well. BSEP was solubilized in large amounts in Fos-Choline-16 and then the detergent was exchanged on the Ni^{2+}-IDA-column, which was used to purify BSEP making use of its his_{6x}-tag. BSEP was further purified using its CBP-tag. With this established protocol, ~1 mg of human BSEP could be obtained in different detergents, after extraction with Fos-Choline-16. This result is in good agreement with other human ABC transporters purified from *P. pastoris* [198,210] and shows its applicability for human ABC transporters, in contrast to cell culture systems, which might only produce µg quantities. The detergent exchange can be advantageous, e.g. in restoring activity, as it might be lost in the solubilizing detergent. For mouse P-gp, the detergent was exchanged after solubilization from *P. pastoris* membranes in Triton-X100 to β-DDM and then crystallized in a functional manner and the structure was solved [105]. The purity of the BSEP preparations was roughly 80 %, most likely due to the use of Fos-Choline-16. This detergent solubilizes membranes very efficiently and other membrane proteins including BSEP get solubilized and are integrated in the same detergent micelle. Different approaches like buffer changes, longer washing steps or different

salt conditions to prevent purifying those contaminants failed. Nonetheless, chapter 5 presents the first published purification procedure for BSEP and MDR3, another bile forming ABC transporter.

3.2.4 Assays to evaluate functional protein purification – BSEP in detergent solution

After the successful purification of a protein, it is inevitable to test the protein for activity, if feasible. This, of course, requires an assay in hand that is specific for the protein. In the case of ABC transporters, the first choice is the ATPase assay [217]. ABC transporters couple the transport of allocrites across membranes to ATP hydrolysis, thus the hydrolysis of ATP can be measured by the release of inorganic phosphate (P_i). ABC transporters are P-loop containing proteins, belonging to the "ASCE" group II of P-loop proteins [11,12]. As a base, they utilize a catalytic glutamate, which abstracts a proton from a catalytic water molecule [13]. The activated water molecule is used to attack the β-γ phosphodiester bond of ATP to hydrolyze it to P_i and ADP. According to the ATP switch model this resets the transporter to its ground state (see Figure 11 in introduction). Furthermore, binding of nucleotides to P-loop proteins can be determined. As in the case of Csn2 (see Figure 15), this has been done for several ABC transporters using wilde-type or tryptophane substituted proteins [218,219]. Instead of modifying the nucleotides binding pocket, fluorescent nucleotide derivatives can be used like 2'(or 3')-O-(2, 4, 6-trinitrophenyl)-adenosine-5'- monophosphate (TNP-ATP) [220,221]. Upon binding, the fluorescence increases and binding constants can be determined, although the affinity might be higher due to the TNP group. Moreover, binding can also be monitored using radioactive labeled nucleotides or nucleotides coupled to agarose beads, which might give more qualitative data. In the case of purified BSEP, no substantial ATPase activity could be measured in its basal state without any allocrite. Furthermore, addition of allocrites like taurocholate or taurochenodeoxycholate did not show an impact. It is not uncommon that purified ABC transporters display no substantial ATPase activity. This has been shown for human ABCD1 purified from *Sf*9 cells or the ABC transporter LmrA from *L. lactis* [217,222]. Likewise, the yeast ABC transporter Pdr5p did not show ATPase activity when solubilized in β-DDM and it could be shown that detergent molecules inhibited transport as well as ATPase

activity [223]. This was ascribed to a lock of the transporter in an ATPase-incompatible state. In this state, the NBDs are close together, so that hydrolysis cannot proceed. Although BSEP is able to bind to ATP agarose in detergent, this could also explain the lost ATPase activity. A crucial aspect during purification is to prevent complete delipidation. This cannot be controlled in detail because it depends on many different aspects like duration of washing during purification or the nature of the detergent used. Therefore, it has been shown that adding lipids back or including them during the purification process retained activity of some membrane proteins [179]. In the case of LmrA for instance, the reconstitution into liposomes drastically increased its specific activity compared to the detergent solubilized state as well as for Pdr5p [217]. Reasons for such a behavior might be either the lateral pressure in the membrane, which applies to the membrane protein, or the presence of annular lipids, tightly bound lipids at the membrane protein, which can have a crucial role in its activity or transport mechanism. Adding back lipids like whole-cell liver lipids to the purified BSEP also did not show an effect, either. For BSEP, it is known that its transport activity is depending on cholesterol as shown by the enrichment of *Sf*9 cell membranes expressing BSEP with cholesterol [224]. A drastical increase in its transport activity could be observed. How this effect is mediated is still unknown. Cholesterol might be bound by BSEP as annular sterol or the binding site of BSEP needs to be filled by cholesterol, to accommodate for bile salt binding according to the cholesterol fill-in theory [109]. MDR3 and BSEP share ~80 % and ~50 % identity to P-gp, a multidrug transporter transporting a wide variety of hydrophobic, structurally unrelated compounds [225]. To explain the differences in the allocrite spectrum, it was assumed that the binding site might be very similar, but binding of cholesterol within the binding cavity forms the three-dimensional binding site capable of binding individual allocrites [109]. In the yeast expression host however, ergosterol is the predominant sterol instead of cholesterol like in other mammalian cells. Both sterols differ by two additional double bonds (in the ring and in the tail) in the case of ergosterol, which might abrogate binding of ergosterol and thus binding of the allocrite and ATPase-activity. To further test, if BSEP is at least able to bind to ATP, binding assays with ATP agarose beads were performed. BSEP was able to bind to ATP in neutral, maltoside detergents like β-DDM and Cymal-5, but not in Fos-Choline-16, the detergent used for its solubilization. This experiment at least shows that the NBDs have to be properly folded so that binding can be restored. Fos-

Choline-16 might lock the transporter more drastically than the neutral detergents. Most likely, BSEP just gets inhibited by detergent molecules, as is the case for other ABC transporters. However, the assumption that a membrane component is missing which has not yet identified cannot be ruled out.

3.2.5 The vesicular transport assay based on yeast membranes to study BSEP mutants

To test whether BSEP is functionally expressed in *P. pastoris* in general, membranes were prepared and a vesicular uptake assay was performed. A protocol for highly-enriched plasma membranes, originally established for the yeast *S. cerevisiae*, was adopted [226]. The vesicular transport assay is a well-established functional test system for ABC transporters [227]. This is the first type of assay which could detect the direct transport of substrates across a membrane [228]. The prerequisite are so called inside-out vesicles, which form under adequate conditions. In this configuration, the NBDs of the ABC transporter are localized to the surrounding medium and the direction of transport is into the lumen of the vesicle. After the transport has taken place, the vesicles are separated from the rest of the liquid solution, in most cases by filtration, and the amount of transported substrate can be determined. This depends on the nature of the test compound. Most common are radioactive labeled substrates, so that the measurement of the radioactivity is the direct output of activity. Also, fluorescent or unlabeled substrates are used. These then can be detected by fluorescence spectroscopy or mass spectrometry, respectively. For BSEP the vesicular uptake assay has been used extensively to characterize its kinetic values as well as the inhibition or modulation of its activity [148]. The vesicular transport assay is mostly based on the BSEP expression in insect cells like *Sf*9 and to a lesser extent in HEK293 cells. In many studies, the substrate spectra as well as potent inhibitors could be revealed. The Michaelis-Menten constants (K_M) for various substrates have been measured in different expression systems and multiple times (Table 1) [148].

Table 1: K_M-values for bile salts determined with different systems for BSEP.

Substrate	K_M / μM	Source
Taurocholate	8, 20, 15	*Sf*9
	4	HighFive
Taurochenodeoxycholate	4, 5, 13	*Sf*9
	7	HEK293

Therefore, to test for functionality of BSEP, the vesicular uptake assay for BSEP was adopted to membranes prepared from *P. pastoris*. The most important value to be determined was the Michaelis-Menten constant, as this is an indirect measure for the affinity of the substrate to BSEP. In chapter 6 two tritium labeled substrates, taurocholate and taurochenodeoxycholate were tested for their transport. BSEP expressed in *P. pastoris* was able to transport both substrates and the K_M-values were 22.7 ± 6.2 µM and 12.9 ± 1.8 µM, respectively. These values are in very good agreement with the observed constants from insect cell culture or mammalian cell culture [148]. This indicates that human BSEP can be functionally expressed in *P. pastoris* and that it requires at least some lipids to be functional compared to detergent-solubilized BSEP. This investigation also shows that ergosterol might replace mammalian cholesterol, which seems to be an important factor for its activity. The canalicular membrane is highly enriched in cholesterol and other ABC transporters like ABCG2 or P-gp have been shown to reside in detergent-resistant microdomains (DRMs) [229,230]. This is also the case for BSEP [231]. Those DRMs are enriched in cholesterol and sphingomyelin and certain proteins have been shown to co-localize with them [232]. DRMs in canalicular membranes might protect the membrane against solubilization by bile salts and thus protect the transporters. There might also be an inherent resistance of BSEP against solubilization, which might be determined by its amino acid sequence or in concert with bound lipids.

