Expression and purification of human connexin 26

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Taras Balandin
from Syrdarya

Düsseldorf, October 2014

from the institute for Physical Biology
at the Heinrich Heine University Düsseldorf

Table of content

Abbreviations

A835	Amphiphilic polymer (amphipol) A-835 (ref.)
B.subtilis	*Bacillus subtilis*
BCIP	5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt
cbb	Coomassie brilliant blue
CDS	coding DNA sequence
CECF	Continuous exchange cell-free expression
CF	Cell-free
CFE	Cell-free expression
CoNTA	NTA in complex with Co^{++} cation
DDM	n-Dodecyl-β-D-Maltopyranoside
DLS	dynamic light scattering
DM	n-Decyl-β-D-Maltopyranoside
DNA	Deoxyribonucleic acid
E.coli	*Escherichia coli*
EK	enterokinase
ER	Endoplasmic reticulum
FC12	n-Dodecylphosphocholine, Fos-Choline-12
FXa	Factor X activated
GJ	Gap junction
GJIC	Gap junctional intercellular communication
hhw	half-height width of a chromatographic peak
IMP	Integral membrane protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LDAO	n-Dodecyl-N,N-Dimethylamine-N-Oxide, Lauryldimethylamine-N-Oxide
mg/L	when used to describe expression yield it refers to liters of cell culture
MISTIC	Initially was suggested as an acronym for "Membrane - Integrating Sequence for Translation of Integral membrane protein Constructs" and was used in all-capitals. Later, it refers to MstX protein, or 'mistic protein' and is used here in all-lower-case as 'mistic'.
MP	Membrane protein
mRNA	Messenger ribonucleic acid
MWST	Molecular weight standard(s)
NBT	nitro-blue tetrazolium chloride
NiNTA	NTA in complex with Ni^{++} cation
NTA	nitrilotriacetic methal chelating group immobilized on sepharose
OG	n-Octyl-β-D-Glucopyranoside
OTG	n-Octyl-β-D-Thioglucopyranoside
PAGE	Polyacrylamide gel electrophoresis
PDS	Potassium dodecyl sulfate
PEG	Polyethylene glycol
pI	isoelectric point
POPC	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate

Srk, srk	Sodium Dodecanoyl Sarcosine, Sarkosyl
TCA	Trichloroacetic acid
TEV	tobacco etch virus
TEVp	TEV protease
tRNA	Transfer ribonucleic acid

Zusammenfassung

In mehrzelligen, lebenden Organismen wie Tieren agieren Zellen in intensiver und dynamischer Kooperation. Diese Kooperation wird in erster Linie durch elektrische und chemische Signale aufrechterhalten. Beide Signale ergänzen einander und sind gleichermaßen erforderlich in komplexen Organismen. Zur Übertragung eines chemischen Signals zwischen entfernten Zellen werden bestimmte Botenmoleküle freigesetzt. Die Oberflächen der Zielzellen tragen entsprechend viele selektive Rezeptoren, welche die Botenmoleküle erkennen und eine Zellreaktion auslösen. Für Zellen, die direkt aneinander angrenzen, gibt es durch diesen direkten Kontakt noch eine weitere Möglichkeit zur Zell-Zell-Kommunikation. In Gewebe oder Organen, wo Zellen als eine Einheit Funktionen erfüllen, ist dieser direkte Kontakt von höchster Wichtigkeit. Im Gegensatz zu Botenstoff-Rezeptor-Signalen werden direkte Kontakte mittels einiger weniger Oberflächenstrukturen auf Zellen implementiert, wobei die Gap Junctions die wahrscheinlich vielseitigste dieser Strukturen ist.

Auf molekularem Level bestehen Gap Junctions aus Connexinen, welche integrale Membranproteine sind, die selbstorganisierend kanalbildende Hexamere bilden. Zwei Hexamere auf der Plasmamembran von aneinander liegenden Zellen sind in der Lage eine enge Bindung einzugehen, in der sich die extrazellulären Teile zu einer isolierten Pore zusammenlagern, die das Cytoplasma beider Zellen verbindet. Dieser dodecamerische Komplex bildet einen Gap-Junction-Kanal. Diese Kanäle neigen dazu große Cluster zu bilden, die benachbarte Zellen miteinander verbinden. Diese Plaques wurden zum ersten Mal vor über 40 Jahren mittels Transmission-Elektronenmikroskopie identifiziert und Gap Junctions (GJ) genannt.

Da sie einen substanziellen Teil des interzellulären Kontakts ausmachen, erfüllen die Gap Junctions elementare und lebensnotwendige Funktionen in multizellulären Organismen. Beschädigte Gap Junctions führen zu abnormaler Organentwicklung, Herzversagen, vaskulären, epidermalen und myelinspezifischen Krankheiten, Myopathie, Taubheit, Katarakt, neurodegenerativen Störungen und Krebs. Eine Entwicklung von Therapien dieser Krankheiten, deren Ansatz Connexine waren, führte bisher zu keinem signifikanten Erfolg.

Die Gap-Junction-Kanäle(GJC) können einen offenen oder geschlossenen Zustand einnehmen und diesen Zustand als Antwort auf ein chemisches oder elektrisches Signal ändern. Die Differenzierung dieser Signale und die Spezifizierung der entsprechenden Antwort sind möglich aufgrund der verschiedenartigen Struktur der Proteine, die die Gap Junctions bilden. Beim Menschen besteht die Gap Junction Proteinfamilie aus 21 Proteinen, die jeweils vier Ausdehnungs-Topologien haben und sich in erster Linie in den Regionen unterscheiden, die dem Cytoplasma zugewandt sind. Die Proteine dieser Familie werden nach ihrem Molekulargewicht benannt, wobei das Kleinste Connexin 23 und das Größte Connexin 62 ist. Während Gap Junctions allgegenwärtig sind, besitzen die einzelnen Connexine bei Säugern eine bestimmte Spezifität in den verschiedenen Organen. Connexin 32 und 43 gehörten zu den ersten, die identifiziert wurden aufgrund ihrer Häufigkeit in Leber und Herz. Später wurde erkannt, dass diese kritisch für Myelin- und Astrozyt-Funktionen sind. Während inzwischen viel über Funktionen von Connexinen bekannt ist, bleiben Fragen bezüglich ihrer Struktur ungeklärt.

Die Herstellung von qualitativ hochwertigen Kristallen integraler Membranproteine von Säugetieren zur 3D Strukturbestimmung bleibt eine herausfordernde Aufgabe. Der Erfolg solcher Kristallisationen hängt maßgeblich von der Qualität des produzierten Proteins ab. Aus natürlichen Quellen isolierte Gap Junctions sind nicht homogen und somit nicht für eine Kristallisation geeignet. Zum Teil liegt das an der Eigenschaft der Connexine gemischte Kanäle zu bilden, wie das Cx32 und Cx26 in Leber-GJs tun. In den frühen 90ern wurde ein Überexpressionssystem in Insektenzellen etabliert, um reines, rekombinantes Connexin 26 und 32 in Milligramm-Mengen zu produzieren. Trotz zweier Dekaden beharrlicher Kristallisationsversuche wurde ein nur bescheidener Erfolg mit Connexin 26, dem kleinsten Protein in der Gruppe, erreicht.

Indessen werden Connexine seit ihrer Entdeckung als sehr vielversprechendes Modell zur Kristallographie integraler Membranproteine (IMP) angesehen. In erster Linie aufgrund von natürlich auftretender Bildung halbkristalliner Gitter in den Gap Junctions Plaques. Die positive Prognose wird noch durch die große Ähnlichkeit von Connexinsequenzen verschiedener Säugertiere unterstützt, was auf eine abgeschlossene evolutionäre Selektion besonders stabiler Proteinfaltungen hindeutet.

Die Widersprüchlichkeit von Erwartungen und Resultaten intensiver Kristallisationsversuche kann sowohl den unvorteilhaften Kristallisationsmethoden, welche hierfür verwendet wurden, als auch der Inhomogenität in chemischer Struktur oder Konformation der Proteinproben zugeschrieben werden. In der Tat weisen Connexine einige Stellen für posttranslationale Modifikationen auf, die den intrazellulären Transport und die Funktionen in den eukaryotischen Wirten regulieren. Aufgrund dessen verursachen die posttranslationalen Modifikationen eine Proteininhomogenität, welche für eine Proteinkristallisation unakzeptabel sind. Eine weitere Behandlung der Proteine zur Beseitigung der posttranslationalen Modifikationen ist erforderlich. Neben ihrer Neigung zur Unvollständigkeit kann die Behandlung eine zusätzliche Minderung der Proteinqualität hervorrufen.

Eine der verbreitetsten Herangehensweisen zur Eliminierung posttranslationaler Modifikationen ist die Etablierung eines Überexpressionssystems in *E.coli*. Den Bakterien fehlen posttranslationale Protein-Modifikations-Systeme, welche typisch für Säugetierzellen sind, und zusätzlich wird ein funktionierendes *E.coli*-Überexpressionssystem mit verschiedenen Vorteilen in Verbindung gebracht, die schlussendlich zu einem Erfolg in der Kristallisation führen könnten. Tatsächlich wurde eine überwiegende Mehrheit der bisher kristallisierten Proteine aus Protein gewonnen, das in Bakterien überexprimiert wurde. Die relativ wenigen Beispiele einer Kristallisation von integralen Membranproteinen aus Protein, welches in Insektenzellen produziert wurde, erforderten enorme Investitionen und Ambitionen zur Vollendung.

Zieht man dieses in Betracht wurden einige Anstrengungen unternommen, um Gap-Junction-Kanäle in *E.coli* zu produzieren. So wurde zum Beispiel Cx 43 mit einem N-terminalem GST-Fusion-Tag produziert, eine Faltung zu funktionalen Kanälen wurde nicht bestätigt. Cx 43 ohne N-terminale Tags wurde *in vitro* hergestellt mittels einer *E.coli* Proteinsynthese Maschinerie; während die Ausbeute der Expression äußerst gering war, konnten funktionale Kanäle gewonnen werden. Cx32 konnte in einem *E.coli* Expressionssystem nicht *in vivo* mit einem N-terminalem Trx-Tag hergestellt werden, obwohl eine in vitro *E.coli* Expression laut TEM korrekt gefaltete Connexinhexamere gebildet hatte. Weiterhin wurde Cx26 in *E.coli* mit einem N-terminalen His-Tag

produziert. Es wurde gezeigt, dass nach Entfernen des His-Tags funktionale Kanäle gebildet wurden. Keine dieser Vorgehensweisen zeigte eine hinreichend hohe Ausbeute an aufgereinigtem Protein. Diese Beispiele zeigen ausführlich, dass eine Produktion von funktionalem Connexin in *E.coli* möglich ist. Jedoch bedarf es weiterer Verbesserungen um für Kristallisation nutzbares Protein zu gewinnen.

Unter den Connexinen ist Cx26 vor allem interessant als Modell zur Etablierung einer Produktion in *E.coli*, da sie für die gesamte Connexinfamilie angewendet werden kann. Das 3D Strukturmodell, das für Cx26 aus Insektenzellen bestimmt wurde, kann mit dem Modell verglichen werden, das für bakterielles Protein ermittelt werden wird und das die Gültigkeit einer Expression in Bakterien bestätigen wird. Die geringe Größe dieses Familienmitglieds verringert die Anzahl der möglichen nachteiligen Strukturmerkmale, die spezifisch für Säugetierproteine sind, und sollte die bakterielle Expression erleichtern.

Somit wurde die Entdeckung neuer Methoden zur Produktion von Connexinen in *E.coli* mit dem Schwerpunkt Cx26 zum Hauptziel der hier präsentierten Arbeit. Wir haben Mistic, ein vor kurzem vorgeschlagenen Fusions-Tag für Membrantargets, eingesetzt, um die Cxs Expression zu erhöhen und haben die Wirksamkeit von ESR als neuen getestet. Der Mistic-Tag zeigte eine ungewöhnliche Resistenz in Bezug auf sein Entfernen, obwohl er ein sehr effizienter Cxs Expressionsbooster ist. Dies hat uns überzeugt intensiv nach Alternativen zu forschen. Eine komplette Genoptimisierung für den Codon bias der *E.colis* hat es möglich gemacht durchschnittlich 0.2 mg reines Cx26 Protein ohne N-terminale Tags aus einem Liter Kultur zu gewinnen. Das Cell free (CF) System ergab eine Ausbeute von 1mg Protein aus 1 mL Reaktionsmix. Weiterhin erlaubte uns das CF System einen originalen Aufreinigungsansatz für Cx 26 ohne Tags zu entwickeln. Das aufgereinigte Protein organisierte sich selbst zu Hexameren nach einer Rekonstitution in unilamellaren Liposomen, die abhängig vom Ca^{++} Level einen offenen oder geschlossenen Zustand eingenommen haben.

1. Summary

In multicellular living organisms, like animals, cells function in an intensive dynamic cooperation. The cooperation is primarily maintained via electrical and chemical signals. Both signalings complement each other and are equally required in complex organisms. For a chemical signals to propagate between distant cells certain messenger molecules are released. There are plenty of selective receptors on the target cells surface to recognize these messengers and mediate cell responses. Yet for cells touching each other an additional way for cell-to-cell communication is possible and provided through direct contacts. These direct contacts become of major importance at a level of tissue or organ when an assembly of cells operates as a whole performing a particular function. Direct contacts, in opposite to messenger-receptor signaling, are implemented via very few type of cell surface structures with gap junctions comprising possibly the most versatile group of them.

Gap junctions at a molecular level are composed of connexins that are integral membrane proteins self-assembling into channel-forming hexamers. Two hexamers located on plasma membranes of apposed cells are able to tightly bind together via their extracellular interfaces and build up thereby an isolated pore between cytoplasms of the cells. This dodecameric complex constitutes a gap junction channel. The channels are prone to cluster into large plaques connecting plasma membranes of adjacent cells. These plaques were identified for the first time over four decades ago by transmission electron microscopy and called gap junctions (GJ).

Making a substantial part in direct intercellular contacts the gap junctions provide the very basic and absolutely indispensable functions for multicellular organisms. Impairing the gap junctions causes abnormal organ development, heart failure, vascular, epidermal, myelin-related diseases, myopathies, deafness, cataract, neurodegenerative disorders, and cancer. No significant success has been achieved so far in the development of therapeutics targeting connexins for the diseases treatment.

The gap junction channels (GJC) can accommodate open or closed state and switch between them in response to as chemical and electrical signals. The differentiation of signals and specificity of responses are possible due to diversified structure of proteins constituting gap junctions. In humans, gap junction protein family is represented by 21

members that share common four membrane span topology and differ primarily in cytoplasm-exposed regions. The family members are designated by their molecular weight, the smallest is connexin 23 and the largest is connexin 62. While gap junctions are ubiquitous in mammals every particular connexin exhibits certain specificity in organ distribution. Connexin 32 and 43 were the ones of the very first identified due to their relative abundance in liver and heart correspondingly. It was found later the ones are critical for myelin and astrocyte functions as well. While much is known about connexins functions their structure determinants remain to be under a question.

Producing of high quality crystals from mammalian integral membrane proteins for the 3D structure determination remains to be a challenging task. The success in crystallization depends substantially on the protein preparation quality. Gap junctions isolated from natural sources are not sufficiently homogeneous and not suitable for crystallization in part due to connexins ability to form mixed channels like Cx32 and Cx26 do in liver GJs. An insect cells overexpression system had been established to produce pure recombinant connexins 26 and 32 in milligram amounts by early 90ths and despite of two decades of persistent 3D crystallization efforts a very minor success has been achieved so far with only connexins 26, the smallest in the group.

Meanwhile, since their discovery, connexins are long considered to be very promising as a model for integral membrane protein (IMP) crystallography mostly because gap junction plaques naturally occur in form of semi-crystalline lattice. The positive prospects are additionally supported by considerable similarity of connexin sequences between mammal species what might indicate accomplished evolutionary selection of rather stable protein folds.

The inconsistence of the expectations and results of extensive crystallization trials might be attributed to as unfavorable crystallization approaches that had been used and the protein samples inhomogeneity in either chemical structure or conformation. Indeed, connexins present a number of sites for post-translational modifications that regulate connexins intracellular trafficking and function in eukaryotic hosts. With that, the post-translational modifications cause the protein inhomogeneity inacceptable in crystallization. An additional protein treatment is required for the post-translational

modifications removal. Besides a propensity for being incomplete, the treatment itself can cause an additional deterioration of the protein preparation quality.

One of the prevalent approaches for elimination of post-translational modifications is establishing the overexpression in *E.coli*. The bacteria lack post-translational protein modification systems typical for mammalian cells and, additionally, a functional *E.coli* overexpression is associated with a number of benefits eventually contributing to a success in crystallization. In fact, a vast majority of the protein crystals reported so far were obtained from proteins overexpressed in bacteria. Relatively few examples of crystallization of integral membrane proteins that were produced in insect cells actually required enormous investments and ambitions for accomplishment.

Taking that into account, several efforts have been undertaken to produce gap junction channels in *E.coli*. Particularly, Cx 43 was produced with an N-terminal GST-fusion tag and its assembling into functional channels was not confirmed. Cx 43 with no N-terminal tags was produced *in vitro* using *E.coli* protein synthesis machinery; while expression yield was extremely low the functional channels have been obtained. Cx 32 failed *in vivo E.coli* expression with N-terminal Trx tag and Cx 32 after *in vitro E.coli* expression was reported to form properly assembled connexin hexamers according to TEM. Cx 26 was produced in *E.coli* with N-terminal His-tag and demonstrated to form functional channels after the His-tag removal. None of the reports mentioned a reasonably high yield of the purified protein. These examples clearly show that functional connexin production is possible in *E.coli* but an additional optimization is required to produce a protein suitable for crystallization.

Among the connexins, Cx 26 is of primarily interest as a target and a model for establishing an *E.coli* production for all the family. The 3D structure model that was determined for the insect cell produced Cx 26 permits a comparison with a structure model that will be determined for the bacterial produced protein and that will validate an approach of bacterial connexin production itself. The minimum size of the family member minimizes number of possibly unfavorable structure features specific for mammalian proteins and should facilitate a bacterial expression.

Thus, exploring the approaches for production of connexins in *E.coli* with emphasis on Cx 26 has become a main focus of the present project. We applied Mistic, the

recently suggested membrane-targeting fusion tag, to boost the Cxs expression. The Mistic tag, while being a very efficient Cxs expression booster, demonstrated a particular resistance to its removal what made us looking for alternatives. Thus, we evaluated bacteriorhodopsin from *Exiguobacterium sibiricum* as a new membrane-targeting expression-enhancing tag. However, the best results have been achieved with cell free (CF) expression system that yielded 1 mg of Cx 26 from 1 mL of reaction mixture. Moreover, CF system permitted us to develop an original purification approach for Cx 26 expressed with no any tags. The purified protein self-assembled into hexamers and after reconstitution into unilamellar liposomes responded to free Ca^{++} level by switching between closed and opened states.

2. Introduction

Intercellular communication and direct contacts

As early as in 1960ths, gap junctions have already received a close attention as particular structures with a distinct morphology in transmission electron microscopy (TEM) images of tissues (1, 2) and primary cell cultures (3). While an average distance between plasma membranes of adjacent cells comprised 16-17 nm a number of regions with 2 nm space between the closely apposed cell membranes were revealed in healthy cells. The proximity made these dense stained regions (Figure 2.1), nexuses, particular and clearly different from structures known as desmosomes which appeared in TEM images of lower quality somewhat similar to the nexuses. Concurrently, it was found that nexuses are rare in carcinoma cells. When cell surface ultrastructures in cervix carcinoma cells were analyzed by TEM to account for their evasion of normal contact growth inhibition a lack of specific connections between adjacent cells that were abundant in the normal cells was observed. Another key structure feature of nexuses that had already been identified at the time was their composition of a numerous identical 10 nm subunits tightly packed into ordered 2D arrays. Later on, the nexus structures of this particular type were more commonly referred to as gap junctions.
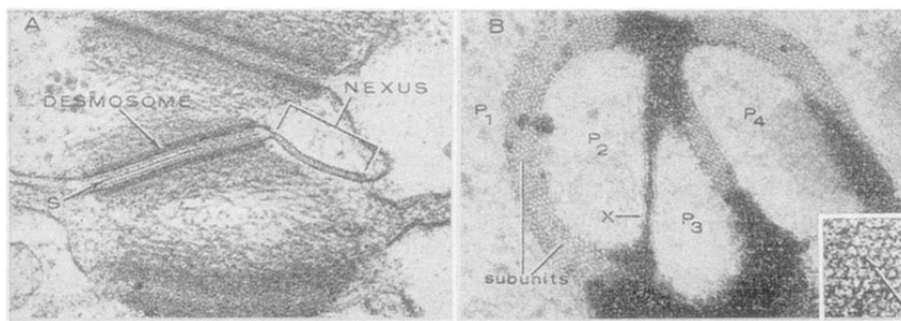
Figure 2.1. Gap junctions in normal cervical epithelium (4). *a*, closely apposed cell membranes are designated as nexus. *b*, a closely packed array of subunits is observed inside the nexus.

By that time an electric coupling between adjacent cells in various epithelium tissues have been investigated (5–12) and much lower resistance was observed between cells being in direct contact comparing to separated cells. However, a malignant epithelium lacked the low-resistance electric coupling. Therefrom an idea of a connection between the intercellular permeability for electrolytes and the number

of junctional direct intercellular contacts emerged. Permeability of the junctional contacts was also demonstrated in free diffusion experiments with relatively large fluorescent tracers (4, 13). Gradually, evidences of the direct implication of gap junctions into direct intercellular communications via various intercellular messengers has grown and supported now by over 10 000 of original reports and 1000 of reviews. Influencing intercellular communications gap junctions are crucial in broad range of biological processes in vertebrates including embryos developing, neuroglia function, cardiac contraction, immunity, oxidative stress resistance, cancer suppression, wound healing, etc. (14).

2.1 Gap junction structure

Distinct lattices observed by electron microscopy in GJ plaques isolated from liver tissues encouraged structure investigation of GJ proteins. Electron diffraction on natural GJ plaques and *in vitro* produced 2D crystals delivered valuable data on gap junction organization.

2.1.1 Levels of GJ organization

Gap junction organization at a molecular level is now well known. Gap junctions are built of gap junction proteins, connexins, ubiquitous in all vertebrates. In humans, connexins comprise a 21-member family of homologous proteins. There are plenty of excellent reviews devoted to the information obtained so far on the structure and function of connexins, so only the very brief description of main connexin features will be given here (14–16).
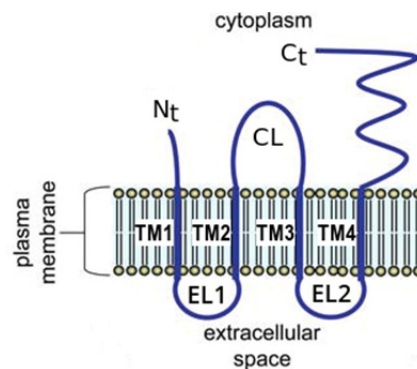
Figure 2.2. Gap junction protein topology (17). Nt, N-terminus; TM1-TM4, transmembrane helices; EL1 and EL2, extracellular loops; CL, cytoplasmic loop; Ct, C-terminus.

All connexins have four rather conservative membrane-spanning alpha-helices, very conservative extracellular loops, variable composition but conservative length N-terminus, variable intracellular loop, and variable C-terminus; the amino acid sequence conservation will be described later. Connexin membrane-spanning topology is represented in Figure 2.2. Specific functions of each structure element will be detailed in sections following further.

Connexins integrate into membranes of endoplasmic reticulum (ER) in a conventional cotranslational manner where they undergo regular quality control and sorting processes. Oligomerization of connexin protomers does not occur under normal condition, i.e. with no overexpression, until protomers are in trans-Golgi network (TGN). All oligomeric assemblies of connexins from dimers to heptamers have been observed, while penta- and heptamers were seldom. Soon after assembling, the hemichannels leave TGN to appear on a cell surface and, with a few exceptions, form full intercellular channel by docking with the opposed hemichannel from the adjacent cell. No monomeric connexins were found in plasma membrane. Ten to thousand full channels cluster into plaques with surface density of 10000 channels per μm^2. The plaques can comprise up to one-half of the total cell surface in lens fiber cells but in most other tissues they take less than 0.1% of the cell surface (18, 19). The described levels of connexin organization are depicted in Figure 2.3.
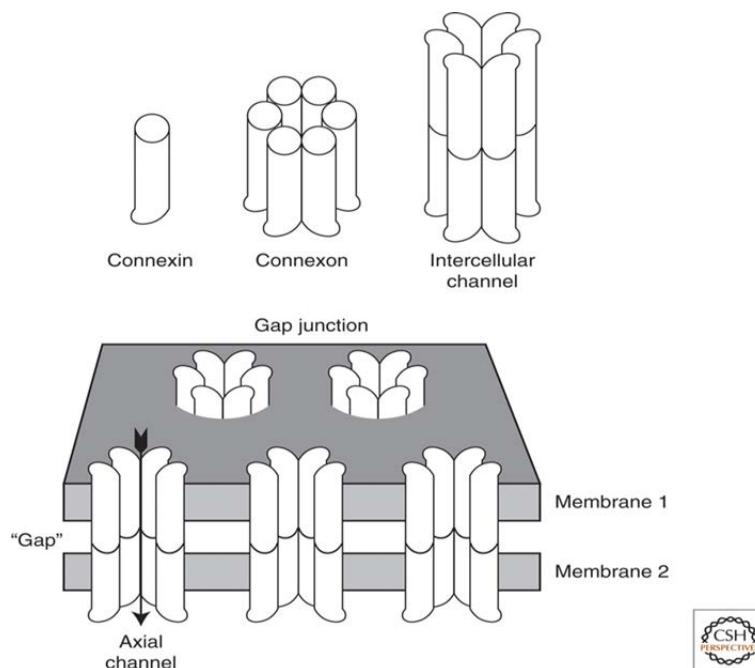
Figure 2.3. Connexin organization levels - gap junction assembly (20).

2.1.2 Connexin nomenclature

Initially, when only a few individual connexins were identified they were designated by their molecular weight, as for example Cx32 and Cx43, the most abundant and best studied gap junction proteins. However, as the family grew this designation rule became inconvenient. Then connexins were divided into five groups A to E, or α to ε, according to their sequence homology. Table 2.1 enumerates all human connexin family members according to modern nomenclature (21). Old nomenclature referring to predicted molecular weight of a protein is still often used.

UPKB_ID	Gene names (primary)	Gene names	Length, aa	MW, kDa
A6NN92	GJE1	GJE1	205	23.8
Q6PEY0	GJB7	GJB7 CX25	223	25.9
P29033	**GJB2**	**GJB2**	**226**	**26**
O95452	GJB6	GJB6	261	30.4
Q9NTQ9	GJB4	GJB4	266	30.4
O75712	GJB3	GJB3 CX31	270	30.8
O95377	GJB5	GJB5	273	31.1
Q8NFK1	GJC3	GJC3 GJE1	279	31.3
P08034	GJB1	GJB1 CX32	283	32
Q8N144	GJD3	GJD3 GJA11 GJC1	294	31.9
Q9UKL4	GJD2	GJD2 GJA9	321	36
P35212	GJA4	GJA4	333	37
P36382	GJA5	GJA5	358	40.4
Q96KN9	GJD4	GJD4 CX40.1	370	40.1
P17302	GJA1	GJA1 GJAL	382	43
P36383	GJC1	GJC1 GJA7	396	45
P48165	GJA8	GJA8	433	48
Q9Y6H8	GJA3	GJA3	435	47.4
Q5T442	GJC2	GJC2 GJA12	439	47
P57773	GJA9	GJA9 GJA10	515	59
Q969M2	GJA10	GJA10 CX62	543	62

Table 2.1. Human connexin family members with corresponding identifiers in UniProt knowledge base (21, 22).

2.1.3 Amino acid sequence conservation

A sequence homology of two largest connexin groups A and B is demonstrated by Figure 2.4. A full sequence alignment for the whole family is given in the appendix A1.

Protein sequence alignment of connexins.

```
                NTH                        TM1            3₁₀ helix        E1                       TM2
         10        20        30        40        50        60        70        80        90
cx26_beta2   ---MDWGTLQTILGGVNKHSTSIGKIWLTVLFIFRIMIVVAAKEVWGDEQADFVCNTLQPGCKNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAY  97
cx32_beta1   ---MNWTGLYTLLSGVNRHSTAIGRVWLSVIFIFRIMVLVVAAESVWGDEKSSFICNTLQPGCNSVCYDQFFPISHVRLWSLQLILVSTPALLVAMHVAH  97
cx31_beta3   ---MDWKTLQALLSGVNKHSTSIGKIWLSVVFVFRVLVYVVAAERVWGDEQKDFDCNTKQPGCTNVCYDNYFPISNIRLWALQLIFVTCPSLLVILHVAY  97
cx30.3_beta4 ---MNWAFLQGLLSGVNKYSTVLSRIWLSVVFIFRVLVYVAAAEEVWGDEQSDFVCNTKQPGCPNVCYDQFFPISHIRLWALQLILVTCPSLLVVMHVAY  97
cx31.1_beta5 ---MNWSIFEGLLSGVNKYSTAFGRIWLSLVFIFRVLVYVLVTAERVWSDDHKDFDCNTRQPGCSNVCFDEFFPVSHVRLWALQLIFVTCPSLLVVMHVAY  97
cx30_beta6   ---MDWGTLHTFIGGVNKHSTSIGKVWITVIFIFRVMILVVAAQEVWGDEQADFVCNTLQPGCKNVCYDHFFPVSHIRLWALQLIFVSTPALLVAMHVAY  97
cx25_beta7   ---MSWMFLRDLLSGVNKYSTGTGWIWLAVVFVFRLLVYMVAAEHVWKDEQKEFECNSFRPGCKNVCFDDFFPISQVRLWALQLIMVSTPSLLVVLHVAY  97
cx43_alpha1  --MGDWSALGKLLDKVQAYSTAGGKVWLSVLFIFRILLLGTAVESAWGDEQSAFRCNTQQPGCENVCYDKSFPISHVRFWVLQIIFVSVPTLLYLAHVFY  98
cx46_alpha3  --MGDWSFLGRLLENAQEHSTVIGKVWLTVLFIFRILVLGAAAEDVWGDEQSDFTCNTQQPGCENVCYDRAFPISHIRFWALQIIFVSTPTLIYLGHVLH  98
cx37_alpha4  --MGDWGFLEKLLDQVQEHSTVVGKIWLTVLFIFRILILGLAGESVWGDEQSDFECNTAQPGCTNVCYDQAFPISHIRYWVLQFLFVSTPTLVYLGHVIY  98
cx40_alpha5  --MGDWSFLGNFLEEVHKHSTVVGKVWLTVLFIFRMLVLGTAAESSWGDEQADFRCDTIQPGCQNVCYDQAFPISHIRYWVLQIIFVSTPSLVYMGHAMH  98
cx50_alpha8  --MGDWSFLGNILEEVNEHSTVVGKVWLTVLFIFRILILGTAAEFVWGDEQSDFVCNTQQPGCENVCYDEAFPISHIRLWVLQIIFVSTPSLMYVGHAVH  98
cx59_alpha9  --MGDWNLLGGILEEVHIHSTMIGKIWLTILFIFRMLVLGVAAEDVWNDEQSGFICNTEQPGCRNVCYDQAFPISLIRYWVLQVIFVSSPSLVYMGHALY  98
cx62_alpha10 --MGDWNLLGGILEEVHSHSTIVGKIWLTILFIFRMLVLRVAAEDVWDDEQSAFACNTRQPGCNNICYDDAFPISLIRFWVLQIIFVSSPSLVYMGHALY  98
cx47_alpha12 MTNMSWSFLTRLLEEIHNHSTFVGKVWLTVLVVFRIVLTAVGGEAIYSDEQAKFTCNDVCYDAFAPLSHVRFWVFQIVVISTPSVMLYGYAVH 100
```

```
                                           CL
         100       110                 120
cx26_beta2   RRHEKKRKFIKGEIKS------------------EFKDIEEIK-TQ-----------------------------------  124
cx32_beta1   QQH-IEKKMLRLEGHG------------------DPLHLEEVK-RH-----------------------------------  123
cx31_beta3   REERERRHRQKHGDQC------------------AKLYDNAGKK-------------------------------------  123
cx30.3_beta4 REERERKHHLKHGPNA------------------PSLYDNLSKK-------------------------------------  123
cx31.1_beta5 REVQEKRRHREAHGENS-----------------GRLYLNPGKK-------------------------------------  123
cx30_beta6   YRHETTRKFRRGEKRN------------------DFKDIEDIK-KQ-----------------------------------  124
cx25_beta7   HEGREKRHRKK-----------------------LYVSPGTM---------------------------------------  116
cx43_alpha1  VMRKEEKLNKKEEELKVAQTDGVNVDMHLKQIEIKKFKYGIEEHG-----------------------------------  143
cx46_alpha3  IVRMEEKKKEREEEQLKRESPSPKE------PPQDNPSSRDDRG-----------------------------------  137
cx37_alpha4  LSRREERLRQKEGELRALPAKDPQVERALAAVERQMAKISVAEDG-----------------------------------  143
cx40_alpha5  TVRMQEKRKLREAERAKEVRGSGSYEYPVA--EKAELSCWEEGNG-----------------------------------  141
cx50_alpha8  YVRMEEKKKSREAEELGQQAGTN--GGPDQ--GSVKKSSGSKGTH-----------------------------------  139
cx59_alpha9  RLRVLEEERQRMKAQLRVELEEVEFEMPRDRRRLEQELCQLEKRK-LN-------------------------------  145
cx62_alpha10 RLRAFEKDRQRKKSHLRAQMENPDLDLE-EQQRIDRELRRLEEQKRIH-------------------------------  145
cx47_alpha12 RLARASEQERRRALRRRPGPRRAPRAHLPPPHAGWPEPADLGEEEPMLGLGEEEEEEETGAAEGAGEEAAEEAGAEEEACTKAVGADGKAAGTPGPTGQHDG 200
```

```
         TM3                          E2                TM4
         130       140       150       160       170       180       190       200       210       220
cx26_beta2   -KVRIEGSLWWTYTSSIFFRVIFEAAFMYVFYVMYDGFSMQRLVKCN-AWPCPNTVDCFVSRPTEKTVFTVFMIAVSGICILLNVTELCYLLIRYCSG--  220
cx32_beta1   -KVHISGTLWWTYVISVVFRLLFEAVFMYVFYLLYPGYAMVRLVKCD-VYPCFNTVDCFVSRPTEKTVFTVFMLAASGICIILNVAEVVYLIIRACAR--  219
cx31_beta3   -----HGGLWWTYLFSLIFKLIIEFLFLYLLHTLWHGFNMPRLVQCANVAPCPNIVDCYIARPTEKKIFTYFMVGASAVCIVLTICELCYLICHRVLRG-  217
cx30.3_beta4 -----RGGLWWTYLLSLIFKAAVDAGFLYIFHRLYKDYDMPRVVACS-VEPCPHTVVISRPTEKKVFTYFMVTTAAICILLNLSEVFYLVGKRCMEI-  216
cx31.1_beta5 -----RGGLWWTYVCSLVFKASVDIAFLYVFHSFYPKYILPPVVKCH-ADPCPNIVDCFISKPSEKNIFTLFMVATAAICILLNLVELIYLVSKRCHEC-  216
cx30_beta6   -KVRIEGSLWWTYTSSIFFRVIFEAAFMYVFYFLYNGYHLPWVLKCG-IDPCPNLVDCFISRPTEKTVFTIFMISASVICMLLNVAELCYLLLKVCFR--  220
cx25_beta7   -----DGGLWYAYLISLIVKTGFEIGFLVLFYKLYDGFSVPYLIKCD-LKPCPNTVDCFISKPTEKTIFILFLVITSCLCIVLNFIELSFLVLKCFIKC-  209
cx43_alpha1  -KVKMRGGLLRTYIISILFKSIFEVAFLLIQWYIY-GFSLSAVYTCK-RDPCPHQVDCFLSRPTEKTIFIIFMLVVSLVSLALNIIELFYVFFKGVKDRV  240
cx46_alpha3  -RVRMAGALLRTYVFNIIFKTLFEVGFIAGQYFLY-GFELKPLYRCD-RWPCPNTVDCFISRPTEKTIFIIFMLAVACASLLNMLEIYHLGWKKLKQGV  234
cx37_alpha4  -RLRIRGALMGTYVASVLCKSVLEAGFLYGQWRLY-GWTMEPVFVCQ-RAPCPYLVDCFVSRPTEKTIFIIFMLVVGLISLVLNLLEVLHLLCRCLSRGM  240
cx40_alpha5  -RIALQGTLLNTYVCSILIRTTMEVGFIVGQYFIY-GIFLTTLHVCR-RSPCPHPVNYVSRPTEKNVFIVFMLAVAAALSLLLSLAELYHLGWKKIRQRF  238
cx50_alpha8  -KFRLEGTLLRTYICHIIFKTLFEVGFIVGHYFLY-GFRILPLYRCS-RWPCPNVVDCFVSRPTEKTIFILFMLSVASVSLFLNVMELGHLGLKGIRSAL  236
cx59_alpha9  -KAPLRGTLLCTYVIHIFTRSVVEVGFMIGQYLLY-GFHLEPLFKCH-GHPCPNIIDCFVSRPTEKTIFLLFMQSIATISLFLNILEIFHLGFKKIKRGL  242
cx62_alpha10 -KVPLKGCLLRTYVLHIILTRSVLEVGFMIGQYILY-GFQMHPLYKCT-QPPCPNAVDCFVSRPTEKTIFMLFMHSIAAISLLLNILEIFHLGIRKIMRTL  242
cx47_alpha12 RRRIQREGLMRVYVAQLVARAAFEVAFLVGQYLLY-GFEVRPFFPCS-RQPCPHVVDCFVSRPTEKTVFLLVMYVVSCLCLLLNLCEMAHLGLGSAQDAV  298
```

```
                    CT
cx26_beta2   ------------------KSKKPV------------------------------------------------------------------------  226
cx32_beta1   ------------------RAQRRSNPPSRKGSGFGHRLSPEYKQN-------------------------------------EINKLLSEQDGSLKDIL---  263
cx31_beta3   ------------------LHKDKPRGGCSPSSSASRASTCRCHH-------------------------------------KLVEAGEVDPDPGN---  257
cx30.3_beta4 ------------------FGPRHRRPRCRECLPDTCPPYVLSQGG-------------------------------------HPEDGNSVLMKAGS---  257
cx31.1_beta5 ------------------LAARKAQAMCTGHHPHGTTSSCKQD-------------------------------------DLLSGDLIFLGSDS---  255
cx30_beta6   ------------------RSKRAQTQKNHPNHAL-----KESKQN-------------------------------------EMNELISD---SGQNAI---  256
cx25_beta7   ------------------CLQKYLK----------KPQVLSV---------------------------------------------------------  223
cx43_alpha1  K------------------GKSDPYHATS-------GALSPAKDCG-------------------------------SQKYAYFNGCSSPTAPLSPM----  281
cx46_alpha3  T------------------SRLGPDASEAPLGTADPPPLPPSSRPP-------------------------------AVAIGFPPYYAHTAAPLGQARAVG  286
cx37_alpha4  R------------------ARQGQDAPPT-------QGTS-SDPYT-------------------------------DQVFFYLP------VGQGPS----  274
cx40_alpha5  V------------------KPRQH-----------MAKCQLSGPS-------------------------------------------VGIVQS----  260
cx50_alpha8  K------------------RPVEQPLGEIPEKSLHSIAVSSIQKAKG-----------------------------YQLLEEEKIVSHYFPLTEVGMVET----  289
cx59_alpha9  WGKYKLKKEHNEFHA--NKAKQNVAKYQSTSANSLKRLPSAPDYN----LLVEKQTHTAVYPSLNSSSVFQPNPDNHSVN-DEKCILDEQETVLSNEI---  333
cx62_alpha10 YKKSSSEGIEDETGPPFHLKKYSVAQQCMICSSLPERISPLQANNQQQVIRNVPKSKTMWQIPQPRQLEVDPSNGKKDWSEKDQHSGQLHVHSPCP---  339
cx47_alpha12 R------------------GRRGPPASAPAPAPRPPPCAFPAAAAG-------------------------------------LACPPDYSLVVRAAERAR---  344
```