3.2.6 The impact of BSEP mutations on its structure and function

BSEP is the main bile salt transporter in vertebrates and responsible for the bile salt dependent bile flow [233]. Thus, dysfunction disrupts the complete bile flow, either by slowing it down or in the worst case by complete abrogation [234]. The reduction of bile flow is called cholestasis. There are two common mechanisms leading to

cholestasis caused by BSEP. BSEP is retrieved from the plasma membrane under certain conditions, either, under hyperosmolaric conditions or by taurolithocholic acid [235,236]. This reduces the number of active transporters in the membrane and bile salt transport is reduced. The other cause for BSEP-dependent cholestatic diseases are mutations within the *BSEP* gene [237]. These can be nonsense mutations leading to a truncated protein by the introduction of a stop codon, missense mutations leading to an amino acid change, and insertion or deletion mutations where nucleotides are introduced or deleted in the gene resulting in a changed protein sequence. Mutations have also been shown to affect the correct splicing of *BSEP* [238]. In those cases, the change in the nucleotide sequence changes the so-called exonic splicing enhancer site (ESE), which is needed to guide the splicing machinery to the corresponding splice site [239].

In principle, three different cholestatic diseases in regard to BSEP mutations can be distinguished. I) Progressive familial intrahepatic cholestasis type 2 (PFIC2) [237], II) benign recurrent intrahepatic cholestasis type 2 (BRIC2) [156] and III) intrahepatic cholestasis of pregnancy (ICP) [159]. PFIC2 symptoms include jaundice, itching and growth failure and is usually diagnosed within the first half year after birth. In general, over 90 % of those mutations lead to the absence of functional BSEP protein in the membrane [154,240]. This leads to elevated levels of bile salts within the cell and can induce apoptotic effects or the induction of carcinogenesis due to reactive oxygen or nitrogen species. The character of the mutation may explain the different causes of the membrane absence. One of them being mutations leading to instable protein after translation. The global BSEP folding is affected and BSEP is retained within the ER and degraded. This could be shown e.g. for the BSEP G238V, G982R, R1153C and R1268Q mutations [241]. Furthermore, mutations can have an impact on trafficking to the membrane itself. This was observed for the E297G and D482G mutations [187]. In addition, mutating of glycosylation sites also lead to trafficking defects and intracellular degradation [242]. Also, mutated BSEP protein, which reaches the plasma membrane to a certain portion, may have a reduced half-life as shown for the D482G and E297G mutations. The half-life could be prolonged by the use of chemical chaperones like 4-phenylbutyrate (4-PBA) [242]. Trafficking mutations may not lead to inactive protein *per se* as folding mutations may do. So it is always necessary to evaluate the transport function of such mutations. On the other hand there are mutations which do not influence folding or trafficking and have a

normal plasma membrane half-life. In these mutations, the binding site of the substrate or the ATP hydrolysis is affected. In chapter 7, a summary of all published mutations leading to either PFIC2 or BRIC2 is given. From this it can be derived that both, PFIC2 and BRIC2 mutations, are both localized in loops located intra- or extracytosolic or within the NBDs of BSEP. This might influence either the stability, trafficking or ATP hydrolysis mechanism of BSEP. 22 known PFIC2 mutations are localized within the TMDs of BSEP. Those mutations can have an influence on the binding site for bile salts in BSEP, as it is presumed that allocrite binding takes place within the TMDs. As only a small portion of all known BSEP mutations are characterized to some extent, either with immunohistochemistry of liver biopsies or rarely with transport assays, the influence on allocrite binding is hard to judge. For some mutations in the TMDs, at least the localization in hepatocytes was determined. It could be shown that e.g. the M62K, Y157C, A167T, G327E or the G982R BSEP mutants were not detected at the canalicular membrane in liver biopsies [243,244,245]. Also, the level was drastically reduced in some cases as for the Y818F or C68Y mutants [238,246]. Those findings likely point to an instable BSEP protein due to misfolding and rapid degradation. But even these findings need to be seen with caution, because patients often are compound heterozygote and contain two different mutant alleles of BSEP, so that the influence of an individual mutation to the clinical finding is hard to establish. In chapter 7, the BSEP G374S mutation was investigated. This mutation was found in a patient suffering from cholestasis. Highly interesting, the mutation seemed to display the first BRIC2 mutation within a TMD. BRIC2 is a milder form of cholestasis, because patients only suffer intermittently from cholestatic effects [156]. This is dependent on the residual transport activity of the mutated BSEP and the patients' way of life. Liver biopsy, as well as expression in HEK293 cells of BSEP G374S showed correct targeting to the plasma membrane, excluding trafficking or degradation effects. It was shown that the transport function of this mutant was strongly reduced compared to wild-type BSEP protein, but still showed minimal transport, which explains the BRIC2 phenotype of the patient. This was tested in the transport assay, established in *P. pastoris* as described above. Because of this finding, it can be assumed that this mutation affects the allocrite binding site. Either because this amino acid directly participates in binding or blocks the way to the binding site. A further mutant investigated, which is not part of this thesis, is the T919Δ mutant. Although one amino acid is missing, the global fold may

not be affected, because it is correctly targeted to the plasma membrane and, very interestingly, shows transport of taurocholate comparable to the wild-type protein (Figure 17). Although the overall bile salt concentration in bile of this patient is only 2.5 % within of healthy individuals (unpublished results), bile salt transport seems to be comparable to wild-type BSEP in *in vitro* systems. Other mechanisms than an impaired bile salt transport have to be considered for this cholestatic disease. Such a characteristic was also observed for the PFIC2-associated mutant BSEP C336S, which is also localized in the TMD. This mutant also displayed transport rates comparable to those of the wild-type BSEP protein when expressed in insect cells [247]. Nonetheless, it could not be excluded that further BSEP mutations in the gene caused this cholestatic symptoms due to imperfect sequencing.
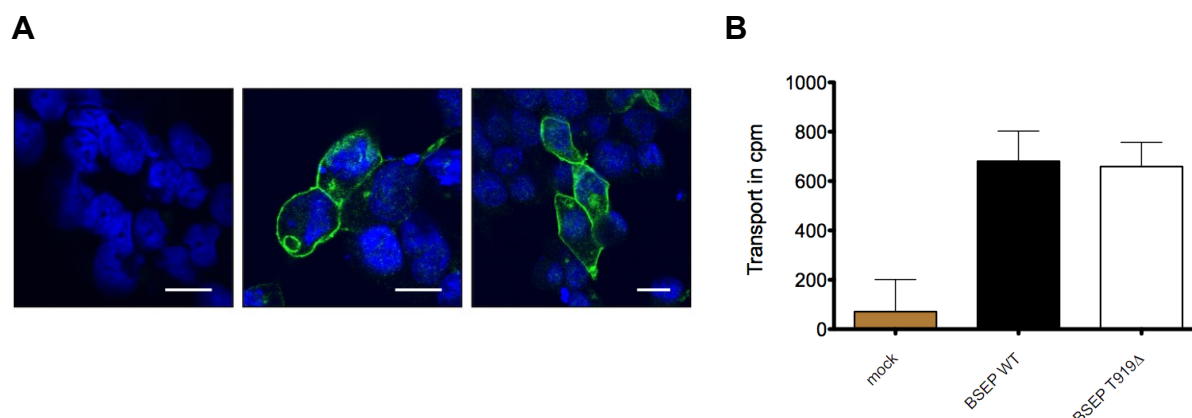
A

B

Figure 17: Characterization of the BSEP T919Δ mutant. A) Expression of human BSEP in HEK293 cells. Left: Mock control (transfected with empty plasmid), middle panel: BSEP wild-type and right: BSEP T919Δ. White bars represent 10 μm. **B)** Transport of [^{3}H]-taurocholate measured with the vesicular uptake assay (cpm = counts per minute) (n=3 ±SD)

In a BSEP homology model based on the crystal structure of P-gp with bound substrates from mouse (PDB code: 3G61), both amino acid side chains (G374 and T919) line the pore (Figure 18). Nonetheless, when compared to the P-glycoprotein structure crystallized with inhibitors, T919 is not part of the binding cavity. The distance to the allocrite is ~18 Å. Therefore, T919 might not have a crucial role and its deletion does not necessarily affect folding or stability of BSEP, so that transport of bile salts is still possible. G374 could potentially interact with its backbone or hydrogen atom of the C_α to substrate binding, but when substituted to serine, there might be repulsions of the hydroxyl group of the serine and the bile salt. Or hydrogen bonding with the hydroxyl groups of bile salts might be abolished. Although

substrates bound by P-gp do not directly interact with G374, the allocrite binding site for BSEP might be localized closer to G374.