20

```
cx26_beta2     ----------------------------------------------------------------------------- 226
cx32_beta1     ---RRSPGTGAGLAEKSDRCSAC------------------------------------------------------ 283
cx31_beta3     ---NKLQASAPNLTPI------------------------------------------------------------- 270
cx30.3_beta4   ---APVDAGGYP----------------------------------------------------------------- 266
cx31.1_beta5   ---HPPLLPDRPRDHVKKTIL-------------------------------------------------------- 273
cx30_beta6     ---TGFPS--------------------------------------------------------------------- 261
cx25_beta7     ----------------------------------------------------------------------------- 223
cx43_alpha1    ---SPPGYK------LVTGDRNNSSCRNYN----KQAS---------------EQNWANYSAEQNRMGQAGSTISNSHAQPFDF--------------- 337
cx46_alpha3    YPGAPPPAADFKMLALTEARGKGQSAKLYNGHHHLLMT----------------EQNWANQAAERQPPALKAYPAASTPAAPSPVGSSSPPLAHEAEAGA 370
cx37_alpha4    ---SPP-------------------CPTYNG---LSSS---------------EQNWANLTTEER--------LASSRPPLF----------------- 308
cx40_alpha5    ---CTPPPD---FNQCLE---NGPGGKFFNPFSNNMAS---------------QQNTDNLVTEQVRGQE--QTPGEGFIQVRYGQKPEVP--------- 324
cx50_alpha8    ---SPLPAKP--FNQFEEKISTGPLGDLSRGYQETLPSYAQVGAQEVEGEGPPAEEGAEPEVGEKKEEAERLTTEEQEKVAVPEGEKVETP-------- 375
cx59_alpha9    ---STLSTSCSHFQHISSNNNKDTHKIFGKELNGNQLMEK------RETEGKDSKRNYYSRGHRSIPGVAIDGENNMRQS-PQTVFSLPAN--CDWKPRW 421
cx62_alpha10   ---WAGSAGNQHLGQQSDHSSFGLQNTMSQSWLGTTTAPRNCPSFAVGTWEQSQDPEPSGEPLTDLHSHCRDSEGSMRESGVWIDRSRPGSRKASFLSRL 436
cx47_alpha12   --AHDQNLANLALQALRDGAAAGDRDRDSSPCVGLPAASR-------------GPPRAGAPASRTGSATSAGTVGEQGRPGTHERPGAKPR-------- 420

cx26_beta2     ----------------------------------------------------------------------------- 226
cx32_beta1     ----------------------------------------------------------------------------- 283
cx31_beta3     ----------------------------------------------------------------------------- 270
cx30.3_beta4   ----------------------------------------------------------------------------- 266
cx31.1_beta5   ----------------------------------------------------------------------------- 273
cx30_beta6     ----------------------------------------------------------------------------- 261
cx25_beta7     ----------------------------------------------------------------------------- 223
cx43_alpha1    -----------------PDDNQNSKKLAAGHELQPLAIVDQRPSS------------RASSRASSRPRPDDLEI----------------------- 382
cx46_alpha3    APLLLDGSGSSLEGSALAGTPEEEEQAVTTAAQMHQPPLPLGDPGRAS------------KASRASSGRARPEDLAI----------------------- 435
cx37_alpha4    ------------------LDPPPQNGQKPPS------------RPSSSASKKQYV----------------------- 333
cx40_alpha5    ------------------NGVSPGHR-------LPHGYHSDKR---------------RLSKASSKARSDDLSV---------------------- 358
cx50_alpha8    ------------------GVDKEGEKEEPQSEKVSKQGLPAEKTPSLCPELTTDDARPLSRLSKASSRARSDDLTV--------------------- 433
cx59_alpha9    LRATWGSSTEHENRGSPPKGNLKGQFRKGTVRTLPPSQGD------SQSLDIPNTADSLG--GLSFEPGLVRTCNNPVCPPNHVVSLTNNLIGRRVPTDL 513
cx62_alpha10   LSEKRHLHSDSGSSGSRNSSCLDFPHWENSPSPLPSVTGHRTSMVRQAALPIMELSQELFHSGCFLFPFFLPGVCMYVCVDREADGGGDYLWRDKIIHSI 536
cx47_alpha12   ------------------------------------------------AGSEKGSASSRDGKTTVWI------------------- 439

cx26_beta2     ------- 226
cx32_beta1     ------- 283
cx31_beta3     ------- 270
cx30.3_beta4   ------- 266
cx31.1_beta5   ------- 273
cx30_beta6     ------- 261
cx25_beta7     ------- 223
cx43_alpha1    ------- 382
cx46_alpha3    ------- 435
cx37_alpha4    ------- 333
cx40_alpha5    ------- 358
cx50_alpha8    ------- 433
cx59_alpha9    QI----- 515
cx62_alpha10   HSVKFNS 543
cx47_alpha12   ------- 439
```

Figure 2.4. A sequence homology of two largest connexin groups A and B, cited from (23). Color coding: red: residues highly conserved among all members; yellow: homologous residues; blue boxes: residues mediating inter-connexon interactions. The secondary structure representation and numbering on the top correspond to Cx26. The cylinders represent α-helices, the small cylinder represents the 3_{10} helix, and the ribbons with arrows represent β-strands.

2.1.4 Role of transmembrane domain
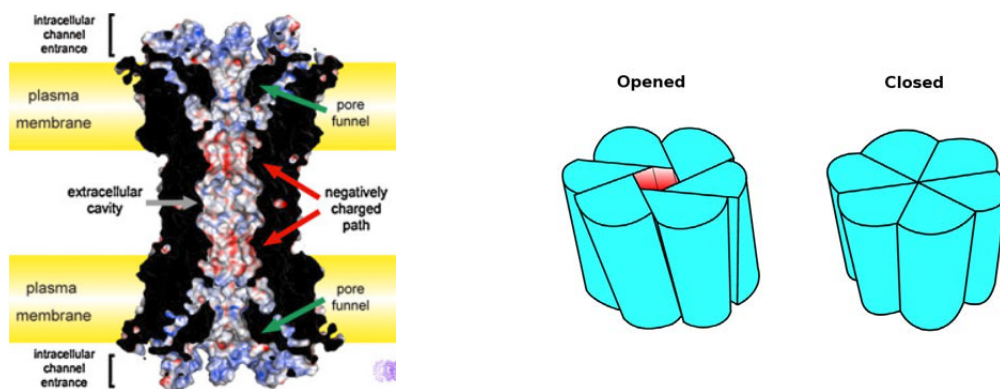
Figure 2.5. Pore architecture of the Cx26 gap junction channel. Left panel, Cx26 gap junction channel is rendered as surface drawing and sectioned along the six-fold axis of symmetry, showing the surface potential distribution of the channel interior (24). Right panel, a model of channel gating mediated by transmembrane domain (25).

Transmembrane domain forms a water-filled channel in cell plasma membrane. The channel is 1.4 nm wide in the opened state and 1.2 nm wide in the closed states respectively (14, 25, 26). Transmembrane domains not only build up the channel but actively mediate gating due to rotation of protomers in response to chemical signal as shown in Figure 2.5 (27).

2.1.5 Role of N-terminal Helix

N-termini of six subunits comprising connexon are able to bind together to form a plug in the channel entrance mediating thereby the trans-junction potential-sensitive gating.
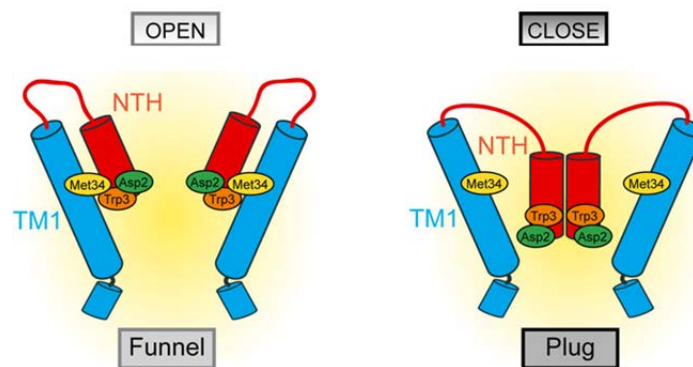
Figure 2.6. Schematic representation of hypothesized plug gating mechanism of gap junctions (24).

Interaction of conservative Trp3 with Met34 explains the most common form of congenital deafness caused by the most frequent natural mutation M34T in connexin 26. Modulating of the channel conductance and the gating polarity by substitutions and deletion of Asp2 confirms its contribution in sensing changes of the electric field along the pore (25, 28).

2.1.6 Role of extracellular loops

Extracellular loops are the structure elements that make connexins particular and different from other channels – they form a rigid tightly isolated tunnel in the extracellular space between adjacent cells (Figures 2.7 and 2.5). Besides the architecture function they contribute to the channel selectivity via the nature of the amino acid side chains lining the pore interior.
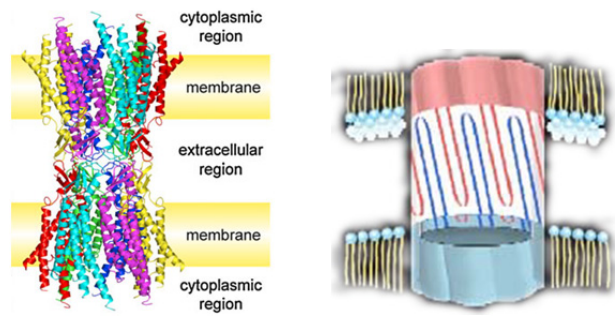
Figure 2.7. Cartoon representation of the atomic model of the connexin 26 gap junctional channel (left), adapted from (24) and schematic representation of the docking interface between hemichannels made up by connexin extracellular loops (right), adapted from (19).

The important structure feature of the extracellular loops is disulfide bridges formed by six conserved cysteines, two bridges per loop and one between the loops. All six cysteines are essential for the proper channel folding and gap junction assembling, since mutation in any of them lead to the loss of channel function (29).

2.1.7 Role of Cytoplasmic loop and C-terminal domain

The cytoplasmic loop is implicated into GJ channel gating via interaction with the C-terminal domain. The interaction influences an occlusion of the channel entrance. The unambiguous description of the interaction remains complicated by loose structure of C-terminal domain (30). However, the definite conformation changes have been detected in the C-terminal domain in response to changes in pH and Ca^{++} concentration (31).
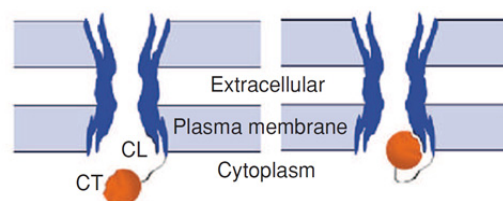
Figure 2.8. Gap junction channel gating by interaction of cytoplasmic loop (CL) and C-terminal domain (CT), adapted from (14).

C-terminal domain also exposes a number of sites for posttranslational modification regulating the connexin intracellular trafficking and contributing to the channel gating as well (32, 33). Short regions responsible for the interactions of connexins with intracellular partners, like zonula occludin and tubulin have been identified as well (14).