P-gp is able to transport a variety of structurally diverse allocrites [225]. This is achieved by the use of many aromatic residues in the binding cavity, which leads to stacking and cation–π interactions and hydrophobic and hydrogen bonding interactions [108]. BSEP is also able to bind to the drug pravastatin, which is able to compete with bile salts [248]. Furthermore, BSEP is one cause for drug induced liver injury (DILI), suggesting that drugs can at least bind to BSEP and inhibit its activity [158]. BSEP may therefore also exhibit polyspecificity to some extent.
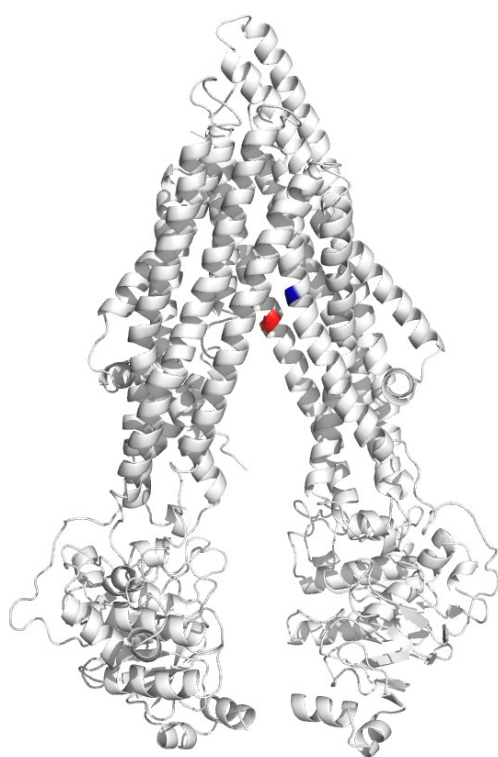
A   **B**

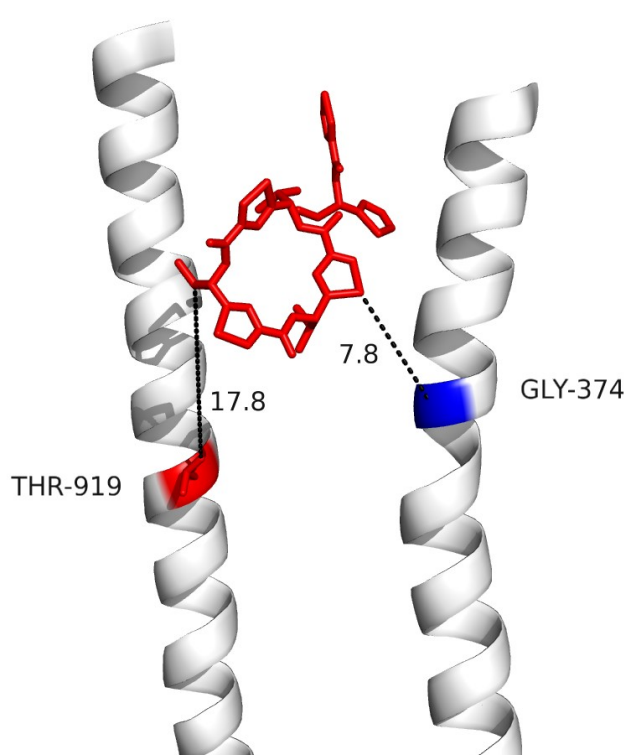**Figure 18: Homology model for human BSEP. A)** Homology model of human BSEP based on the crystal structure of mouse P-gp (PDB code: 3G61) generated with the Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). In blue the residue G374 and in red the residue T919 are depicted. **B)** An inhibitor co-crystallized with mouse P-gp is aligned with the BSEP homology model. For the sake of clarity, only the two helices with the G374 and T919 residues are shown in blue and red, respectively, and the inhibitor in red as sticks. The length is measured in Å.

The only solved structure of a bile salt transporter is a bacterial homolog of the sodium-taurocholate cotransporting peptide (NTCP) from *Neisseria meningitides*, ASBT [249]. In the co-crystal structure with taurocholate, there is only one direct hydrogen bonding between asparagine 295 and the 7α-hydroxyl group of taurocholate. Although the structure represents the inward-facing conformation,

which should display the low-affinity binding site, there might be only weak interactions between taurocholate and ASBT, in general. This could also be true for BSEP explaining its "polyspecificity" in regard to different bile salts and drugs. Further structure-function analysis of BSEP might give a more specific answer to the location of the allocrite binding site and coupling mechanism. Clinically mutations are a valuable tool, but the 3D structure of BSEP will greatly contribute to an understanding of the transport mechanism. This will require further efforts. This doctoral thesis has laid the fundament for the future biochemical and structural characterization of human BSEP in its isolated form to achieve this goal. A cloning and mutagenesis procedure was established, which simplifies BSEP cloning and the introduction of clinically, relevant mutations. In principle, this system can be used for any other instable cDNA. Furthermore, the BSEP expression system in *P. pastoris* can be used for its purification on the one hand and for the subsequent functional characterization of BSEP mutants in a transport assay on the other hand.

4. References

1. Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, 5th edition (publisher:WH Freeman).
2. Westheimer FH (1987) Why nature chose phosphates. Science 235: 1173-1178.
3. Shikama K (1971) Standard free energy maps for the hydrolysis of ATP as a function of pH, pMg and pCa. Archives of biochemistry and biophysics 147: 311-317.
4. Vetter IR, Wittinghofer A (1999) Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. Quarterly reviews of biophysics 32: 1-56.
5. Koonin EV, Wolf YI, Aravind L (2000) Protein fold recognition using sequence profiles and its application in structural genomics. Advances in protein chemistry 54: 245-275.
6. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO journal 1: 945-951.
7. Schulz GE, Schirmer RH Principles of Protein Structure, Springer, Heidelberg, 1979.
8. Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. Trends in biochemical sciences 15: 430-434.
9. Matte A, Delbaere LTJ (2001) ATP-binding Motifs. In: Encyclopedia of Life Sciences (ELS), John Wiley & Sons, Ltd: Chister.
10. Schulz GE (1992) Binding of nucleotides by proteins. Current opinion in structural biology 2: 61-67.
11. Leipe DD, Koonin EV, Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. Journal of molecular biology 333: 781-815.
12. Leipe DD, Wolf YI, Koonin EV, Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. Journal of molecular biology 317: 41-72.
13. Oldham ML, Chen J (2011) Snapshots of the maltose transporter during ATP hydrolysis. Proceedings of the National Academy of Sciences of the United States of America 108: 15152-15156.
14. Forde A, Fitzgerald GF (1999) Bacteriophage defence systems in lactic acid bacteria. Antonie van Leeuwenhoek 76: 89-113.
15. Hyman P, Abedon ST (2010) Bacteriophage host range and bacterial resistance. Advances in applied microbiology 70: 217-248.
16. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, et al. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic acids research 31: 1805-1812.
17. Chopin MC, Chopin A, Bidnenko E (2005) Phage abortive infection in lactococci: variations on a theme. Current opinion in microbiology 8: 473-479.
18. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of bacteriology 169: 5429-5433.
19. Jansen R, Embden JD, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Molecular microbiology 43: 1565-1575.
20. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551-2561.
21. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution 60: 174-182.
22. Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151: 653-663.
23. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709-1712.
24. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nature reviews Genetics 11: 181-190.
25. Wiedenheft B, Sternberg SH, Doudna JA (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482: 331-338.
26. Westra ER, Swarts DC, Staals RH, Jore MM, Brouns SJ, et al. (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. Annual review of genetics 46: 311-339.

27. Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome biology 8: R61.

28. Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, et al. (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Molecular microbiology 75: 1495-1512.

29. Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS computational biology 1: e60.

30. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. (2011) Evolution and classification of the CRISPR-Cas systems. Nature reviews Microbiology 9: 467-477.

31. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155: 733-740.

32. Babu M, Beloglazova N, Flick R, Graham C, Skarina T, et al. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. Molecular microbiology 79: 484-502.

33. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, et al. (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure 17: 904-912.

34. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, et al. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. The Journal of biological chemistry 283: 20361-20371.

35. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic acids research 40: 5569-5576.

36. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471: 602-607.

37. Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, et al. (2012) The crystal structure of the CRISPR-associated protein Csn2 from Streptococcus agalactiae. Journal of structural biology 178: 350-362.

38. Nam KH, Kurinov I, Ke A (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. The Journal of biological chemistry 286: 30759-30768.

39. Pougach K, Severinov K (2012) Use of semi-quantitative Northern blot analysis to determine relative quantities of bacterial CRISPR transcripts. Methods in molecular biology 905: 73-86.

40. Agari Y, Sakamoto K, Tamakoshi M, Oshima T, Kuramitsu S, et al. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. Journal of molecular biology 395: 270-281.

41. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321: 960-964.

42. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature structural & molecular biology 18: 529-536.

43. Gesner EM, Schellenberg MJ, Garside EL, George MM, Macmillan AM (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nature structural & molecular biology 18: 688-692.

44. Sashital DG, Jinek M, Doudna JA (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nature structural & molecular biology 18: 680-687.

45. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science 329: 1355-1358.

46. Haurwitz RE, Sternberg SH, Doudna JA (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. The EMBO journal 31: 2824-2832.

47. Nam KH, Haitjema C, Liu X, Ding F, Wang H, et al. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. Structure 20: 1574-1584.

48. Wang R, Preamplume G, Terns MP, Terns RM, Li H (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure 19: 257-264.

49. Wang R, Zheng H, Preamplume G, Shao Y, Li H (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. Protein science : a publication of the Protein Society 21: 405-417.

50. Carte J, Pfister NT, Compton MM, Terns RM, Terns MP (2010) Binding and cleavage of CRISPR RNA by Cas6. RNA 16: 2181-2188.

51. Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes & development 22: 3489-3496.

52. Hale CR, Majumdar S, Elmore J, Pfister N, Compton M, et al. (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Molecular cell 45: 292-302.

53. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139: 945-956.

54. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816-821.

55. Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468: 67-71.

56. Westra ER, van Erp PB, Kunne T, Wong SP, Staals RH, et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. Molecular cell 46: 595-605.

57. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, et al. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proceedings of the National Academy of Sciences of the United States of America 108: 10098-10103.

58. Beloglazova N, Petit P, Flick R, Brown G, Savchenko A, et al. (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. The EMBO journal 30: 4616-4627.

59. Sinkunas T, Gasiunas G, Waghmare SP, Dickman MJ, Barrangou R, et al. (2013) In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. The EMBO journal 32: 385-394.

60. Hale C, Kleppe K, Terns RM, Terns MP (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. RNA 14: 2572-2579.

61. Zhang J, Rouillon C, Kerou M, Reeks J, Brugger K, et al. (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. Molecular cell 45: 303-313.

62. Schuchat A (1999) Group B streptococcus. Lancet 353: 51-56.

63. Edwards MS, Baker CJ (2005) Group B streptococcal infections in elderly adults. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 41: 839-847.

64. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America 102: 13950-13955.

65. Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, et al. (2002) Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease. Molecular microbiology 45: 1499-1513.

66. Lopez-Sanchez MJ, Sauvage E, Da Cunha V, Clermont D, Ratsima Hariniaina E, et al. (2012) The highly dynamic CRISPR1 system of Streptococcus agalactiae controls the diversity of its mobilome. Molecular microbiology 85: 1057-1071.

67. Zhang M, Stauffacher CV, Lin D, Van Etten RL (1998) Crystal structure of a human low molecular weight phosphotyrosyl phosphatase. Implications for substrate specificity. The Journal of biological chemistry 273: 21714-21720.

68. Kinoshita K, Sadanami K, Kidera A, Go N (1999) Structural motif of phosphate-binding site common to various protein superfamilies: all-against-all structural comparison of protein-mononucleotide complexes. Protein engineering 12: 11-14.

69. Koo Y, Jung DK, Bae E (2012) Crystal structure of Streptococcus pyogenes Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. PloS one 7: e33401.

70. Lee KH, Lee SG, Eun Lee K, Jeon H, Robinson H, et al. (2012) Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. Proteins 80: 2573-2582.

71. Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. Science 175: 720-731.

72. Nagle JF, Scott HL, Jr. (1978) Lateral compressibility of lipid mono- and bilayers. Theory of membrane permeability. Biochimica et biophysica acta 513: 236-243.

73. Engelman DM (2005) Membranes are more mosaic than fluid. Nature 438: 578-580.

74. Gram C (1884) Über die isolierte Färbung der Schizomyceten in Schnitt-und Trockenpräparaten. Fortschritte der Medicin: 185-189.

75. Jain MK, Wagner RC (1988) Introduction to Biological Membranes, 2nd edition (Publisher: Wiley).

76. Saier MH, Jr., Tran CV, Barabote RD (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic acids research 34: D181-186.

77. Higgins CF (1992) ABC transporters: from microorganisms to man. Annual review of cell biology 8: 67-113.

78. Blight MA, Holland IB (1990) Structure and function of haemolysin B,P-glycoprotein and other members of a novel family of membrane translocators. Molecular microbiology 4: 873-880.

79. Higgins CF, Hiles ID, Salmond GP, Gill DR, Downie JA, et al. (1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. Nature 323: 448-450.

80. Chen CJ, Chin JE, Ueda K, Clark DP, Pastan I, et al. (1986) Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. Cell 47: 381-389.

81. Gros P, Croop J, Housman D (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47: 371-380.

82. Gerlach JH, Endicott JA, Juranka PF, Henderson G, Sarangi F, et al. (1986) Homology between P-glycoprotein and a bacterial haemolysin transport protein suggests a model for multidrug resistance. Nature 324: 485-489.

83. Gilson E, Higgins CF, Hofnung M, Ferro-Luzzi Ames G, Nikaido H (1982) Extensive homology between membrane-associated components of histidine and maltose transport systems of Salmonella typhimurium and Escherichia coli. The Journal of biological chemistry 257: 9915-9918.

84. Higgins CF, Haag PD, Nikaido K, Ardeshir F, Garcia G, et al. (1982) Complete nucleotide sequence and identification of membrane components of the histidine transport operon of S. typhimurium. Nature 298: 723-727.

85. Holland IB (2011) ABC transporters, mechanisms and biology: an overview. Essays in biochemistry 50: 1-17.

86. Locher KP (2009) Review. Structure and mechanism of ATP-binding cassette transporters. Philosophical transactions of the Royal Society of London Series B, Biological sciences 364: 239-245.

87. Hebbeln P, Rodionov DA, Alfandega A, Eitinger T (2007) Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. Proceedings of the National Academy of Sciences of the United States of America 104: 2909-2914.

88. Rodionov DA, Hebbeln P, Eudes A, ter Beek J, Rodionova IA, et al. (2009) A novel class of modular transporters for vitamins in prokaryotes. Journal of bacteriology 191: 42-51.

89. Erkens GB, Berntsson RP, Fulyani F, Majsnerowska M, Vujicic-Zagar A, et al. (2011) The structural basis of modularity in ECF-type ABC transporters. Nature structural & molecular biology 18: 755-760.

90. Karpowich NK, Wang DN (2013) Assembly and mechanism of a group II ECF transporter. Proceedings of the National Academy of Sciences of the United States of America 110: 2534-2539.

91. Erkens GB, Majsnerowska M, ter Beek J, Slotboom DJ (2012) Energy coupling factor-type ABC transporters for vitamin uptake in prokaryotes. Biochemistry 51: 4390-4396.

92. Zolnerciks JK, Andress EJ, Nicolaou M, Linton KJ (2011) Structure of ABC transporters. Essays in biochemistry 50: 43-61.

93. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B (2010) A structural classification of substrate-binding proteins. FEBS letters 584: 2606-2617.

94. Biemans-Oldehinkel E, Doeven MK, Poolman B (2006) ABC transporter architecture and regulatory roles of accessory domains. FEBS letters 580: 1023-1035.

95. Holland IB (2003) ABC proteins: from bacteria to man (Publisher: Academic Press).

96. Oldham ML, Khare D, Quiocho FA, Davidson AL, Chen J (2007) Crystal structure of a catalytic intermediate of the maltose transporter. Nature 450: 515-521.

97. Hvorup RN, Goetz BA, Niederer M, Hollenstein K, Perozo E, et al. (2007) Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. Science 317: 1387-1390.

98. Dawson RJ, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. Nature 443: 180-185.

99. Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nature reviews Molecular cell biology 10: 218-227.

100. Hollenstein K, Frei DC, Locher KP (2007) Structure of an ABC transporter in complex with its binding protein. Nature 446: 213-216.

101. Kadaba NS, Kaiser JT, Johnson E, Lee A, Rees DC (2008) The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation. Science 321: 250-253.

102. Locher KP, Lee AT, Rees DC (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 296: 1091-1098.

103. Pinkett HW, Lee AT, Lum P, Locher KP, Rees DC (2007) An inward-facing conformation of a putative metal-chelate-type ABC transporter. Science 315: 373-377.

104. Ward A, Reyes CL, Yu J, Roth CB, Chang G (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proceedings of the National Academy of Sciences of the United States of America 104: 19005-19010.

105. Aller SG, Yu J, Ward A, Weng Y, Chittaboina S, et al. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323: 1718-1722.

106. Hohl M, Briand C, Grutter MG, Seeger MA (2012) Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nature structural & molecular biology 19: 395-402.

107. Linton KJ (2007) Structure and function of ABC transporters. Physiology 22: 122-130.

108. Gutmann DA, Ward A, Urbatsch IL, Chang G, van Veen HW (2010) Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. Trends in biochemical sciences 35: 36-42.

109. Kimura Y, Kodan A, Matsuo M, Ueda K (2007) Cholesterol fill-in model: mechanism for substrate recognition by ABC proteins. Journal of bioenergetics and biomembranes 39: 447-452.

110. Lewinson O, Lee AT, Locher KP, Rees DC (2010) A distinct mechanism for the ABC transporter BtuCD-BtuF revealed by the dynamics of complex formation. Nature structural & molecular biology 17: 332-338.

111. Oldham ML, Davidson AL, Chen J (2008) Structural insights into ABC transporter mechanism. Current opinion in structural biology 18: 726-733.

112. Schmitt L, Tampe R (2002) Structure and mechanism of ABC transporters. Current opinion in structural biology 12: 754-760.

113. Oswald C, Holland IB, Schmitt L (2006) The motor domains of ABC-transporters. What can structures tell us? Naunyn-Schmiedeberg's archives of pharmacology 372: 385-399.

114. Chen J, Lu G, Lin J, Davidson AL, Quiocho FA (2003) A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Molecular cell 12: 651-661.