2.1.8 Electron microscopy

Transmission electron microscopy for a long time was the primary technic for the determination of the gap junction protein structure. Due to the large amount of collected data and very modest requirement for the sample amount, TEM remains the method of choice to control the quality of connexin preparations in terms of their structural integrity and size-homogeneity prior to their detailed biophysical studies. In the properly prepared samples and properly adjusted radiation condition it is possible to distinguish a solvent-filled pore within the particles with approximate size of 8 nm. This "doughnut" appearance (Figure 2.9, *a*) is considered as a hallmark of the correct channel assembly (19, 34–45). Microphotograph on Figure 2.9, *b* with the same sample demonstrates that lack of typical doughnut appearance does not necessarily suggest incorrect protein folding but can be the analysis artifact.

a.   *b.*
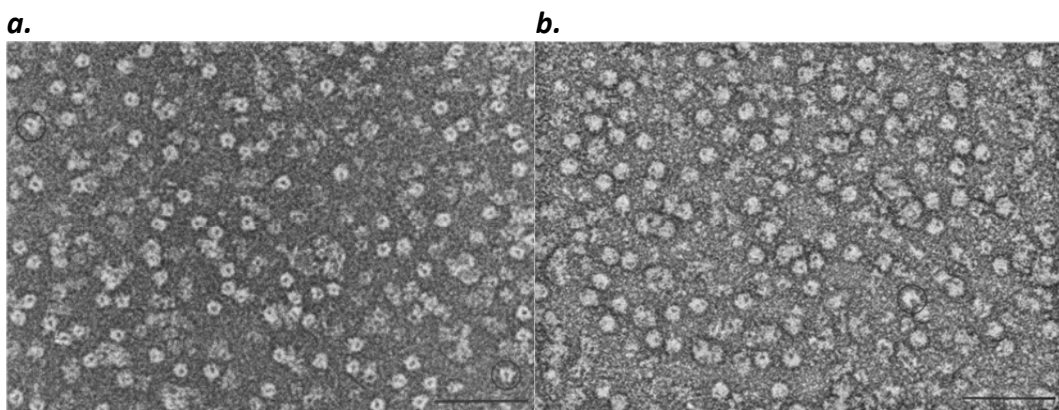
Figure 2.9. Purified recombinant connexons negatively stained with uranyl acetate, adapted from (46). (a) Connexons oriented predominantly with the channel axis perpendicular to the support film, resulting in a doughnut-shaped appearance. (b) Connexons in a thicker stain which are predominantly tilted so that the stain-filled indentation (corresponding to the extracellular entrance of the channel) is no longer in the center of the particle. Some edge-on views (circled) are present in both figures. Bars, 50 nm. Adapted from (46).

3. Results and discussion

Membrane targeting in *E.coli* expression

Native Cx26 coding sequence does not result in a reasonable expression in *E.coli*. A number of particular N-terminal extensions were suggested as general drivers to meliorate poor expression. An essential role in membrane proteins folding plays their insertion into lipid membrane. MP integration sequesters the Sec system that is limited and nearly just sufficient for live essential 'household' proteins. Sec depletion causes toxicity to host. Sec is not always necessary; some MPs intrinsically shows propensity to self-integration into a cell membrane. These MPs can serve as membrane-integrating drivers for other proteins of interest.

A few dedicated expression tags facilitating membrane insertion of the integral membrane protein have been reported so far with a protein "*mistic*" from *B.subtilis* being the best characterized (47). Mistic (48) is a 13 kDa four alpha helixes protein enriched with lipophilic and negatively charged amino acids. It binds to phospholipid cell membrane probably via its numerous surface-exposed carboxylic groups and integrates into the membrane via hydrophobic amino acid side chains, independent on host cell membrane translocation machinery. Both, N- and C- termini of the protein are loose and available for protein conjugation by gene fusion presumably without impairing the protein core structure. Mistic can readily be overexpressed in *E.coli* at levels above 20 mg/L and served as an expression and lipid-membrane associating driver for dozens of integral membrane proteins.

3.1 Mistic tag-driven connexin overexpression.

We have explored a potential of the described above protein, mistic, for connexin overexpression in *E.coli*. It has been shown (49–55) that mistic can assist in membrane association of the cargo protein and only in the case of N-terminal fusion when mistic precedes the protein of interest in a fusion construct. Meanwhile, the sensitivity of the connexin N-terminus to modifications, its impact on the channel assembling and function were outlined in the section 2.1.5. Therefore a linkage between mistic and connexin should be considered with care in the fusion protein design. In the following sections we will focus on the linker features.

3.1.1 Mistic tag boosts connexin expression in *E.coli*

3.1.1.1 Initial constructs design

As mentioned above, the length of the mistic-to-connexin linker can be critical. Roosild et al. mentioned (49) the use of linkers made of up to 35 amino acids for different targets. An attempt to find an optimal linker length was undertaken in (47) for the truncated potassium channel Kav1.1 from *A.californica* as a target protein.

In accordance with the comparison made by Roosild (49) on the example of the mistic-lnk$_x$-Kav1.1 fusion proteins we have initially tested linkers of 5, 7, and 8 amino acids length. Apparently, not only the length but also a composition of the linkage might be of importance. We started with general linkers composed of glycine repeats endowing the flexibility and serine providing hydrophilicity that are listed below with their designations:

Designation	sequence
L5	GSGGS
L5p	GPGGS
L7	GSGSGGS
L8pp	GPGSSGPG

Following the linker we placed a protease recognition site to cleave mistic out of the fused protein. We have chosen enteropeptidase from bovine intestine (EC 3.4.21.9) for the high cleavage specificity and tolerance for the amino acid at P' position following attacked peptide bond. This would provide an opportunity to avoid any extra amino acids at connexin N-terminus after cleavage. To raise a chance for these relatively short linkers to circumvent possible conformational constrains and adopt a favorable conformation to expose the protease recognition site for better accessibility we added proline-flanked linkers in the listed series as well.

Hexahistidine tag preceded the mistic N-terminus to make possible an obtaining of connexin free of any tag after a metal-affinity purification and cleaving mistic out.

The complete construct coding sequences were placed under the transcriptional control of promoter phi10 of the major capsid protein 10A of phage T7 (56) combined with *lac* operator from *E.coli* lactose operon provided by pSCodon1.2 expression vector (Eurogentec S.A.). To compensate for the exogenous target protein coding DNA sequence codon bias the vector encodes also tRNAs that are rarely presented in *E.coli*

comparing to mammalian cells. As a particular feature, the chosen vector implements the plasmid stabilization system developed by Szpirer C. as described in (57). This separate component stabilization system is based on toxin gene *ccdB* and its deactivator *ccdA*, both derived from F-plasmid *ccd* operon. The *ccdB* gene was incorporated into the genome of BL21(DE3) cells resulting in 'suicidal' SE1 cells. The CcdB protein binds DNA gyrase GyrA and arrests host cells propagation. The pSCodon plasmid carries gene for CcdA protein abrogating CcdB toxicity in cells carrying the plasmid. The system allows the positive selection in antibiotic-free medium of host cells bearing the plasmid. Although, the plasmid features also beta-lactamase gene and can therefore be used with regular non-suicidal hosts as well.

The first series of constructs was designated as listed below and can be schematically depicted as in Figure 3.1.1.1.

Titles	Short titles
pSCH6M110L5EKhCx26	HML5Cx26
pSCH6M110L5pEKhCx26	HML5pCx26
pSCH6M110L7EKhCx26	HML7Cx26
pSCH6M110L8ppEKhCx26	HML8ppCx26
pSChCx26	Cx26

Here, the last construct represents 'negative' control, connexin 26_{M34A} CDS including stop codon and having no any tags, inserted into the same pSCodon expression vector.
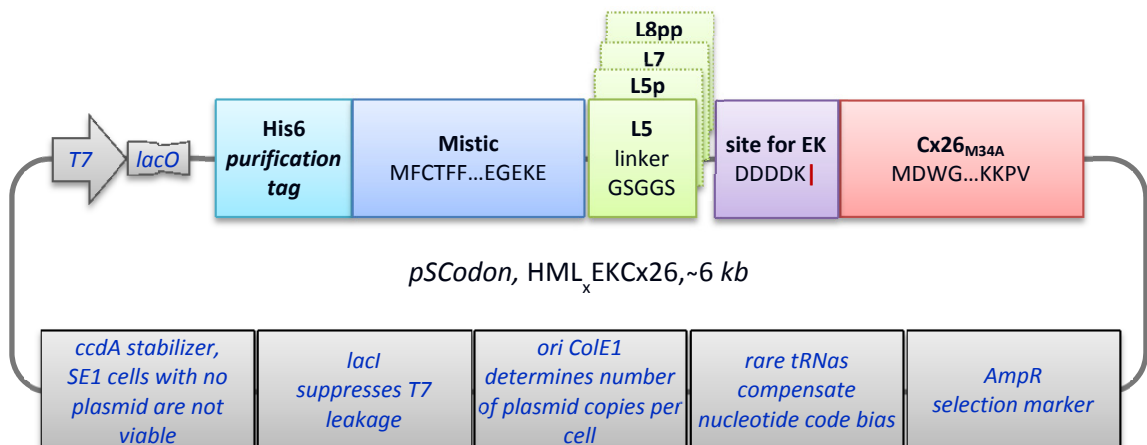
Figure 3.1.1.1 Scheme of connexin 26 expression plasmids representing general design of the initial constructs.

3.1.1.2 Mistic tag boosts connexin 26 expression in E.coli

The *E.coli* cells of strain SE1 were transformed with the plasmids described above. Connexin expression was induced by lactose in autoinduction media according to (58) at 37°C for 15 hours. Total cell lysates were analyzed for Cx26 level by western blotting with specific rabbit polyclonal antibodies, N19, raised against a synthetic peptide mapping at the N-terminus of human connexin 26. After staining with enhanced combined NBT-BCIP precipitating chromogenic substrate for alkaline phosphatase, the nitrocellulose membrane was stained with ponceau S to reveal overall electrotransfer efficiency and loading equivalence by unspecific staining of total proteins.
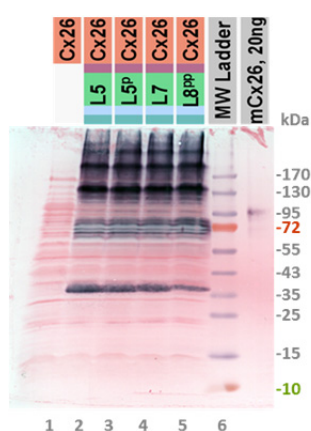
Figure 3.1.1.2 Western-blot analysis of the expression level of connexin 26 produced in *E.coli* in fusion with mistic (lanes 2-5) or with no any tag (lane 1). Lane 6: molecular weight standards; lane 7: positive immunostaining control, a commercial standard of natural connexin 26 purified from mouse liver. Sample lanes were loaded with equal amount of lysed cells estimated by optical density of cultures. Label color code: turquoise, His-tag; blue, mistic protein; green, linkers; purple, protease recognition sites; salmon, connexin 26; grey, reference proteins.

The expected molecular weight for the H6-mistic-Lnk_x-EK-Cx26 fusion constructs is 41 kDa. It should be mentioned also that functional human connexin 26 purified from insect cells as hexameric hemichannel demonstrates relatively elevated electrophoretic mobility in discontinuous SDS-PAGE and shows up after coomassie brilliant blue staining as dominating monomeric band with an apparent molecular weight of ~20 kDa accompanied by the faint dimeric band of ~40 kDa (59). However, after specific immunostaining, the dimeric band became more prominent while remained being minor.

Here, Figure 3.1.1.2 demonstrates as N-terminal mistic tag tremendously boosted the connexin 26 expression level in *E.coli*. Strong immunopositive signals are observed

for the bands of apparent molecular weight of ~37 kDa corresponding to the expected mobility of the monomeric form of the fusion constructs if the abovementioned elevated mobility of Cx26 is taken into account. Remarkably, bands corresponding to the fusion construct oligomers are as well developed as monomeric ones. The most prominent oligomer bands correspond probably to tetramers and hexamers according to the apparent molecular weight. An analysis of bands corresponding to dimers is hampered by the overlapping with host cell proteins. A calculation of band specific staining intensity made with ImageJ software package (60) relative to the positive control of mouse connexin 26 gives the expression level of HML5Cx26 construct of 4 mg per 1L of bacterial culture considering monomeric species only. When all oligomeric forms are taken into account the cumulative expression yield can be estimated as high as 12 mg per 1L of bacterial culture. The HML5Cx26, HML5pCx26, and HML7Cx26 constructs are not notably different in the expression level. The only difference between linkers under the test can be observed in the level of monomeric form of HML8ppCx26 that is half the monomeric form of HML5Cx26.

For the control construct without mistic the expression level is below the immunodetection limit what complicates an evaluation of an expression gain factor caused by mistic fusion. To make a rough estimation we can consider 10 ng as a detection limit taking into account the staining intensity of used positive control. We can assume also that the expression level of connexin 26 control construct without mistic tag does not exceed the detection limit and hence is below 0.27 mg per 1L of bacterial culture. Therefore, we can conclude that mistic tag can provide minimum 40-fold gain in expression level of connexin 26 in *E.coli*.

It is also worth to note that N-terminal connexin extension did not block an antibody binding to the connexin N-terminus suggesting its accessibility under western blotting conditions.

Although being encouraging, the observed oligomerization should be considered with care. Indeed, incomplete disruption of integral membrane protein native oligomers under SDS-PAGE conditions have already been described (61). It is also true for connexins 32 and 43 (35, 45, 62–64). The fraction and an order of oligomers resisted in regular SDS-PAGE analysis varied for connexin 26 from case to case (38, 39,

59, 65–67) . The SDS resistance of connexins was also used for their purification (38, 68–70). At the same time, mistic protein itself is also known to form oligomers in solution (71) that are persisted in part under SDS-PAGE conditions (71). It should be mentioned also that in discontinuous SDS-PAGE system of Laemmli (72) a protein concentrating factor at the border of stacking and separating gels might reach two orders of magnitude that together with sub-CMC concentration of SDS and consequent jump in local protein to detergent concentration ratio can readily induce prompt oligomerization and, moreover, aggregation of highly hydrophobic integral membrane proteins. Thus, the observed oligomerization of the expressed connexin fusion proteins might represent their original native state inside cells but it might be an analysis artifact as well.

3.1.1.3 Mistic targeting: subcellular fractionation

Besides increasing an expression level, the mistic tag was applied to target fused connexin to the cell lipid membrane and to provide thereby a semi-native environment where connexin could fold correctly. To test the targeting efficiency we subjected cells expressing fused proteins to subcellular fractionation by differential centrifugation. This approach is not as effective in cell fractions separation as density gradient centrifugation but allows higher throughput and better suits for the preliminary analysis of multiple samples. Since we used different length linkers between mistic and connexin we could expect a difference in the effect of mistic on connexin localization in cells.

The constructs listed in section 3.1.1.1 were produced in 50-mL scale cultures as described in section 3.1.1.2. Bacteria were harvested, treated with lysozyme, DNase, and disrupted in CAPS buffer at pH 10.5 as recommended in (23) by mechanical stress applied with a French Press. Unopened cells were removed by 10 min centrifugation at 1000 g and the supernatant was centrifuged at 8000 g for 30 min to precipitate heavier cell fragments, and then at 100000 g for 1 hour to harvest cell membranes. The pellets were solubilized in sarkosyl with 6M urea, clarified by centrifugation at 100000 g, and purified further by Co^{++}-affinity chromatography using NTA-sepharose.