115. Karpowich N, Martsinkevich O, Millen L, Yuan YR, Dai PL, et al. (2001) Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. Structure 9: 571-586.

116. Schmitt L, Benabdelhak H, Blight MA, Holland IB, Stubbs MT (2003) Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. Journal of molecular biology 330: 333-342.

117. Zaitseva J, Jenewein S, Jumpertz T, Holland IB, Schmitt L (2005) H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. The EMBO journal 24: 1901-1910.

118. Zaitseva J, Oswald C, Jumpertz T, Jenewein S, Wiedenmann A, et al. (2006) A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. The EMBO journal 25: 3432-3443.

119. Cui J, Davidson AL (2011) ABC solute importers in bacteria. Essays in biochemistry 50: 85-99.

120. Higgins CF, Linton KJ (2004) The ATP switch model for ABC transporters. Nature structural & molecular biology 11: 918-926.

121. Linton KJ, Higgins CF (2007) Structure and function of ABC transporters: the ATP switch provides flexible control. Pflugers Archiv : European journal of physiology 453: 555-567.

122. Jardetzky O (1966) Simple allosteric model for membrane pumps. Nature 211: 969-970.

123. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. Genome research 11: 1156-1166.

124. Dean M, Hamon Y, Chimini G (2001) The human ATP-binding cassette (ABC) transporter superfamily. Journal of lipid research 42: 1007-1017.

125. Piehler AP, Hellum M, Wenzel JJ, Kaminski E, Haug KB, et al. (2008) The human ABC transporter pseudogene family: Evidence for transcription and gene-pseudogene interference. BMC genomics 9: 165.

126. Moradpour D (2006) Leber, in: Klinische Pathophysiologie (Eds. Siegenthaler W., Blum HE.), Thieme, Stuttgart, Chapter 32.

127. Haussinger D, Kubitz R, Reinehr R, Bode JG, Schliess F (2004) Molecular aspects of medicine: from experimental to clinical hepatology. Molecular aspects of medicine 25: 221-360.

128. Meier PJ, Stieger B (2002) Bile salt transporters. Annual review of physiology 64: 635-661.

129. Hofmann AF (2009) Bile acids: trying to understand their chemistry and biology with the hope of helping patients. Hepatology 49: 1403-1418.

130. Hofmann AF, Hagey LR (2008) Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. Cellular and molecular life sciences : CMLS 65: 2461-2483.

131. Russell DW (2009) Fifty years of advances in bile acid synthesis and metabolism. Journal of lipid research 50 Suppl: S120-125.

132. Russell DW (2003) The enzymes, regulation, and genetics of bile acid synthesis. Annual review of biochemistry 72: 137-174.

133. Trauner M, Fickert P, Halilbasic E, Moustafa T (2008) Lessons from the toxic bile concept for the pathogenesis and treatment of cholestatic liver diseases. Wiener medizinische Wochenschrift 158: 542-548.

134. Northfield TC, Kupfer RM, Maudgal DP, Zentler-Munro PL, Meller ST, et al. (1980) Gall-bladder sensitivity to cholecystokinin in patients with gall stones. British medical journal 280: 143-144.

135. Hofmann AF (2009) The enterohepatic circulation of bile acids in mammals: form and functions. Frontiers in bioscience : a journal and virtual library 14: 2584-2598.

136. Hofmann AF (1999) Bile Acids: The Good, the Bad, and the Ugly. News in physiological sciences : an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society 14: 24-29.

137. Fausto N, Campbell JS (2003) The role of hepatocytes and oval cells in liver regeneration and repopulation. Mechanisms of development 120: 117-130.

138. Faber KN, Muller M, Jansen PL (2003) Drug transport proteins in the liver. Advanced drug delivery reviews 55: 107-124.

139. Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM (2003) P-glycoprotein: from genomics to mechanism. Oncogene 22: 7468-7485.

140. Ni Z, Bikadi Z, Rosenberg MF, Mao Q (2010) Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). Current drug metabolism 11: 603-617.

141. Nies AT, Keppler D (2007) The apical conjugate efflux pump ABCC2 (MRP2). Pflugers Archiv : European journal of physiology 453: 643-659.

142. Zeng H, Liu G, Rea PA, Kruh GD (2000) Transport of amphipathic anions by human multidrug resistance protein 3. Cancer research 60: 4779-4784.

143. Akita H, Suzuki H, Hirohashi T, Takikawa H, Sugiyama Y (2002) Transport activity of human MRP3 expressed in Sf9 cells: comparative studies with rat MRP3. Pharmaceutical research 19: 34-41.

144. Schaap FG, van der Gaag NA, Gouma DJ, Jansen PL (2009) High expression of the bile salt-homeostatic hormone fibroblast growth factor 19 in the liver of patients with extrahepatic cholestasis. Hepatology 49: 1228-1235.

145. Gradhand U, Lang T, Schaeffeler E, Glaeser H, Tegude H, et al. (2008) Variability in human hepatic MRP4 expression: influence of cholestasis and genotype. The pharmacogenomics journal 8: 42-52.

146. Boyer JL, Trauner M, Mennone A, Soroka CJ, Cai SY, et al. (2006) Upregulation of a basolateral FXR-dependent bile acid efflux transporter OSTalpha-OSTbeta in cholestasis in humans and rodents. American journal of physiology Gastrointestinal and liver physiology 290: G1124-1130.

147. Keppler D (2011) Multidrug resistance proteins (MRPs, ABCCs): importance for pathophysiology and drug therapy. Handbook of experimental pharmacology: 299-323.

148. Stieger B (2011) The role of the sodium-taurocholate cotransporting polypeptide (NTCP) and of the bile salt export pump (BSEP) in physiology and pathophysiology of bile formation. Handbook of experimental pharmacology: 205-259.

149. Trauner M, Boyer JL (2003) Bile salt transporters: molecular characterization, function, and regulation. Physiological reviews 83: 633-671.

150. Gerloff T, Stieger B, Hagenbuch B, Madon J, Landmann L, et al. (1998) The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver. The Journal of biological chemistry 273: 10046-10050.

151. Noe J, Stieger B, Meier PJ (2002) Functional expression of the canalicular bile salt export pump of human liver. Gastroenterology 123: 1659-1666.

152. Byrne JA, Strautnieks SS, Mieli-Vergani G, Higgins CF, Linton KJ, et al. (2002) The human bile salt export pump: characterization of substrate specificity and identification of inhibitors. Gastroenterology 123: 1649-1658.

153. Keitel V, Burdelski M, Warskulat U, Kuhlkamp T, Keppler D, et al. (2005) Expression and localization of hepatobiliary transport proteins in progressive familial intrahepatic cholestasis. Hepatology 41: 1160-1172.

154. Strautnieks SS, Byrne JA, Pawlikowska L, Cebecauerova D, Rayner A, et al. (2008) Severe bile salt export pump deficiency: 82 different ABCB11 mutations in 109 families. Gastroenterology 134: 1203-1214.

155. Bull LN, Carlton VE, Stricker NL, Baharloo S, DeYoung JA, et al. (1997) Genetic and morphological findings in progressive familial intrahepatic cholestasis (Byler disease [PFIC-1] and Byler syndrome): evidence for heterogeneity. Hepatology 26: 155-164.

156. van Mil SW, van der Woerd WL, van der Brugge G, Sturm E, Jansen PL, et al. (2004) Benign recurrent intrahepatic cholestasis type 2 is caused by mutations in ABCB11. Gastroenterology 127: 379-384.

157. Jacquemin E (2012) Progressive familial intrahepatic cholestasis. Clinics and research in hepatology and gastroenterology 36 Suppl 1: S26-35.

158. Dawson S, Stahl S, Paul N, Barber J, Kenna JG (2012) In vitro inhibition of the bile salt export pump correlates with risk of cholestatic drug-induced liver injury in humans. Drug metabolism and disposition: the biological fate of chemicals 40: 130-138.

159. Dixon PH, van Mil SW, Chambers J, Strautnieks S, Thompson RJ, et al. (2009) Contribution of variant alleles of ABCB11 to susceptibility to intrahepatic cholestasis of pregnancy. Gut 58: 537-544.

160. Han D, Lehmann K, Krauss G (2009) SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA. FEBS letters 583: 1928-1932.

161. Samai P, Smith P, Shuman S (2010) Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. Acta crystallographica Section F, Structural biology and crystallization communications 66: 1552-1556.

162. Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP, et al. (2012) Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. The Journal of biological chemistry 287: 35943-35952.

163. Plagens A, Tjaden B, Hagemann A, Randau L, Hensel R (2012) Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax. Journal of bacteriology 194: 2491-2500.

164. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology direct 1: 7.

165. Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic acids research 38: W545-549.

166. Tuteja N, Tuteja R (2004) Unraveling DNA helicases. Motif, structure, mechanism and function. European journal of biochemistry / FEBS 271: 1849-1863.

167. Frick DN, Richardson CC (2001) DNA primases. Annual review of biochemistry 70: 39-80.

168. Kojima KK, Kanehisa M (2008) Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Molecular biology and evolution 25: 1395-1404.

169. Gulbis JM, Kelman Z, Hurwitz J, O'Donnell M, Kuriyan J (1996) Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. Cell 87: 297-306.

170. Tessmer I, Moore T, Lloyd RG, Wilson A, Erie DA, et al. (2005) AFM studies on the role of the protein RdgC in bacterial DNA recombination. Journal of molecular biology 350: 254-262.

171. Pitcher RS, Brissett NC, Doherty AJ (2007) Nonhomologous end-joining in bacteria: a microbial perspective. Annual review of microbiology 61: 259-282.

172. Shuman S, Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. Nature reviews Microbiology 5: 852-861.

173. Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, et al. (2002) Identification of a DNA nonhomologous end-joining complex in bacteria. Science 297: 1686-1689.

174. Weller GR, Doherty AJ (2001) A family of DNA repair ligases in bacteria? FEBS letters 505: 340-342.

175. Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. Nature 412: 607-614.

176. Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology 72: 211-222.

177. Wagner S, Bader ML, Drew D, de Gier JW (2006) Rationalizing membrane protein overexpression. Trends in biotechnology 24: 364-371.

178. Midgett CR, Madden DR (2007) Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies. Journal of structural biology 160: 265-274.

179. Hunte C (2005) Specific protein-lipid interactions in membrane proteins. Biochemical Society transactions 33: 938-942.

180. Bibi E, Gros P, Kaback HR (1993) Functional expression of mouse mdr1 in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 90: 9209-9213.

181. Linton KJ, Higgins CF (2002) P-glycoprotein misfolds in Escherichia coli: evidence against alternating-topology models of the transport cycle. Molecular membrane biology 19: 51-58.

182. Pozza A, Perez-Victoria JM, Di Pietro A (2010) Insect cell versus bacterial overexpressed membrane proteins: an example, the human ABCG2 transporter. Methods in molecular biology 654: 47-75.

183. Loll PJ (2003) Membrane protein structural biology: the high throughput challenge. Journal of structural biology 142: 144-153.

184. Martinez Molina D, Cornvik T, Eshaghi S, Haeggstrom JZ, Nordlund P, et al. (2008) Engineering membrane protein overproduction in Escherichia coli. Protein science : a publication of the Protein Society 17: 673-680.

185. Monne M, Chan KW, Slotboom DJ, Kunji ER (2005) Functional expression of eukaryotic membrane proteins in Lactococcus lactis. Protein science : a publication of the Protein Society 14: 3048-3056.

186. Hayashi H, Takada T, Suzuki H, Onuki R, Hofmann AF, et al. (2005) Transport by vesicles of glycine- and taurine-conjugated bile salts and taurolithocholate 3-sulfate: a comparison of human BSEP with rat Bsep. Biochimica et biophysica acta 1738: 54-62.

187. Hayashi H, Takada T, Suzuki H, Akita H, Sugiyama Y (2005) Two common PFIC2 mutations are associated with the impaired membrane trafficking of BSEP/ABCB11. Hepatology 41: 916-924.

188. Kubitz R, Sutfels G, Kuhlkamp T, Kolling R, Haussinger D (2004) Trafficking of the bile salt export pump from the Golgi to the canalicular membrane is regulated by the p38 MAP kinase. Gastroenterology 126: 541-553.

189. Mita S, Suzuki H, Akita H, Hayashi H, Onuki R, et al. (2006) Vectorial transport of unconjugated and conjugated bile salts by monolayers of LLC-PK1 cells doubly transfected with human NTCP and BSEP or with rat Ntcp and Bsep. American journal of physiology Gastrointestinal and liver physiology 290: G550-556.

190. Hagmann W, Nies AT, Konig J, Frey M, Zentgraf H, et al. (1999) Purification of the human apical conjugate export pump MRP2 reconstitution and functional characterization as substrate-stimulated ATPase. European journal of biochemistry / FEBS 265: 281-289.

191. Kodan A, Shibata H, Matsumoto T, Terakado K, Sakiyama K, et al. (2009) Improved expression and purification of human multidrug resistance protein MDR1 from baculovirus-infected insect cells. Protein expression and purification 66: 7-14.

192. McDevitt CA, Collins RF, Conway M, Modok S, Storm J, et al. (2006) Purification and 3D structural analysis of oligomeric human multidrug transporter ABCG2. Structure 14: 1623-1632.

193. Freigassner M, Pichler H, Glieder A (2009) Tuning microbial hosts for membrane protein production. Microbial cell factories 8: 69.

194. Jidenko M, Nielsen RC, Sorensen TL, Moller JV, le Maire M, et al. (2005) Crystallization of a mammalian membrane protein overexpressed in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 102: 11687-11691.

195. Palanivelu DV, Kozono DE, Engel A, Suda K, Lustig A, et al. (2006) Co-axial association of recombinant eye lens aquaporin-0 observed in loosely packed 3D crystals. Journal of molecular biology 355: 605-611.

196. Mao Q, Scarborough GA (1997) Purification of functional human P-glycoprotein expressed in Saccharomyces cerevisiae. Biochimica et biophysica acta 1327: 107-118.

197. Lee SH, Altenberg GA (2003) Expression of functional multidrug-resistance protein 1 in Saccharomyces cerevisiae: effects of N- and C-terminal affinity tags. Biochemical and biophysical research communications 306: 644-649.

198. Chloupkova M, Pickert A, Lee JY, Souza S, Trinh YT, et al. (2007) Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry 46: 7992-8003.

199. Zehnpfennig B, Urbatsch IL, Galla HJ (2009) Functional reconstitution of human ABCC3 into proteoliposomes reveals a transport mechanism with positive cooperativity. Biochemistry 48: 4423-4430.

200. Johnson BJ, Lee JY, Pickert A, Urbatsch IL (2010) Bile acids stimulate ATP hydrolysis in the purified cholesterol transporter ABCG5/G8. Biochemistry 49: 3403-3411.

201. Wang Z, Stalcup LD, Harvey BJ, Weber J, Chloupkova M, et al. (2006) Purification and ATP hydrolysis of the putative cholesterol transporters ABCG5 and ABCG8. Biochemistry 45: 9929-9939.

202. Wang G, Christensen LA, Vasquez KM (2006) Z-DNA-forming sequences generate large-scale deletions in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America 103: 2677-2682.

203. Miroux B, Walker JE (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. Journal of molecular biology 260: 289-298.

204. Vu K, Bautos J, Hong MP, Gelli A (2009) The functional expression of toxic genes: lessons learned from molecular cloning of CCH1, a high-affinity Ca2+ channel. Analytical biochemistry 393: 234-241.

205. Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58: 201-216.

206. Braman J, Papworth C, Greener A (1996) Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods in molecular biology 57: 31-44.

207. Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology 60: 523-533.

208. Seddon AM, Curnow P, Booth PJ (2004) Membrane proteins, lipids and detergents: not just a soap opera. Biochimica et biophysica acta 1666: 105-117.

209. Zeder-Lutz G, Cherouati N, Reinhart C, Pattus F, Wagner R (2006) Dot-blot immunodetection as a versatile and high-throughput assay to evaluate recombinant GPCRs produced in the yeast Pichia pastoris. Protein expression and purification 50: 118-127.

210. Lerner-Marmarosh N, Gimi K, Urbatsch IL, Gros P, Senior AE (1999) Large scale purification of detergent-soluble P-glycoprotein from Pichia pastoris cells and characterization of nucleotide binding properties of wild-type, Walker A, and Walker B mutant proteins. The Journal of biological chemistry 274: 34711-34718.

211. Beaudet L, Urbatsch IL, Gros P (1998) Mutations in the nucleotide-binding sites of P-glycoprotein that affect substrate specificity modulate substrate-induced adenosine triphosphatase activity. Biochemistry 37: 9073-9082.

212. Bai J, Swartz DJ, Protasevich, II, Brouillette CG, Harrell PM, et al. (2011) A gene optimization strategy that enhances production of fully functional P-glycoprotein in Pichia pastoris. PloS one 6: e22577.

213. Urbatsch IL, Beaudet L, Carrier I, Gros P (1998) Mutations in either nucleotide-binding site of P-glycoprotein (Mdr3) prevent vanadate trapping of nucleotide at both sites. Biochemistry 37: 4592-4602.

214. Rosenberg MF, Bikadi Z, Chan J, Liu X, Ni Z, et al. (2010) The human breast cancer resistance protein (BCRP/ABCG2) shows conformational changes with mitoxantrone. Structure 18: 482-493.

215. Lewinson O, Lee AT, Rees DC (2008) The funnel approach to the precrystallization production of membrane proteins. Journal of molecular biology 377: 62-73.

216. Kawate T, Gouaux E (2006) Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14: 673-681.

217. Infed N, Hanekop N, Driessen AJ, Smits SH, Schmitt L (2011) Influence of detergents on the activity of the ABC transporter LmrA. Biochimica et biophysica acta 1808: 2313-2321.

218. Eckford PD, Sharom FJ (2008) Functional characterization of Escherichia coli MsbA: interaction with nucleotides and substrates. The Journal of biological chemistry 283: 12840-12850.

219. Swartz DJ, Weber J, Urbatsch IL (2013) P-glycoprotein is fully active after multiple tryptophan substitutions. Biochimica et biophysica acta 1828: 1159-1168.