Figure 3.1.1.3 demonstrates the distribution of the connexin fused protein between the fractions obtained.
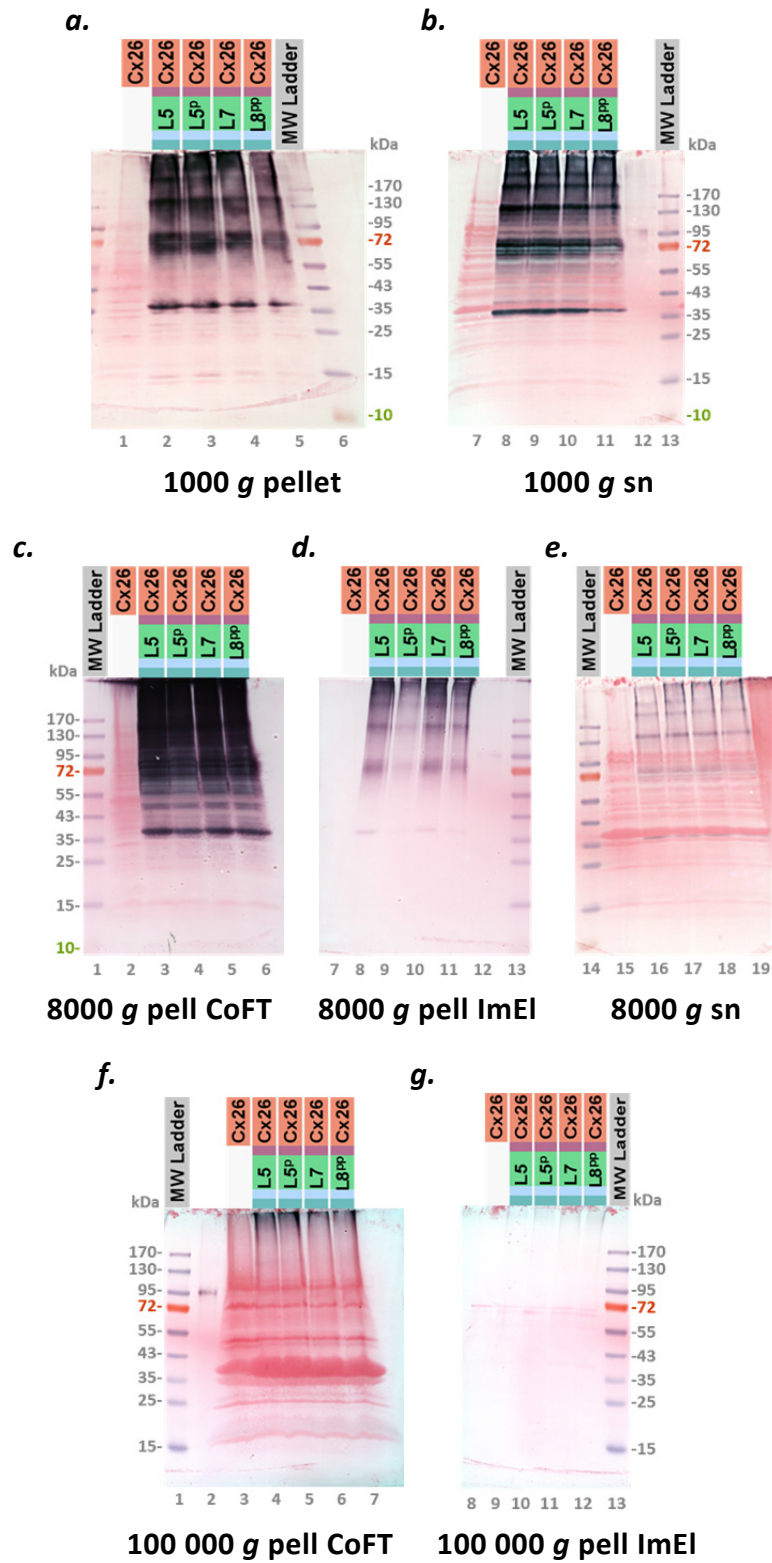
Figure 3.1.1.3 Subcellular fractionation by differential centrifugation of *E.coli* expressed connexin 26. *a*, 1000 *g* pellet; *b*, 1000 *g* supernatant; *c*, CoNTA non-binding flow-through fraction of solubilized 8000 *g* pellet; *d*, CoNTA-binding imidazole-eluted fraction of solubilized 8000 *g* pellet; *e*,

8000 *g* supernatant; *f*, CoNTA non-binding flow-through fraction of solubilized 100000 *g* pellet; *g*, CoNTA-binding imidazole-eluted fraction of solubilized 100000 *g* pellet.

As can be seen from panes *a* and *b*, a considerable fraction of the expressed connexin precipitated from the crude cell lysate at low speed centrifugation. The 1000 *g* pellet did not contain typical intact cells layer but resembled inclusion bodies. A formation of inclusion bodies could be anticipated for the fusion constructs under the test due to their distinct dipolar nature. Mistic, being extremely enriched with aspartate and glutamate residues, has pI of 4.5 and pI of connexin 26 is 9.1. In case of misfolding such a charge distribution can induce a nonproductive aggregation followed by inclusion bodies formation especially with the effect of high hydrophobicity of both proteins. A high biosynthesis rate at 37°C could also contribute to the phenomena.

Approximately half the produced target protein remained in 1000 *g* supernatant yet. Precipitation at 6000 *g* to 10000 *g* is recommended for removal of large pieces of cell debris prior to isolation of purified cell membranes (52, 73, 74). As panes *c* to *e* demonstrate, a little of connexin left in the 8000 *g* supernatant that should have represented cell lipid membranes together with cytoplasmic soluble proteins. Instead, most of the connexin that remained in the 1000 *g* supernatant precipitated at this step together with a lipid-like pellet. The 8000 *g* pellets were almost completely soluble in sarkosyl, a detergent that is used for selective solubilization of *E.coli* inner membrane (75–78).

The bacterial lipid membrane fraction isolated by precipitation at 100000 *g* still represented a typical pattern of major membrane proteins(73, 74) as confirmed by unspecific ponceau S staining shown in pane *f* that validate lipid membrane fraction isolation procedure.

Although the simple experiment described above did not clarified a definite distribution of connexin over cellular fractions, it permitted to observe that 1) none of the expressed fusion protein remained in a soluble form similar to cytoplasmic proteins, 2) its significant fraction probably formed inclusion bodies, 3) its overexpression altered sedimentation behavior of *E.coli* membranes, 4) the protein solubilized from isolated membranes did not bind CoNTA sepharose reasonably well.

A lipid bilayer insertion expected from the association of the expressed integral membrane protein of interest with the cell lipid membrane fraction is considered as a prerequisite for the correct folding of this protein and can suggest possible feasible purification and further crystallization. Another easy to test and generally positive sign for the favorable membrane insertion is amenability to solubilization in mild detergents.

3.1.1.4 Detergent solubility screening

To screen a detergent solubility we expressed HML8^{pp}Cx26 as described in section 3.1.1.2 harvested and disrupted bacteria as in section 3.1.1.3 and isolated by 100 000 *g* centrifugation an insoluble cell fraction including total lipid membranes and inclusion bodies in case they were present. We tested a detergent-solubility of the insoluble fraction in a small set of the best performing detergents following the summary of the membrane protein crystallization experience (79, 80). Sarkosyl was included in the test for the comparison with the previously obtained results. To discriminate a possible presence of inclusion bodies we set up the alternative solubilization in the presence of 6M urea as well. Detergent insoluble material was removed by centrifugation at 100 000 *g* and supernatants were analyzed for connexin 26 presence by western blotting. The results are presented in figure 3.1.1.4.
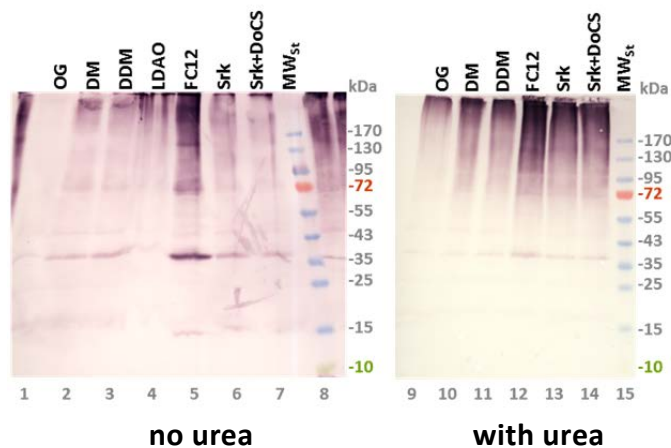
Figure 3.1.1.4 Solubilization of *E.coli* produced connexin 26 in selected detergents. Insoluble fraction of *E.coli* produced connexin 26 was extracted with indicated detergents in the presence of 6M urea (right pane) and without urea (left pane). Supernatants of clarified solubilizates were analyzed by western-blotting.

The very mild maltoside detergents with no urea were insufficient to solubilize connexin 26 fused with mistic, although DDM was shown to be the best of dozens detergents tested for solubilization and stabilization of connexin 26 produced in insect cells (81). Urea did raise a fraction of solubilized connexin for the most of detergents but for the price of the increased aggregation and diminishing of distinct monomeric band of the fused protein. Meanwhile, connexons within gap junctions isolated from liver are known to resist 6M urea (68). The observed effect of urea might indicate insufficient protein protection by lipids and thereby its incomplete membrane integration and/or consequently partial protein unfolding. This does not reveal clearly any presence of regular inclusion bodies though; that would require expression up-scaling and gradient density centrifugation analysis but the CoNTA binding inefficiency mentioned in the conclusion of section 3.1.1.2 was of higher priority to address. In the case if a urea-assisted approach had been chosen for the further preparative purification the aggregation would be insuperable and would completely nullify the yield of a properly folded protein at the end. Therefore, urea was considered as exerting negative effect and to be avoided further.

Zwitterionic *n*-dodecylphosphocholine (FC12) showed up here as the best candidate for the further solubilization and purification. This detergent is often referred to as very powerful in integral membrane protein solubilization (82) while preserving protein functionality, and the power is attributed to the fact that its chemical structure closely resembles structure of phospholipids building up cell membranes (83). Albeit, examples of its denaturing effect also should not be ignored (84).

3.1.1.5 CoNTA binding in FC12 is not satisfying still

Taking into account inefficient CoNTA-sepharose binding of the constructs previously purified in sarkosyl we tested a purification efficiency of HML8^{pp}Cx26 solubilized in FC12 prior to up-scaling. Figure 3.1.1.5 demonstrates that as little as approximately 10% of the solubilized protein bound the CoNTA-sepharose. The poor binding can be explained by insufficient exposition of hexahistidine tag (85). In our constructs the hexahistidine tag is followed by mistic protein immediately with no any spacer amino acids as it was originally designed by Noirclerc-Savoye (86) in the vector

pLIM14 that was suggested for high throughput screenings of mistic-tagged constructs and used as a donor of hexahistidine tagged mistic in our constructs. As the mistic N-terminus is not involved into maintaining of the protein tertiary structure (49) it seemed being sufficient itself as a flexible spacer arm to expose a histidine tag. At the same time, the mistic N-terminus is enriched with aromatic residues that can interact with some of the preceding histidines and prevent them from entering into metal ion coordination sphere directly or by reducing an effective number of those that are still able to interact with an immobilized metal ion.
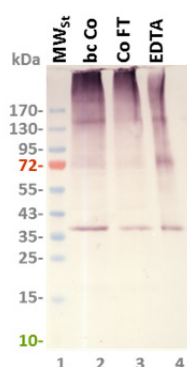
Figure 3.1.1.5 Western-blot analyses of CoNTA binding efficiency in FC12. Lanes 2 and 3 were loaded with 1/100 of total fraction applied onto a CoNTA, lane 4 was loaded with 1/10 of total eluted fraction. Lane 2, insoluble cell fractions solubilized in FC12, clarified, before loading on CoNTA. Lane 3, CoNTA non-binding flow-through fraction. Lane 4, CoNTA eluate.

In contrast with Noirclerc-Savoye, Roosild (49) surrounded histidine tag with serine repeats aside from introducing the thrombin recognition site between histidine tag and mistic; both could favor a better exposition of histidine tag into solution.

3.1.1.6 Exposing His tag for better CoNTA binding

In a systematic comparison of NiNTA binding efficiency of short peptides of different length and composition that were enriched with histidines it was established that a single stretch of five histidine repeats binds immobilized Ni^{++} nearly as good as six histidines (87–89). Simple raising the number of histidines in the tag that is often recommended to circumvent poor binding (90) might not be sufficient to improve the histidine tag exposition as the local environment in mistic N-terminus vicinity would provide sufficient hydrophobic amino acid side chains for the added histidines to

interact with and remain withdrawn from the desired interaction with immobilized metal ions. Therefore we have introduced a short hydrophilic linker between His tag and mistic as depicted at Fig. 3.1.1.6. The generated constructs were designated HgssML7Cx26 and HgssML8^{pp}Cx26, respectively.

Figure 3.1.1.6 Constructs with His tag exposed for better binding, only CDS is depicted, vector body remained the same as in Figure 3.1.1.1 and is omitted here.

3.1.1.7 Exposed His tag demonstrates better binding

The fused proteins HML7Cx26 and HgssML7Cx26 were expressed as described in section 3.1.1.2. The total insoluble cell fraction was isolated as described in section 3.1.1.4 and expressed proteins were solubilized in FC12 at pH 8.0 and purified on CoNTA-sepharose. A comparison of the proteins distribution between the fractions obtained in the course of purification is given in figure 3.1.1.7.
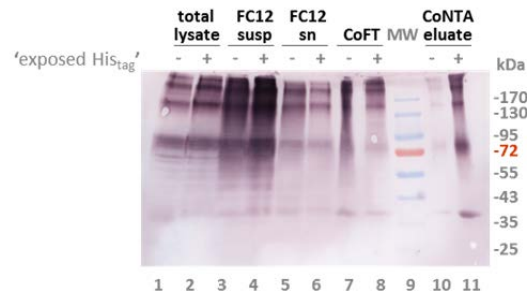
Figure 3.1.1.7 Western-blot analyses of CoNTA binding efficiency in FC12. Constructs with and with no spacer between His$_6$tag and mistic were compared. Samples were loaded in equivalent amounts excepting lanes 1 and 2 that were 3-fold diluted. Lanes 3 and 4, insoluble cell fractions solubilized in FC12, non-clarified. Lanes 5 and 6, insoluble cell fractions solubilized in FC12, clarified, before loading on CoNTA. Lanes 7 and 8, CoNTA non-binding flow-through fractions. Lanes 10 and 11, CoNTA eluates.

Improving the histidine tag position relative to the original one (86) allowed higher fraction of the solubilized protein to be eluted from the CoNTA column (compare lanes 10 and 11). Yet it is the solubilization what showed up here as the next critical step limiting the protein purification yield (compare lanes 3 and 4 with 5 and 6). The

solubilization efficiency could depend significantly on the protein folding state and on its local environment. Both can correlate with the subcellular localization of the protein. Inner and outer membranes of *E.coli* are known to have different lipid composition (91–94) that can influence as protein folding and solubility of a lipid membrane in detergent solution required for the protein solubilization. We attempted to accomplish the analysis of the protein distribution over the cell fractions that we have started in section 3.1.1.3 using an alternative approach.

3.1.1.8 Subcellular fractionation by isopycnic centrifugation

HgssML7Cx26 was expressed as in section 3.1.1.2. Bacteria were harvested, treated with lysozyme, DNase, and disrupted in TrisHCl buffer at pH 8.0 by mechanical stress applied with a French Press. The obtained total lysate was layered on a 20% glycerol layer that was underlaid with 80% glycerol cushion in ultracentrifuge tubes. After the 100 000 *g* overnight centrifugation insoluble cell components settled under 20% glycerol layer and soluble ones remained above 20% glycerol. The supernatant and the most of separating 20% glycerol layer were discarded. The insoluble cell fraction was resuspended in a buffer to dilute residual glycerol approximately below 30% and layered over a 'separating' series of glycerol layers, 40% to 100% in 20% steps, in ultracentrifuge tubes. After 100000 *g* centrifugation insoluble cell components distributed between 40% and 80% glycerol layers. Obtained layers were collected into fractions 1 to 6 by their appearance and according to (95, 96): lower-density brown fractions were considered as enriched with inner cell membranes and higher-density turbid white fractions were collected as outer cell membranes, possibly contaminated with inclusion bodies. No connexin was detected in a soluble protein fraction (< 20% glycerol), in consistence with our intention to target connexin into membranes. Additional washing step that have been done by dilution of the membranes harvested in 80% glycerol cushion just after cell lysis turned out to be useful in removal of contaminant proteins loosely bound with cell membranes as can be concluded from fractions F1 and F2 composition.

As follows from fractions F3 to F6 composition all the produced connexin distributed between 40% and 80% glycerol layers corresponding to inner and outer

membrane fractions respectively. And major amount of the protein is localized within outer membrane fraction if each fraction volume is taken into account.
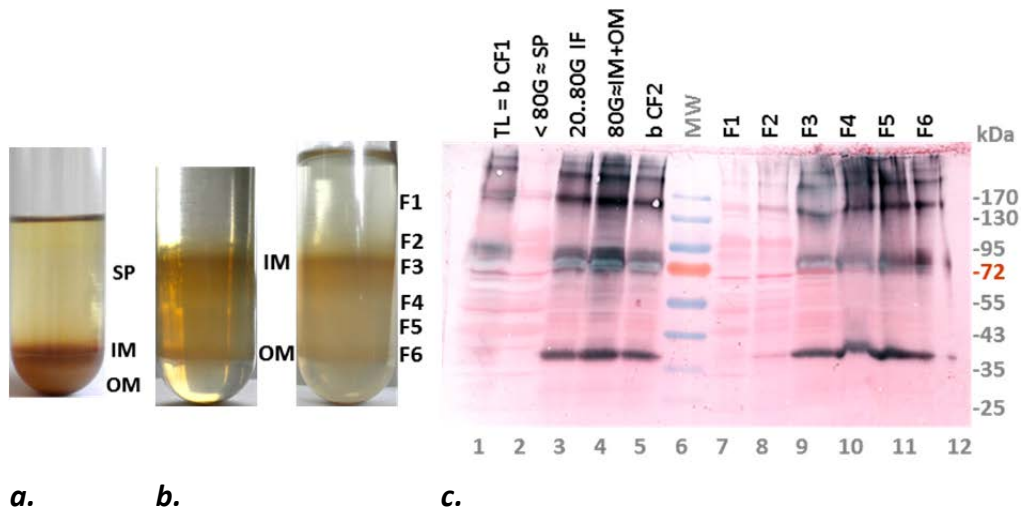
Figure 3.1.1.8 Subcellular fractionation of *E.coli* expressing connexin 26 in fusion with mistic. ***a.*** Separation of soluble and insoluble cell fraction using 80% glycerol cushion, 28-mL ultracentrifuge tube with the cell lysate after ultracentrifugation is presented. SP, soluble fraction of bacterial lysate. IM, inner (cytoplasmic) membrane fraction. OM, outer membrane fraction. ***b.*** Isopycnic fractionation of insoluble cell fraction. An ultracentrifuge tube with the fractionated sample was backlighted on the left pane and was front side illuminated on the right pane. Distinct layers were collected as fractions F1-F6. ***c.*** Anti-His-tag western blotting of the fractions obtained in buoyant density separation in glycerol gradient centrifugation. Lane 1, total non-fractionated cell lysate. Lane 2, supernatant above 20% glycerol separator layer after the first ultracentrifugation (cf1); it corresponds to soluble cell fraction. Lane 3, 20 % glycerol separating layer after cf1. Lane 4, 80% glycerol cushion layer after cf1. Lane 5, 80% glycerol cushion layer diluted approximately 3-fold for the isopycnic fractionation (cf2). Lane 6, molecular weight standards. Lanes 7 to 12 correspond to fractions F1 to F6 after cf2. Every sample lane was loaded with the representative amount of the total volume of the corresponding fraction.

Pure inclusion bodies were expected to settle under 100% glycerol as an apparent density of solvent-inaccessible constituent of inclusion bodies can be as high as 1.37 g/mL (97) what is much above the density of 100% glycerol of 1.27 g/mL (98), but the nature of the protein precipitated into inclusion bodies might influence that apparent density as well (99). We did not observe in this analysis any significant amount of connexin fusion protein in 100 % glycerol fraction suggesting no connexin containing inclusion bodies were present in this fraction.

3.1.1.9 Membrane separation provides no benefits in purification

It was discussed in section 3.1.1.7 that detergent solubility can correlate with the subcellular localization of the protein. To test this out we purified connexin fused with mistic from cellular fractions F3 and F6 obtained as described in section 3.1.1.8 and enriched with inner and outer cell membranes correspondingly. Membranes were solubilized in FC12 at pH 8.0, clarified by centrifugation at 100 000 *g*, and purified on CoNTA-sepharose. Fractions obtained in course of purification were analyzed by western blotting for the presence of connexin 26. Figure 3.1.1.9 demonstrates the distribution of the protein between the fractions.
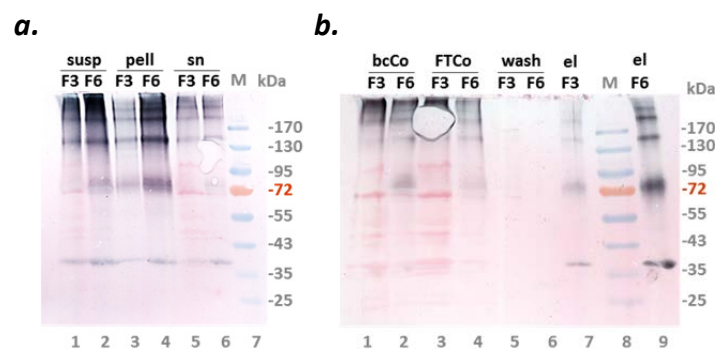
Figure 3.1.1.9 Comparison of IM ('F3') and OM ('F6') fractions. *a.* The solubilization efficiency of connexin 26 analyzed by western-blotting with anti-CxN19 antibodies. Lanes 1 and 2, membrane suspensions in FC12 prior to centrifugation. Lanes 3 and 4, FC12-insoluble fractions of IM and OM membranes. Lanes 5 and 6, FC12-soluble fraction of IM and OM membranes. *b.* CoNTA purification of FC12-solubilized connexin 26. Lanes 1 and 2, FC12-soluble fraction of inner and outer membranes prior to loading on CoNTA. Lanes 3 and 4, CoNTA non-binding flow-through fractions. Lanes 5 and 6, CoNTA wash fractions. Lanes 7 and 9, CoNTA eluates. Lane 8, molecular weight standards. Every sample lane was loaded with the representative amount of the total volume of the corresponding fraction.

A fraction of the detergent-soluble connexin was much higher in inner membranes than in outer membranes. However, a total amount of the solubilized protein is very similar. In contrast, a fraction of connexin that bound metal-affinity resin was higher for the protein solubilized from outer membranes comparing to inner ones. Pursuing a higher purification yield we chose to combine the membrane fractions for further purification.

3.1.1.10 Entherokinase cleavage of the purified proteins for mistic tag removal

As it was outlined in the section 2.1.5 the connexin 26 N-terminus is involved in the channel function and sensitive to modifications. To eliminate the expression and purification tag influence on the channel assembling and function these tags are to be removed from the produced connexins by a specific proteolysis. To test the tag removal efficiency fused proteins HgssML7Cx26 and HgssML8^{pp}Cx26 were purified from total membrane fraction in FC12 as it is described in section 3.1.1.7. According to Vergis et.al. (100) FC12 inhibits the enterokinase activity, yet not completely. To favor the enterokinase activity the protein samples were equilibrated with a cleavage buffer supplemented with either DDM or OTG prior to cleavage. The equilibration was accompanied by the protein precipitation. Precipitated protein was removed by high-speed centrifugation and the supernatants were subjected to enterokinase proteolysis. The samples were analyzed in SDS-PAGE presented in Fig. 3.1.1.10.
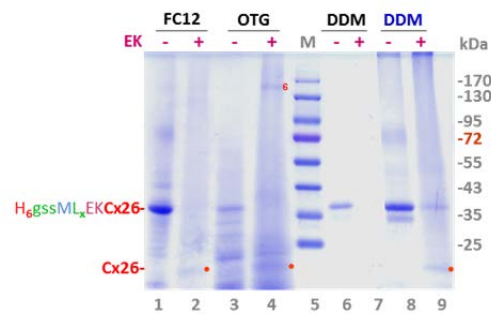
Figure 3.1.1.10 Samples of HgssML8^{pp}Cx26 (lanes 1-4, 6, 7) and HgssML7Cx26 (lanes 8 and 9) treated (**+**) and non-treated (**-**) with enterokinase (EK) were loaded in pairwise equivalents on 8%-16% SDS-PAGE and stained with cbb after the electrophoresis. The cleavage product positions corresponding to connexin 26 monomer (•) and presumable hexamer (**6**, petite) are indicated on the right of the bands in lanes 2, 4, 9. Detergent supplements (FC12, OTG, DDM) used in the EK cleavage buffer are indicated above the lanes. M - protein molecular weight standards.

The protease was active in all tested detergents including FC12 and permitted the expression tag elimination. At the same time, HgssML8^{pp}Cx26 treated in FC12 and DDM undergo heavy unspecific proteolysis. Indeed, the intensity of the highlighted (•) band corresponding to the desired cleavage product in lane 2 is much lower than that of non-cleaved fusion protein in lane 1. Along with that, in EK-treated samples extensive smearing emerged (lanes 2, 4, 9) additionally indicating the unspecific proteolysis accompanied by aggregation. In OTG, the integral intensity of the bands corresponding to connexin 26 monomer and hexamer that appeared (lane 4) as a

result of cleavage fairly correlates with the intensity of the initial non-cleaved fusion protein band (lane 3). The mobility of the hexameric band correlates with that observed for the hexamers purified from insect cells (39). The formation of hexamers as result of N-terminal tag removal was expected and accords with (70, 101).

In the case of HgssML7Cx26 the same amount of enterokinase (lane 9) was insufficient for complete specific cleavage of the fused protein. At the same time non-specific cleavage occurred as can be judged from smearing and aggregation appeared after cleavage as well as from lower total staining intensity of cleaved and non-cleaved bands comparing to that of control sample non-treated with protease (lane 8). Similar effect was observed with human connexins 32 and 43 fused with N-terminal mistic tag that are also under a study in our group. These homologous proteins allowed purification directly in DDM, which, comparing to FC12, proteases tolerate better (100, 102, 103). Yet, also in those cases variations in cleavage conditions such as proteases used for cleavage (Factor Xa, thrombin, TEV protease), their concentrations, reaction temperature (4°C, 20°C, 30°C), treatment time (1h to 48h), and detergent supplement resulted in either incomplete or non-specific cleavage. Besides, no separation of cleavage product from initial fusion protein was achieved by size-exclusion, ion-exchange, and metal-chelating chromatography with samples of cleavage reaction mixtures. Cleavage products remained associated with the non-cleaved substrate till SDS-PAGE electrophoretic separation (104, 105).

3.1.2 Mistic tag removal *in vivo*

To surmount the resistance of the expression tag to cleavage from a purified fusion protein a specific proteolysis *in vivo* prior to purification can be employed. Mistic is advantageous in targeting a cargo protein to membrane but complicates a subsequent purification. To cleave mistic of connexin in living cells we made use of an auxiliary plasmid encoding a cysteine protease from tobacco etch virus (TEV). We had to update the connexin-encoding plasmid too to make the fusion protein susceptible to TEV protease. Figure 3.1.2 depicts an expression system employing *in vivo* cleavage.
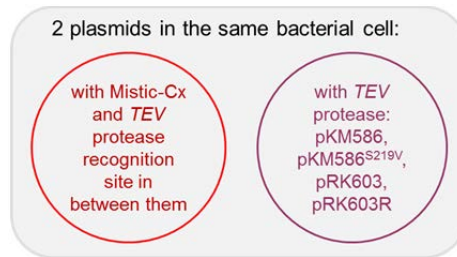
Figure 3.1.2 A scheme for TEV protease co-expression for *in vivo* cleavage.

3.1.2.1 Helper plasmids for in vivo Mistic removal

An auxiliary plasmid must coexist with the plasmid encoding connexin in the same cell and should be maintained therefore via a different origin of replication. Besides the compatibility, this enables the control of the plasmid copy number in a bacterial cell and thereby a level of TEV protease and its proteolytic activity. Adjusting a level of the proteolytic activity might be necessary in the case when non-specific cleavage takes place (106). Suitable plasmids, pKM586 and pRK603 (Figure 3.1.2.1), were developed by Kapust (107). The plasmids share the same P_L / tetO − rrnB T1 transcription cassette incorporating an S219D mutant of TEV protease and differ in the replication origins providing approximately 3 and 30 plasmid copies per bacterial cell (108), correspondingly. The S219D mutation amends the TEV protease by alleviating its auto inactivation (109). Both plasmids enable selection by endowing kanamycin resistance to bacterial hosts and belong to a versatile and modular pZ vector family (108) from *Expressys*. In opposite to *Expressys* expression hosts, SE1 suicide cells selected for pSCodon driven expression lack tetracycline repressor *tetR* gene what leaves P_L promoter in a constantly active state.
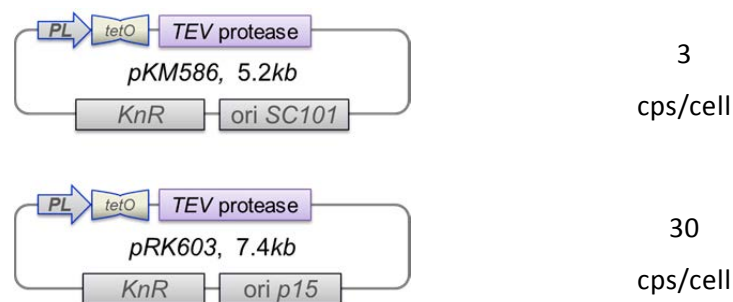
Figure 3.1.2.1 Schemes of auxiliary plasmids for TEV protease co-expression.

3.1.2.2 Optimization of the Mistic to connexin linker to facilitate cleavage

One of the probable cause for the incomplete proteolysis described in the section 3.1.1.10 could be an insufficient exposition of the protease recognition site. To improve the cleavage site accessibility we further elongated the mistic to connexin linker by 24 amino acids via incorporating an additional recognition site for thrombin protease and streptag II (110), an additional immunochemical detection and purification tag (Figure 3.1.2.2). The limited accessibility of the cleavage site within the connexin close proximity may also be associated with a particular orientation of the connexin N-terminus that folds inside the entrance of a properly assembled connexin hexamer. An additional proteolytic cleavage site recognized by thrombin and placed closer to mistic C-terminus could help to examine this hypothesis. In this position, the thrombin site is known to be well accessible for the specific proteolysis (50).
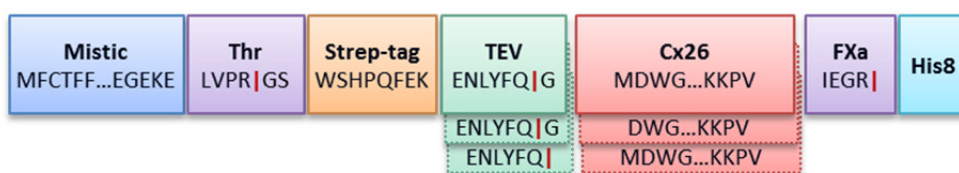
Figure 3.1.2.2 Mistic-connexin fusion constructs updated for *in vivo* cleavage. For the comments on the variations in TEV protease recognition site to connexin 26 junction see Table 3.1.2.2.

The TEV protease activity, in opposite to Enterokinase or Factor Xa applied before, is influenced by amino acid residue in P' position of its recognition sequence and tiny residues like glycine or serine are preferable (111, 112). There were reports (111, 113) suggesting that methionine in the P' position is well tolerated too.

Table 3.1.2.2

construct short alias	sequence*	features
TeV	ENLYFQ|GMDWG	canonical TEV protease recognition site, intact connexin 26 N-terminus
TeG	ENLYFQ|GDWG	canonical TEV protease recognition site, connexin 26 N-terminus is truncated by M1
TeM	ENLYFQ|MDWG	non-canonical TEV protease recognition site with methionine in P' position, intact connexin 26 N-terminus

*) vertical bar **|** indicates where cleavage occurs. TEV protease recognition site is shown in green and connexin 26 N-terminus is in red.

Taking into account expected connexin sensitivity to the N terminus extensions and modifications we have assembled and tested different junctions of the TEV protease recognition site and the connexin N-terminus that are presented in Table 3.1.2.2 and Figure 3.1.2.2 as well.

3.1.2.3 Optimization of the His tag position to facilitate CoNTA binding

Since a reasonable binding with CoNTA resin have not been achieved using N-terminal His tag in the constructs described in section 3.1.1 we moved the His tag to a C-terminus and make it longer. According to the available crystallization data (23) the connexin C-terminus is rather flexible and should provide an arm sufficient to expose the His tag outside a detergent micelle; actually His tag at this position have already been used for connexin 26 purification from insect cells (38, 39, 59). Still, it should be kept in mind that even a few amino acid elongation at a connexins C-terminus can significantly impair their function (114) and stability (81). Factor Xa (115) recognition site preceding the purification tag should enable removal of C-terminal connexin extension (Figure 3.1.2.2).

3.1.2.4 Tev protease intracellular level influences cleavage efficiency

To set up the expression system described above SE1 *E.coli* cells were transformed first with pSCodon1.2 plasmids bearing either of TeG, TeM, or TeV constructs and selected for ampicillin resistance. Then, the obtained transformed cells were made competent, transformed with either pKM586 or pRK603 plasmids, and selected for kanamycin and ampicillin resistance. Expression was carried out in lactose-based autoinduction media according to (58) in 25 mL scale in shaking flasks. Inoculates were initially incubated at 37°C and when either of the cultures reached an optical density of approximately 2.0 AU_{600} the culture was transferred to 20°C and incubated for 16 to 18 hours longer. Bacteria were harvested and analyzed by western blotting using anti-His antibodies for the specific detection of the produced connexin (Figure 3.1.2.4).
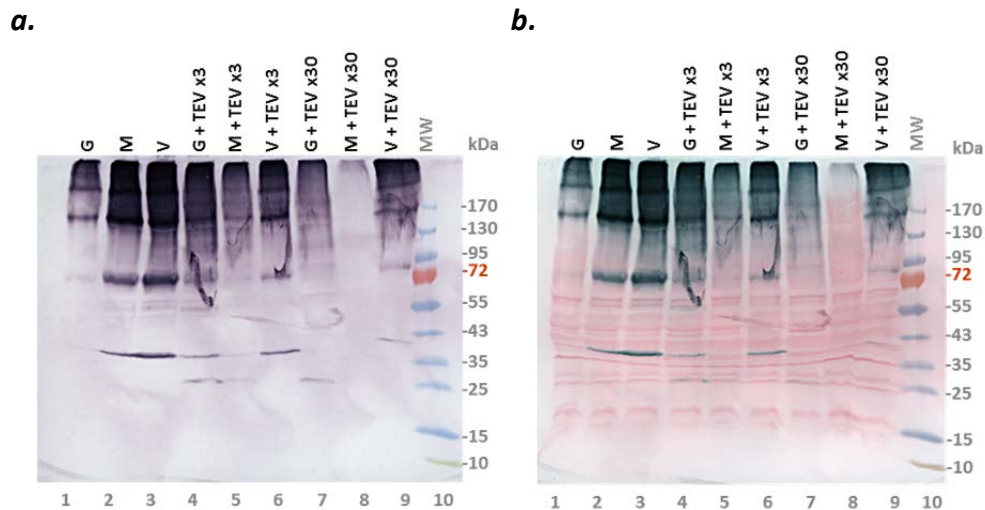
Figure 3.1.2.4 *In vivo* cleavage of mistic-connexin fusion constructs TeG, TeM, and TeV (the designations are explained in Figure 3.1.2.2 and Table 3.1.2.2). 'G', 'M', and 'V' correspond to fusion constructs TeG, TeM, and TeV respectively. 'TEV x3' and 'TEV x30' correspond to auxiliary plasmids pKM586 and pRK603 respectively. Lane 10, molecular weight standards. *a.* Immunoblot with anti-His-tag antibodies. *b.* The same immunoblot stained with ponceau S to demonstrate equivalent loading of cell lysates.