220. Zaitseva J, Jenewein S, Wiedenmann A, Benabdelhak H, Holland IB, et al. (2005) Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter. Biochemistry 44: 9680-9690.

221. Sauna ZE, Ambudkar SV (2001) Characterization of the catalytic cycle of ATP hydrolysis by human P-glycoprotein. The two ATP hydrolysis events in a single catalytic cycle are kinetically similar but affect different functional outcomes. The Journal of biological chemistry 276: 11653-11661.

222. Morita M, Kurisu M, Kashiwayama Y, Yokota S, Imanaka T (2006) ATP-binding and -hydrolysis activities of ALDP (ABCD1) and ALDRP (ABCD2), human peroxisomal ABC proteins, overexpressed in Sf21 cells. Biological & pharmaceutical bulletin 29: 1836-1842.

223. Ernst R (2007) Molekulare Analyse von ABC-Transportern am Beispiel von Pdr5p aus S. cerevisiae und der isolierten NBD des humanen TAP1. Dissertation, Heinrich-Heine-Universität Düsseldorf.

224. Kis E, Ioja E, Nagy T, Szente L, Heredi-Szabo K, et al. (2009) Effect of membrane cholesterol on BSEP/Bsep activity: species specificity studies for substrates and inhibitors. Drug metabolism and disposition: the biological fate of chemicals 37: 1878-1886.

225. Sharom FJ (2011) The P-glycoprotein multidrug transporter. Essays in biochemistry 50: 161-178.

226. Ernst R, Kueppers P, Klein CM, Schwarzmueller T, Kuchler K, et al. (2008) A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. Proceedings of the National Academy of Sciences of the United States of America 105: 5069-5074.

227. van Staden CJ, Morgan RE, Ramachandran B, Chen Y, Lee PH, et al. (2012) Membrane vesicle ABC transporter assays for drug safety assessment. Current protocols in toxicology / editorial board, Mahin D Maines Chapter 23: Unit 23 25.

228. Steck TL, Weinstein RS, Straus JH, Wallach DF (1970) Inside-out red cell membrane vesicles: preparation and purification. Science 168: 255-257.

229. Storch CH, Ehehalt R, Haefeli WE, Weiss J (2007) Localization of the human breast cancer resistance protein (BCRP/ABCG2) in lipid rafts/caveolae and modulation of its activity by cholesterol in vitro. The Journal of pharmacology and experimental therapeutics 323: 257-264.

230. Lavie Y, Liscovitch M (2000) Changes in lipid and protein constituents of rafts and caveolae in multidrug resistant cancer cells and their functional consequences. Glycoconjugate journal 17: 253-259.

231. Guyot C, Stieger B (2011) Interaction of bile salts with rat canalicular membrane vesicles: evidence for bile salt resistant microdomains. Journal of hepatology 55: 1368-1376.

232. Lucero HA, Robbins PW (2004) Lipid rafts-protein association and the regulation of protein activity. Archives of biochemistry and biophysics 426: 208-224.

233. Stieger B, Meier Y, Meier PJ (2007) The bile salt export pump. Pflugers Archiv : European journal of physiology 453: 611-620.

234. Nicolaou M, Andress EJ, Zolnerciks JK, Dixon PH, Williamson C, et al. (2012) Canalicular ABC transporters and liver disease. The Journal of pathology 226: 300-315.

235. Cantore M, Reinehr R, Sommerfeld A, Becker M, Haussinger D (2011) The Src family kinase Fyn mediates hyperosmolarity-induced Mrp2 and Bsep retrieval from canalicular membrane. The Journal of biological chemistry 286: 45014-45029.

236. Crocenzi FA, Mottino AD, Sanchez Pozzi EJ, Pellegrino JM, Rodriguez Garay EA, et al. (2003) Impaired localisation and transport function of canalicular Bsep in taurolithocholate induced cholestasis in the rat. Gut 52: 1170-1177.

237. Strautnieks SS, Kagalwalla AF, Tanner MS, Knisely AS, Bull L, et al. (1997) Identification of a locus for progressive familial intrahepatic cholestasis PFIC2 on chromosome 2q24. American journal of human genetics 61: 630-633.

238. Byrne JA, Strautnieks SS, Ihrke G, Pagani F, Knisely AS, et al. (2009) Missense mutations and single nucleotide polymorphisms in ABCB11 impair bile salt export pump processing and function or disrupt pre-messenger RNA splicing. Hepatology 49: 553-567.

239. Carlini DB, Genut JE (2006) Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. Journal of molecular evolution 62: 89-98.

240. Jansen PL, Strautnieks SS, Jacquemin E, Hadchouel M, Sokal EM, et al. (1999) Hepatocanalicular bile salt export pump deficiency in patients with progressive familial intrahepatic cholestasis. Gastroenterology 117: 1370-1379.

241. Wang L, Dong H, Soroka CJ, Wei N, Boyer JL, et al. (2008) Degradation of the bile salt export pump at endoplasmic reticulum in progressive familial intrahepatic cholestasis type II. Hepatology 48: 1558-1569.

242. Mochizuki K, Kagawa T, Numari A, Harris MJ, Itoh J, et al. (2007) Two N-linked glycans are required to maintain the transport activity of the bile salt export pump (ABCB11) in MDCK II cells. American journal of physiology Gastrointestinal and liver physiology 292: G818-828.

243. Davit-Spraul A, Fabre M, Branchereau S, Baussan C, Gonzales E, et al. (2010) ATP8B1 and ABCB11 analysis in 62 children with normal gamma-glutamyl transferase progressive familial intrahepatic cholestasis (PFIC): phenotypic differences between PFIC1 and PFIC2 and natural history. Hepatology 51: 1645-1655.

244. Pauli-Magnus C, Kerb R, Fattinger K, Lang T, Anwald B, et al. (2004) BSEP and MDR3 haplotype structure in healthy Caucasians, primary biliary cirrhosis and primary sclerosing cholangitis. Hepatology 39: 779-791.

245. Liu LY, Wang ZL, Wang XH, Zhu QR, Wang JS (2010) ABCB11 gene mutations in Chinese children with progressive intrahepatic cholestasis and low gamma glutamyltransferase. Liver international : official journal of the International Association for the Study of the Liver 30: 809-815.

246. Matte U, Mourya R, Miethke A, Liu C, Kauffmann G, et al. (2010) Analysis of gene mutations in children with cholestasis of undefined etiology. Journal of pediatric gastroenterology and nutrition 51: 488-493.

247. Wang L, Soroka CJ, Boyer JL (2002) The role of bile salt export pump mutations in progressive familial intrahepatic cholestasis type II. The Journal of clinical investigation 110: 965-972.

248. Hirano M, Maeda K, Hayashi H, Kusuhara H, Sugiyama Y (2005) Bile salt export pump (BSEP/ABCB11) can transport a nonbile acid substrate, pravastatin. The Journal of pharmacology and experimental therapeutics 314: 876-882.

249. Hu NJ, Iwata S, Cameron AD, Drew D (2011) Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. Nature 478: 408-411.

5. Abbreviations

aa	amino acid
AAA+	ATPases Associated with diverse cellular Activities
ABC	ATP binding cassette
ASBT	apical sodium-dependent bile acid transporter
ADP	adenosine-5'-diphosphate
ANP	phosphoaminophosphonic acid-adenylate ester
ATP	adenosine-5'-triphosphate
BCRP	Breast Cancer Resistance Protein
bp	base pair
BRIC2	benign recurrent intrahepatic cholestasis type 2
BSEP	Bile Salt Export Pump
Cas	CRISPR-associated
Cascade	CRISPR- associated complex for antiviral defence
CBP	calmodulin binding peptide
cDNA	complementary DNA
CH	coupling helix
cmc	critical micellar concentration
cpm	counts per minute
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR-RNA
CSD	catalytic subdomain
3D	three-dimensional
β-DDM	n-dodecyl β-D-maltoside
β-DM	n-decyl β-D-maltoside
DIC	drug-induced cholestasis
DNA	deoxyribonucleic acid
DRM	detergent resistant microdomain
DSB	double-strand break
dsDNA	double-stranded DNA
E. coli	*Escherichia coli*
E. faecalis	*Enterococcus faecalis*
e.g.	exempli gratia
ESE	exonic splicing enhancer
FSEC	fluorescence-detection size-exclusion chromatography

g	gram
GNP	phosphoaminophosphonic acid-guanylate ester
GTP	guanosine-5'-triphosphate
H^+	proton
HEK293	human embryonic kidney 293 cells
HepG2	hepatocellular carcinoma G2 cell line
his_{6x}-tag	histidine-tag
HR	homologous recombination
HSD	helical subdomain
ICP	intrahepatic cholestasis during pregnancy
IMP	integral membrane protein
IUBMB	International Union of Biochemistry and Molecular Biology
K_d	dissociation constant
kDa	kilo Dalton
kg	kilogram
kJ	kilo Joule
K_M	Michaelis-Menten constant
L. lactis	*Lactococcus lactis*
LLC PK1	pig kidney epithelial cells
MDCK	Madin Darby canine kidney
MDR	multidrug resistance
mg	milligram
µg	microgram
µM	micromolar
MRP	multidrug resistance related protein
NBD	nucleotide binding domain
NHEJ	non-homologous end-joining
Ni^{2+}-IDA	nickel(2+)- iminodiacetic acid
nt	nucleotide
NTCP	sodium-taurocholate cotransporting polypeptide
OATP	organic anion-transporting polypeptide
OST	organic solute transporter
P_i	inorganic phosphate
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PEBD	partial external biliary diversion
PFIC2	progressive familial intrahepatic cholestasis type 2