The most prominent bands corresponding to connexin 26 with no mistic are observed in lanes 4 and 7 corresponding to TeG connexin fusion construct expressed in the presence of low copy pKM586 and higher copy pRK603 TEV protease encoding helper plasmids. While increasing the TEV protease activity in the case of higher copy helper plasmid did not elevated a total amount of the desired product of cleavage it eliminated the non-cleaved connexin fused with mistic that made major complications in purification described previously in section 3.1.1. Surprisingly, the TeV construct implementing quite a typical design of junction of the TEV protease to a protein of interest (116) showed no desired product of cleavage at all. The TeM construct providing non-canonical methionine in P' position of TEV protease recognition site demonstrated reduced susceptibility to the specific proteolysis and possibly underwent non-specific proteolytic degradation as faint immuno-staining in lane 8 could suggest. The non-specific proteolysis by the overproduced TEV protease could also explain the depletion of high-order oligomeric forms of connexin in lane 8. Therefore, a choice between high- and low-copy helper plasmid should be done with care. Here, TeG construct was selected for the further purification trials.

3.1.2.5 In vivo cleavage permitted complete removal of the mistic expression tag

To choose between high- and low-copy helper plasmid we tried both of them to purify connexin free of mistic. For a small scale preliminary purification trial the TeG construct was expressed as described in section 3.1.2.4 with pKM586, pRK603, or pRK603R plasmids or alone. The pRK603R plasmid included in the test as an additional negative control was derived from pRK603 by insertion of a tetracycline repressor *tetR* gene upstream the *kanR* gene. Bacteria were harvested and solubilized in buffered 1% DDM solution with sonication. Non-solubilized material was removed by low speed centrifugation for 10 min at 1000 *g* and the supernatant was purified on CoNTA magnetic beads. Figure 3.1.2.5 represents the SDS-PAGE and western blotting analyses of the fractions obtained.
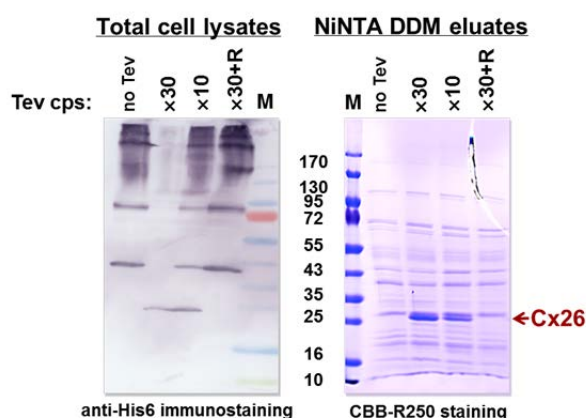
Figure 3.1.2.5 Purification of connexin free of mistic ('Cx26') that was expressed in fusion with mistic using 'TeG' construct and cleaved *in vivo* using pRK603 ('x30'), pKM586 ('x10'), or pRK603R ('x30+R') helper plasmids. Left pane, immunoblotting of crude cell lysates. Right pane, SDS-PAGE analysis of the NiNTA-purified samples. Every sample lane was loaded with the representative amount of the total volume of the corresponding elution fraction. M, molecular weight standards.

With this approach we provided the TEV protease residing in the cytoplasm with the detergent-solubilized connexin fusion construct. If membranes had been isolated prior to solubilization the TEV protease would be washed out. The TEV protease was reported to tolerate well (100, 102, 103) the detergent used for solubilization. Detergent micelles represent an alternative environment that could alter an exposition of the TEV protease recognition site comparing to lipid bilayers. Additionally, those cleavage sites that escaped the cytoplasmic TEV protease due to translocation into periplasmic or extracellular space during expression could get an additional

opportunity for cleavage after membrane solubilization. Thereby, we tried to use the entire proteolytic activity of the TEV protease produced during expression and to find out whether the level of TEV protease is the only factor responsible for the incomplete cleavage observed. But, we did not find any significant changes in a cleavage pattern after the additional treatment described above (compare Figure 3.1.2.5, lanes 2 and 3 with Figure 3.1.2.4, lanes 7 and 4, correspondingly). Furthermore, omitting the membrane isolation and washing steps resulted in a significant level of contamination of connexin 26 preparations after CoNTA purification (Figure 3.1.2.5, panel *b*).

As one could anticipate, the TEV protease activity indeed can be a limiting factor responsible for the incomplete removal of the mistic tag from the connexin fusion construct when samples with low copy helper plasmid (Figure 3.1.2.5, lane 3) is compared to the higher copy one (Figure 3.1.2.5, lane 2).

3.1.2.6 In vivo cleavage becomes less efficient in large-scale cultures

Pursuing the completeness of mistic tag removal we have chosen the TeG construct in combination with pRK603 helper plasmid to upscale the protein production. The 5-L culture was grown as described in section 3.1.2.4. Bacteria were harvested, treated with lysozyme, DNase, and disrupted in TrisHCl buffer at pH 8.0 by mechanical stress applied with a French Press. Total cell membranes were isolated by ultracentrifugation with a glycerol cushion as described in section 3.1.1.8. The collected membranes were further fractionated in two sequential isopycnic centrifugations at 100 000 *g* in ultracentrifuge tubes with a series of 40%, 60%, and 80% glycerol layers. Figure 3.1.2.6 represents the immunostaining analysis of the fractions obtained.

Lane 3 representing total membrane fraction demonstrates that the cleavage efficiency is far not as good in the large scale culture as in the small scale one. At the same time, a membrane fractionation revealed that uncleaved connexin fusion protein probably accumulates in outer membranes (compare lanes 9 to 8). The NiNTA purification of the solubilized inner membrane fraction improved the ratio between cleaved and non-cleaved forms further (pane *c*, lane 1), still we did need to have no uncleaved form at all. As can be seen from pane *a* the non-cleaved product emerges after the some amount of connexin completely free of mistic is accumulated.
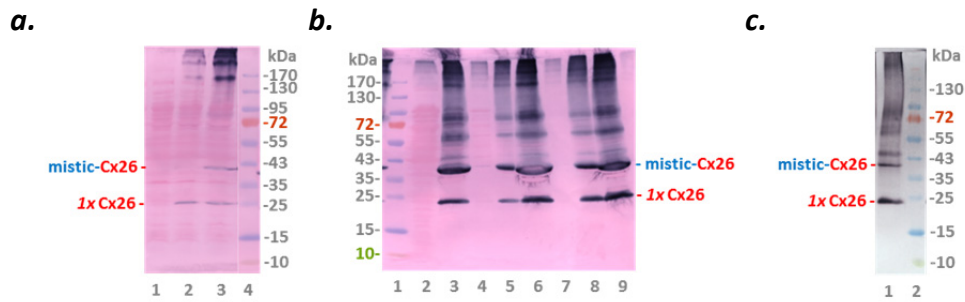
Figure 3.1.2.6 *In vivo* cleavage in large-scale SE1 cells culture producing TeG construct analyzed by immunoblotting with anti-His-tag antibodies. *a.* Time-dependence of connexin expression level and cleavage efficiency. Lanes 1, 2, and 3 correspond to 2, 4, and 25-hour AIM cultures. Sample lanes were loaded with total cell lysates in equal amounts as demonstrated by non-specific staining with ponceau S. Lane 4, molecular weight standards. *b.* Isolation of inner and outer membranes. Lane 2, soluble cell fraction. Lane 3, insoluble cell fraction. Lanes 4, 5, and 6 correspond to fractions containing <60%, 60%, and >60% of glycerol after first round of isopycnic fractionation. Lanes 7, 8, and 9 correspond to fractions containing <60%, 60%, and >60% of glycerol after second round of isopycnic fractionation. Every sample lane was loaded with the representative amount of the total volume of the corresponding fraction. Lane 1, molecular weight standards. Immunonegative proteins were stained with ponceau S. *c.* Lane 1, connexin 26 sample from inner membranes purified NiNTA in DDM. Lane 2, molecular weight standards.

This did not happen in case of small scale cultures possibly due to lower overall expression yield. Explaining this effect by insufficient level of TEV protease activity we have recloned the TEV protease gene under control of T7 promoter in case of homologous connexin 43 (collaborative work, see (105)). The major effect of the further increasing the TEV protease activity level was a considerable decrease of the total yield of produced connexin. Therefore we paid more attention to the exposition of the TEV protease cleavage side in our connexin fusion protein to cytoplasmic TEV protease. Indeed, a ratio between cleaved and non-cleaved forms in inner membranes is close to 1:1 corresponding to a random orientation of the cleavage side relative to the membrane plane after the protein insertion. The abundance of the non-cleaved fusion construct in the outer membrane fraction further support the hypothesis of non-oriented insertion that was actually not expected taking into account results of Roosild (49). This observation made us looking for another driver for connexin expression.

3.2 ESR tag

3.2.1 ESR as a potential expression and membrane targeting driver

Besides mistic, a number of other integral membrane proteins amenable for overexpression at level of dozens of milligram per liter of culture were identified last years. None of them was explored as much as mistic in a role of membrane-targeting expression driver for other poorly expressed proteins. We decided to try one of these proteins that we produce in our lab, a bacteriorhodopsin from *Exiguobacterium sibiricum* (117), for connexin expression. The protein demonstrates stable high yield of 20 mg per litter of culture after purification under mild condition from *E.coli* membranes (118, 119). The expressed protein predominately adopts functional conformation. The functional insertion into *E.coli* membrane generally implies C-terminus of ESR facing cytoplasm (120–124). According to the structure data available (125) the ESR C-terminus is not involved in the maintaining of the protein integrity and is available for fusion with connexin N-terminus.

3.2.2 Connexin fusion constructs with ESR as an expression driver

If ESR can be an efficient driver for the targeted overexpression of membrane proteins it would be reasonable to identify a minimal fraction of the driver that is responsible for the effect. To test this potential of ESR we assembled the following constructs.
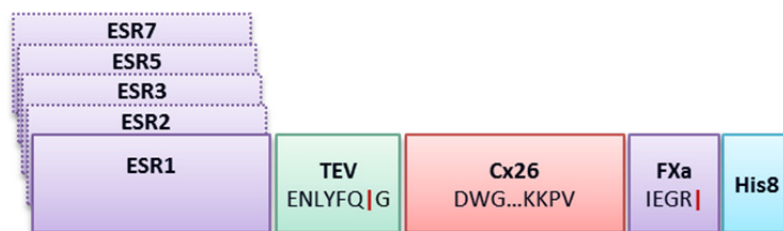

Figure 3.2.2 Scheme of connexin 26 constructs for testing ESR as N-terminal expression and membrane insertion tag. As full length (ESR7) and its truncations representing first single (ESR1), two (ESR2), tree (ESR3), or five (ESR5) transmembrane helices were tested.

Designing the depicted truncated ESR constructs we assumed that the insertion of ESR into lipid membrane starts from its N-terminus and therefore it is the N-terminal fraction what directs and determines the insertion efficiency and orientation of ESR in *E.coli* membrane. Oriented insertion of ESR in *E.coli* membrane would control as an

exposition of TEV protease recognition site and an orientation of connexin in bacterial membrane.

3.2.3 ESR did not assist connexin expression in *E.coli*.

The listed above ESR-connexin constructs were expressed in SE1 *E.coli* cells with or without the TEV protease provided via the pRK603 plasmid as described in section 3.1.2.4. Bacteria were harvested and analyzed by western blotting using anti-His antibodies for the specific detection of the produced connexin (Figure 3.2.3.)
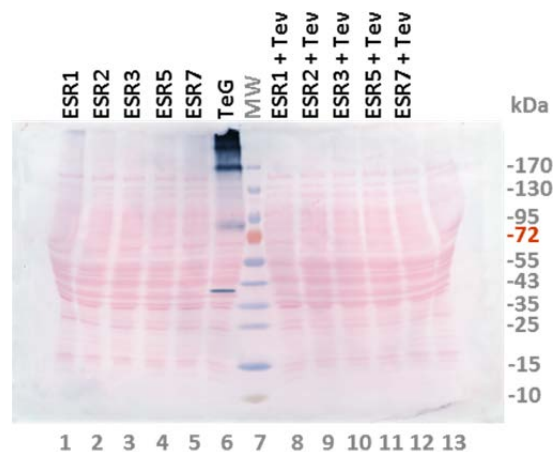

Figure 3.2.3 Expression level of connexin 26 in fusion with bacteriorhodopsin from *Exiguobacterium sibiricum* analysed by immunoblotting with anti-His-tag antibodies. Lanes 1-5 correspond to ESR1, ESR2, ESR3, ESR5, and ESR7 tags. Lane 6, 'TeG' construct with mistic tag as a positive expression control, see Figure 3.1.2.2 and Table 3.1.2.2 for the description. Lane 7, molecular weight standard. Lanes 8-12 correspond to the same constructs as Lanes 1-5 but co-expressed with TEV protease.

Neither of the ESR to connexin fusion constructs resulted in any detectable expression product. As it turned out that ESR (118), in spite of its reported stability and high level expression in *E.coli*, is dramatically sensitive to as long C-terminal extension as connexin is. The observed effect is rather surprising since the closely related to ESR protein, bacteriorhodopsin I from *Haloarcula marismortui*, was reported to be capable to significantly improve an expression yield of undecaprenyl pyrophosphate phosphatase and carnitine / butyrobetaine antiporter (126). However, it should also be taken into account that both cargo proteins were of *E.coli* origin.

3.3 *In vitro* expression

Connexin 26 has already been produced *in vitro* using wheat-germ and reticulocyte lysates (127, 128). The yields achieved were sufficient to study the protein incorporation into membranes yet too low for crystallization trials. A close homologue of connexin 26, connexin 32, yielded above 0.5 mg of the protein per mL of *in vitro* reaction mixture based on *E.coli* "S30" lysate (129). And formation of hexameric channels observed by TEM was reported in that work. Connexin 43 was also produced *in vitro* with *E.coli* translational machinery assembled from individual constituents known as PURE system (130). The particular feature of the PURE system is lack of endogenous lipids that are usually presented in S30 lysate. It was demonstrated that when lipids in form of liposomes are presented during the biosynthesis connexin 43 integrates into liposomes co-translationally and form dye-permeable channels. These examples of connexin 26 homologous proteins demonstrate that *in vitro* expression system derived from *E.coli* is capable to deliver a reasonable amount of functional connexin.

3.3.1 Selecting an approach

When the protein of interest imposes specific requirements on an expression system which living *E.coli* does not intrinsically meet it is the format of cell-free expression what could provide a convenient alternative for identification and optimization of critical expression parameters (131). And a number of reports grows where producing of milligrams of functional integral membrane protein was made affordable in cell-free system (82, 129, 132–135).

An opened and versatile format of cell-free expression prompted the development of a variety of implementations (136, 137). We have used a continuous exchange reaction (138, 139) to feed working ribosomes with fresh low molecular weight substrates and to remove the biosynthesis by-products away from the ribosomes (140). The system tolerates supplements that are beneficial for the membrane protein folding including some detergents and lipids (141, 141–145).

Basically, the strategy for producing membrane proteins in this system includes adjusting the composition of the reaction mixture for the best translation efficiency in the absence of any detergent and then screening of amphiphilic supplements to

optimize the yield of a soluble and functional protein. The cell-free expression of membrane proteins performed without any typical membrane mimetics like detergents or lipids is often referred to as pellet-mode reaction (pCFE) since the produced membrane protein usually precipitates in these conditions (132). Similarly, when the *in-vitro* reaction mixture is supplemented with detergents, lipids, or both the dCFE, lCFE, or dlCFE designations are used (135).

3.3.2 Connexin construct for *in vitro* expression.

For cell free expression we have inserted the connexin gene into the high-copy number plasmid pIVEX2.3. The plasmid provides ColE1 replication origin, beta-lactamase gene for ampicillin resistance, phage T7 promoter/terminator transcription cassette with ribosome-binding site, lacking any operator. A new connexin 26 construct had no any N-terminal expression tag and consisted of *E.coli* optimized cDNA for the wild type connexin 26 with the C-terminal extension of 19 amino acids including Factor Xa recognition sequence and octa-histidine stretch as a purification tag (Figure 3.3.2).
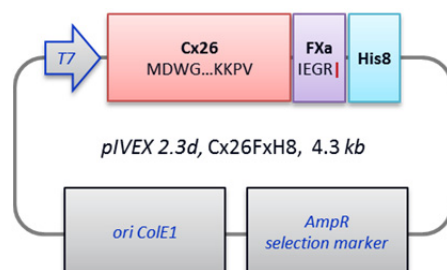

Figure 3.3.2 Scheme of the plasmid for *in vitro* expression of connexin 26.

3.3.3 Detergent selection for best solubility

The cell free expression mixtures based on *E.coli* S30 in-house made extract were prepared in an analytical scale as described in the methods section 4.3.10 "*In vitro* expression" and summarized in the appendix A5. The mixtures were provided with the plasmid encoding Cx26FxH8 construct (Figure 3.3.2). A series of the selected detergents which typically endow better solubility to membrane proteins expressed *in vitro* (139, 146, 147) were tested to identify those that are best suited for connexin 26. The detergents were added to the reaction mixtures according to the Table 3.3.3.

name	Brij35 (C12E23)	Brij58 (C16E20)	Brij78 (C18E20)	DDM	Digitonin	DHPC
Working conc, %	0.1	1.5	1.0	0.1	0.4	0.2
CMC, % (mM)	0.01 (0.08)	0.008 (0.075)	0.005 (0.046)	0.008 (0.17)	0.09 (0.73)	0.07 (1.4)

Table 3.3.3 Detergent supplements for *in vitro* synthesis and their properties. Brij detergents share the common polyoxyethylene-(y)-alkyl(x) ether structure that can be designated as CxEy. The CxEy notations are given together with the corresponding Brij*NN* names.