P. aeruginosa	*Pseudomonas aeruginosa*
P-gp	permeability glycoprotein 1
P-loop	phosphate-binding loop
P. pastoris	*Pichia pastoris*
R-M system	restriction-modification system
RNA	ribonucleic acid
S. agalactiae	*Streptococcus agalactiae*
S. cerevisiae	*Saccharomyces cerevisiae*
SF1/2	superfamily 1/2
*Sf*9	*Spodoptera frugiperda 9*
SLC	solute carrier
SNP	single nucleotide polymorphism
S. pyogenes	*Streptococcus pyogenes*
ssRNA	single-stranded RNA
S. solfataricus	*Sulfolobus solfataricus*
S. thermophilus	*Streptococcus thermophilus*
TMD	transmembrane domain
TNP-ATP	2′,3′-O-(2,4,6-trinitrophenyl) adenosine 5′-triphosphate
tracrRNA	tracer-RNA

Amino acid	Three/One letter code	Amino acid	Three/One letter code
alanine	Ala / A	leucine	Leu / L
arginine	Arg / R	lysine	Lys / K
asparagine	Asn / N	methionine	Met / M
aspartic acid	Asp / D	phenylalanine	Phe / F
cysteine	Cys / C	proline	Pro / P
glutamic acid	Glu / E	serine	Ser / S
glutamine	Gln / Q	threonine	Thr / T
glycine	Gly / G	tryptophan	Trp / W
histidine	His / H	tyrosine	Tyr / Y
isoleucine	Ile / I	valine	Val / V

6. Acknowledgement

An dieser Stelle möchte ich zunächst Prof. Dr. Lutz Schmitt danken. Lutz hat den Glauben an das BSEP Projekt nie verloren und stetig versucht mich zu motivieren und mit mir über das Projekt zu diskutieren. Trotz seines vollen Terminkalenders war er immer ansprechbar und hatte eine offene Tür. Außerdem danke ich Lutz für die gewährte wissenschaftliche Freiheit, die meiner Meinung nach in dieser Form selten zu finden ist. Danke Lutz! Ohne Dich wäre diese erfolgreiche Dissertation so nicht zustande gekommen.

Weiterhin danke ich Prof. Dr. Ralf Kubitz für die feundliche Übernahme des Korreferats, sowie die tolle Kooperation innerhalb des BSEP Projekts und für die erfolgreiche Leitung der klinischen Forschergruppe 217.

„Trinken wir nen Kaffeee? – Ja!" Sander, vielen Dank für all die Jahre (ein gefühltes Drittel davon auf dem Balkon), die Diskussionen und Deinen stetigen Einsatz, aus gefühltem „Nichts" eine Publikation zu machen. Es hat mir sehr viel Spaß bereitet mit Dir und ich danke Dir für Deinen stetigen Einsatz.

PD Dr. Ulrich (Ulli) Schulte danke ich für die tolle Betreuung des Biochemie Studiengangs, die maßgeblich mit dazu beigetragen hat aus ihm das zu machen was er heute ist. Und danke für die vielen kritischen Blicke, wenn ich ein Seminar gehalten habe, sodass ich immer dachte ich erzähle nur Märchen.

Ein ganz besonderer Dank gilt Dr. Jan Stindt. Jan war mein Betreuer, Mentor und „wissenschaftlicher Übervater"! Ohne Dich Jan hätte auch so einiges nicht funktioniert. Obwohl Du mich immer kurz vor den Herzinfarkt getrieben hast, bist Du mir ein guter, treuer Freund geworden und diese Freundschaft will ich nicht missen und hoffe, dass sie noch lange hält! Danke für Deine Hilfe.

André Abts, Dr. Christian Schwarz und Dr. Britta Ries danke ich für eine tolle Zeit. Hilfsbereitschaft, viele Gespräche über den Sinn von Versuchen, Grillen und das eine oder andere Bier.

Meinen weiteren Büropartnern Miroslav Kiro, Dr. Nils Hanekop und Marianne Kluth danke ich für die entspannte und lustige Atmosphäre und die gute Zusammenarbeit. Marianne, halt die Stange für die humanen Leber ABC Transporter hoch! Ich wünsch Dir damit weiterhin viel Erfolg.

Michael Hubert Horst Lenders danke ich für viele Anekdoten aus seinem Leben und dafür, dass er mit mir um sechs Uhr morgen schon im Schwimmbad Bahnen zieht. Viel Spaß beim Lehrer werden!

"Rakesh, how´s it going? – It´s going and going as ususal!" I wish you all the best, you are a great guy!

219

7. Curriculum Vitae

Personal Data

Date of Birth	24th March 1984 in Solingen, Germany
Nationality	German

Work Experience

since Oct 2009	**Ph.D. student** at the Institute of Biochemistry with Prof. Dr. Lutz Schmitt, Heinrich Heine University Düsseldorf
Oct 2008 – Mar 2009	**Research Internship** at the Dartmouth Medical School, Departement of Biochemistry, Lab of Prof. Dr. Dean R. Madden, New Hampshire, USA
Aug – Sep 2007	**Internship** at QIAGEN GmbH in the biological production unit

Academic Education

Oct 2007 – Sep 2009	Master studies in Biochemistry at the Heinrich Heine University Düsseldorf Degree: *Master of Science Biochemistry* (Grade: 1,2)
Oct 2004 – Sep 2007	Bachelor studies in Biochemistry at the Heinrich Heine University Düsseldorf Degree: *Bachelor of Science Biochemistry* (Grade: 1,2)

Civilian Service

Sep 2003 – Jun 2004	**Central laboratory for clinical chemistry** at the municipal hospital Solingen

School Education

Aug 1990 – Jul 2003	Elementary and secondary school in Solingen, Germany Degree: **Abitur** (Grade: 2.8)

Stipends and Awards

Jun 2010	**Master price** of the German Society for Biochemistry and Molecular Biology (GBM) e.V.
Jan 2008	**Stipend of the Dr. Jost Henkel Foundation** for good achievements in the study year 2008/2009

Further Training and Commitments

since Jan 2010   Member of the graduate school iGRAD

since Oct 2009   Member of the integrated graduate school iGK974 of the CRC974 (former CRC575)

Apr 2006 – Sep 2007   Member of the student council for biochemistry

List of Publications

Stindt J[#], **Ellinger P**[#], Stross C, Keitel V, Häussinger D, Smits SH, Kubitz R, Schmitt L. Heterologous overexpression and mutagenesis of the human bile salt export pump (ABCB11) using DREAM (Directed REcombination-Assisted Mutagenesis). **PLoS ONE**. (2011); 6(5):e20562.

Ellinger P, Kluth M, Pryzbylla S, Smits SH, Schmitt L. Structure and function of hepatic ABC-transporters. In: Hepatobiliary Transport in Health and Disease (Eds. Häussinger D, Kubitz R, Keitel V), de Gruyter, Berlin, 23-41.

Cihil K, **Ellinger P**, Fellows A, Beer Stolz D, Madden DR, Swiatecka-Urban A. Dab2 Facilitates AP-2 Independent Recruitment of CFTR to Endocytic Vesicles in Polarized Human Airway Epithelial Cells. **Journal of Biological Chemistry**. (2012); 287(18):15087-99.

Ellinger P, Arslan Z, Wurm R, Tschapek B, MacKenzie C, Pfeffer K, Panjikar S, Wagner R, Schmitt L, **Gohlke** H, Pul Ü, Smits SH. The Crystal Structure of the CRISPR-associated protein Csn2 from *Streptococcus agalactiae*. **Journal of Structural Biology**. (2012); 178(3):350-62.

Ellinger P[#], Kluth M[#], Stindt J, Smits SH, Schmitt L. Detergent Screening and Purification of the Human Liver ABC Transporters BSEP (ABCB11) and MDR3 (ABCB4) expressed in the yeast *Pichia pastoris*. **PLoS ONE**. (2013); 8(4):e60620.

Arslan Z, Wurm R, Brener O, **Ellinger P**, Nagel-Steger L, Oesterhelt F, Schmitt L, Willbold D, Wagner R, Gohlke H, Smits SH, Pul Ü. Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. **Nucleic Acids Research, in press**

Stindt J, **Ellinger P**, Weissenberger K, Dröge C, Herebian D, Mayatepek E, Homey B, Braun S, Horacek M, Canbay A, Schulte am Esch J, Schmitt L, Häussinger D, Kubitz R. A novel mutation within a transmembrane helix of the bile salt export pump (BSEP) with delayed development of liver cirrhosis. **Liver International (under revision)**

[#] equally contributed

8. Statement

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den

_____

Philipp Ellinger