The used detergents concentrations were limited by their inhibitory effect on the transcription/translation machinery and by inefficiency of the further concentration raise. After the reaction had completed the mixtures were separated by centrifugation at 10000 *g* for soluble and insoluble fractions and analyzed in 12% SDS-PAGE (Figure 3.3.1).
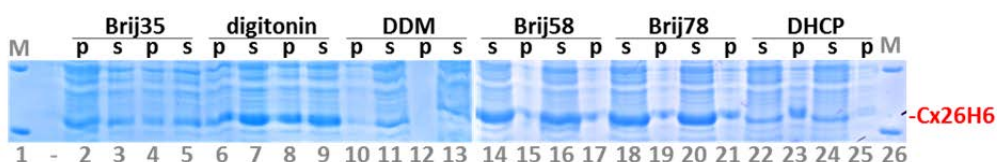

Figure 3.3.3 *In vitro* expression level and solubility of connexin 26 in the presence of selected detergents (listed in Table 3.3.3) analyzed by SDS-PAGE. p, insoluble fraction. s, soluble fraction.

It is remarkable here that the cell-free synthesis conditions resulted primarily in monomeric form of connexin 26, unlike expression of mistic-fusion constructs in living *E.coli*. The detergents with larger hydrophobic moieties, as Brij78, Brij58, and digitonin, allowed higher level of overall synthesis and of a soluble form of connexin as well comparing to those with moderate size hydrophobic tails. While it is tempting to conclude that longer hydrophobic tails of Brij58 and Brij78 provided also a higher fraction of a soluble form comparing to Brij35 or DDM the direct comparison is not possible here because of the significant difference in the detergent concentrations applied. Due to their poor tolerance by the cell-free synthesis machinery, DDM and Brij35 had to be used at so low concentration that possibly was just ineffective in maintaining the synthesized connexin in solution.

Regarding the best performance of Brij58 and Brij78 it is also worth mentioning that their hydrophobic chains mimic the major fatty acids of cell plasma membrane in their

length that could better satisfy a potential lipid binding valences of connexin and could contribute to the effect. Also, detergents sharing a common CxEy structure are often highly rated (140, 148, 149) by their ability to maintain membrane proteins in solution under conditions of cell free protein synthesis.

3.3.4 dCFE up-scaling

The prep scale reactions were carried out in a 1-3 mL scale with a 14 fold excess of a feeding mixture for continuous exchange as described in the methods section "*In vitro* expression" and summarized in the appendix A5. As it was selected in the small scale screening (section 3.3.3), Brij78 was supplied to the reaction and feeding mixtures to maintain connexin in a soluble form. After 12h-incubation at 32°C, the reaction mixtures were separated for soluble and insoluble fractions by 10000 *g* centrifugation. The soluble fraction was purified on NiNTA sepharose as described further.

3.3.5 Purification of the *in vitro* - produced Cx26His6 in Brij35.

Since the *in vitro* reaction mixtures contained low molecular weight compounds compromising downstream purification the macromolecules from the soluble fraction produced as described in the section 3.3.3 were transferred into NiNTA binding buffer by group-separation mode size exclusion chromatography on 10 mL Sephadex G25 column. The binding buffer was supplemented with 0.5% Brij35 to maintain the connexin solubility. This connexin sample was applied on NiNTA sepharose and extensively washed with the binding buffer. Proteins bound to the resin were eluted then with 300 mM imidazole in binding buffer. The fractions collected were analyzed in 12.5% SDS-PAGE with cbb-G250 staining (Figure 3.3.5).

The total yield of connexin 26 was estimated as 1 mg per 1 mL of the reaction mixture and no significant lost occurred in course of purification. Yet, using non-ionic Brij35 we did not obtain homogenous connexin 26 from the soluble fraction of *in vitro* reaction mixtures by metal affinity chromatography. There was also a special interest of replacing Brij35 detergent bound with the protein for DDM or OG due to their higher success rate in crystallization trials (80) comparing to polyoxyethylene-alkyl ethers and better suitability for connexin channel activity tests.
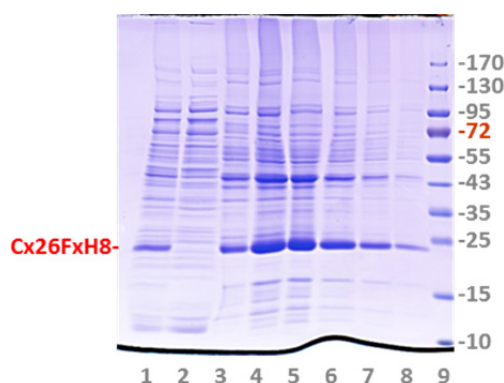
Figure 3.3.5 SDS-PAGE analysis of the fractions obtained in the course of NiNTA purification in a mild detergent of connexin 26 produced *in vitro* with Brij78. Lane 1, the soluble fraction of the *in vitro* reaction mixture prior to loading on NiNTA sepharose. Lane 2, the unbound sample flown through the NiNTA resin. Lanes 3-8, NiNTA bound proteins elutes with 300 mM imidazole, serial fractions. Lane 9, molecular weight standards. Staining: cbb.

3.3.6 Purification of the *in vitro* - produced connexin 26 in OG.

The soluble fraction of the *in vitro* produced connexin was purified on NiNTA sepharose as described in section 3.3.5, but after the sample loading and washing with the binding buffer supplemented with Brij35 the column was washed with 0.8% OG and bound proteins were eluted with imidazole followed by EDTA, both supplemented with OG. The collected fractions were analyzed by SDS-PAGE and western-blotting (Figure 3.3.6, *a* and *b*). The comparison of loaded, unbound, and eluted amounts of connexin suggested that a significant fraction of loaded connexin remained on column. A fraction of the NiNTA-resin remained after the purification with the associated connexin was loaded directly in a polyacrylamide gel for SDS-PAGE (Figure 3.3.6, *c*).

We can conclude that the exchange of Brij35 for OG caused connexin "binding" to the column in course of chromatography possibly due to fast aggregation followed by precipitation inside the pores of sepharose beads.
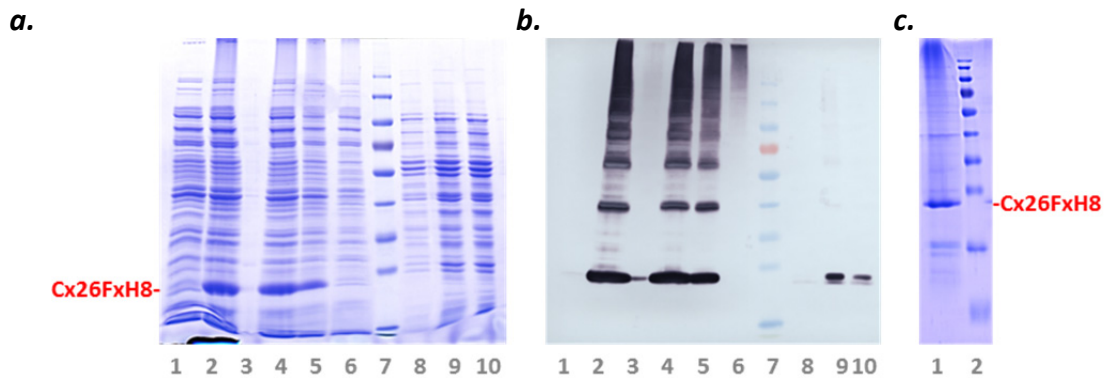
Figure 3.3.6 SDS-PAGE (panels **a, c**) and immunoblotting (panel **b**) analyses of the fractions obtained in the course of NiNTA purification of connexin 26 produced *in vitro* with Brij78 and using OG in purification. **a.** and **b.** Lane 1, 0h CFE mix. Lane 2, 16h CFE mix. Lane 3, CFE mix, insoluble fraction. Lane 4, CFE mix, soluble fraction. Lane 5, soluble fraction, desalted, before loading on NiNTA. Lane 6, NiNTA flow-through. Lane 7, MWST. Lane 8-10, NiNTA binding fractions eluted with 300 mM imidazole, serial fractions. **c.** Lane 1, NiNTA resin after imidazole and EDTA elution. Lane 2, MWST. Staining: **a** and **c**, cbb; **b**, anti-His-tag antibodies.

An anionic detergent was required to solubilize and elute these aggregates under the electrophoresis conditions. The same effect was observed when Brij35 was substituted for DDM. DDM due to longer hydrophobic alkyl chain and bulkier hydrophilic sugar moiety is less aggressive comparing to OG which is well-known aggregation promoter (150). Moreover, both, DDM and Brij35, have the same twelve-carbon hydrophobic tail. Apparently, the reason for the precipitation upon these detergents exchange should lie in the difference between polyoxyethylene and maltose moieties.

Since the metal affinity chromatography turned out to be inefficient in preparing a suitable for crystallization protein from a soluble fraction of the cell free synthesized connexin 26, we had to look for an alternative non-chromatographic approach.

3.3.7 Purification of Cx26FxH8 produced in pCF mode

As it was already mentioned in section 3.3.1, detergent supplements are not always used in the *in vitro* synthesis of the membrane proteins. With no any supplements the produced membrane proteins usually precipitate. Despite the precipitate might include a trace amount of co-precipitated contaminant, it is often enriched with the protein of interest sufficiently to be used in solubilization and refolding trials. Therefore, we produced connexin 26 as described in section 3.3.4 but with no any

membrane mimetics added. Then we washed the obtained pellet gently to find out whether the purity suitable for downstream applications can be obtained. The washed pellet was solubilized in 0.5% SDS. The obtained fractions were analyzed in SDS-PAGE.
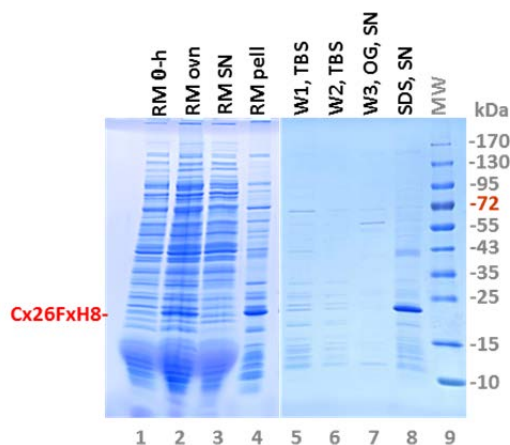

Figure 3.3.7 SDS-PAGE analysis of the fractions obtained in cell free expression of Cx26FxH8 with no detergents. Lane 1, 0-hour reaction mixture (RM). Lane 2, RM incubated overnight. Lane 3, soluble fraction of RM. Lane 4, insoluble fraction of RM. Lane 5 and 6, supernatant after washing the RM pellet with no detergents. Lane 7, washing of the RM pellet with 1% OG. Lane 8, solubilization of RM pellet in 1% SDS. Staining with cbb.

As it was already observed in section 3.3.3, the total connexin yield dropped significantly, in the presented case 3 to 4 times, after excluding the favorable detergent from the reaction mixture, and this does not accord with the previous observations (134). Nascent peptide chains probably start aggregating soon after synthesis, before leaving ribosomes, and block them. Washing the obtained pellet with buffer with no detergents removed a considerable fraction of contaminants (figure 3.3.7, lanes 5,6). Supplementing the washing buffer with OG, DDM, or DHPC did not result into solubilization of connexin but additionally removed a minor fraction of contaminants. The purity of the connexin that remained in the washed pellet was found to be insufficient for crystallization and further purification was required.

3.3.8 An alternative to NiNTA purification

Since mild detergents did not solubilize the precipitated connexin as good as SDS, could the purification be accomplished in SDS? The samples of connexin 26 described in sections 3.3.4 to 3.3.7 demonstrated abnormal electrophoretic mobility similar to the connexin produced in live cells as it have already been discussed in section 3.1.1.2.

Indeed, while the expected molecular weight for the Cx26FxH8 construct is 28.28 kDa, under SDS-PAGE condition it shows up as 22 kDa. The presence of intact N- and C-termini was confirmed immunochemically with N19 antibodies, developed against peptide mapping 12 N-terminal amino acids of connexin 26, and anti-His-tag antibodies. Therefore, proteolytic degradation does not account for the connexin abnormal mobility. The elevated electrophoretic mobility of membrane proteins had already been reported in (61, 151). It was suggested that membrane proteins, unlike many soluble cytoplasmic proteins (152), do not unfold in Laemmli electrophoresis buffer completely (72) and retain, at least in part, their secondary structure. It should be also mentioned that soon after SDS-PAGE is started an excess of SDS from the sample buffer runs out of the protein sample and the running buffer, which contains just half the CMC of SDS, replaces the sample buffer in the protein sample. This detergent depletion can favor the membrane protein refolding, at least partial.

On the other side, a membrane protein produced as a precipitate in cell-free expression in the absence of membrane-mimicking supplements can hold significant part of native structure of the expressed membrane protein as it was demonstrated by NMR (153). For those membrane proteins that do have sites for high affinity lipid binding and do need a limited number of lipid molecules to adopt a native conformation the fact of partial presence of correct folding in the discussed cell-free expression condition can be attributed to the trace amount of lipids that remains in S30 extract (132, 135, 149, 154). Indeed, the S30 extract was estimated (139) to contain approximately 0.1 mg of lipids per equivalent of 1 mL of *in-vitro* reaction mixture; and 1 mL of the reaction mixture can result in 1-2 mg of the synthesized membrane protein in a good case. This corresponds to protein to lipid ratio of about 1:2 that might already be close to the protein requirement in essential lipids.

Connexin plaques are notoriously resistant to solubilization in detergents including SDS in low concentration (68). Gap junction solubilization in as high as 2% SDS resulted in connexon dissociation into monomers (69). The incomplete unfolding could contribute to the discussed abnormal mobility of connexin.

Taking together the discussed above observations we assumed that the expressed *in vitro* connexin after solubilization in SDS retains, at least in part, its folding and we could try preparative scale electrophoresis to isolate a high purity connexin.

3.3.9 Purification of the *in-vitro* produced Cx26 by SDS-PAGE electrophoresis

We produced connexin 26 as described in section 3.3.4. We chose a detergent supplemented *in-vitro* reaction since omitting detergents reduced significantly the reaction yield and cost-efficiency. After the reaction had completed, the mixture was desalted by dialysis to reduce the potassium concentration and supplemented with 0.5% SDS. The mixture was then loaded in 'preparative scale' SDS-PAGE with a 2 mL sample pocket. The electrophoresis was developed in the electric field of 10 V/cm, current density of 20 mA/cm^2, and dissipated power limitation of 0.2 W/cm^3. A temperature in the electrophoresis chamber was maintained at 12°C with a circulating water thermostat.

After finishing the protein separation (Figure 3.3.8), the monomeric connexin was identified by precipitation of dodecyl sulfate inside the gel with potassium chloride. Major protein bands appeared as more clear (darker) areas at a turbid background.
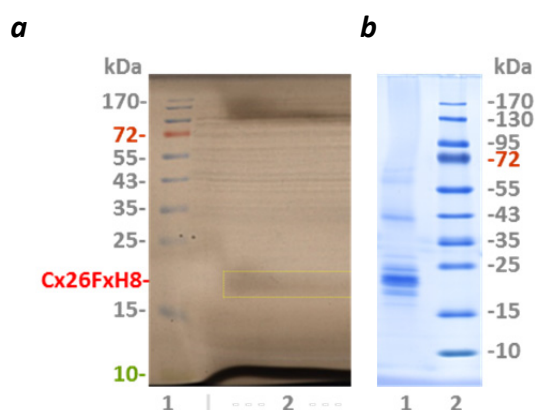

Figure 3.3.8 SDS-PAGE purification and analysis of the fractions obtained. *a.* a fragment of 'preparative gel' developed in 3M KCl. Lane 1, MWST. Lane 2, *in-vitro* reaction mixture, 1 ml. Area indicated by a yellow frame was excised for electroelution. *b.* Analysis of the purified sample. Lane 1, electroeluted sample. Lane 2, MWST. Note monomeric connexin self-association resulting in SDS-resistant dimers and trimers. Staining: cbb.

A peace of the gel corresponding to the monomeric connexin band was excised and the protein was extracted from the gel by electroelution, dialyzed against TBS buffer

with 0.2% SDS, concentrated 5 to 10 folds by ultrafiltration, and analyzed in SDS-PAGE. The separation in SDS-PAGE turned out to be very promising for connexin purification. Proteins with molecular weight significantly different from monomeric connexin, roughly above 35 kDa and below 15 kDa, were completely eliminated. Consequently, an appearance of the dimeric and trimeric forms of connexin clearly suggests their assembling from the isolated previously monomers. Significant unfolding of integral membrane proteins often leads to random oligomerization that eventually ends up with severe aggregation that is observed regardless of presented in electrophoresis buffers SDS. Note that no connexin aggregates of high molecular weight showed up in the purified and concentrated sample.

Still, proteins of the molecular weight close to the monomeric connexin remained in the prepared sample. This could be explained in part by broadening of the major connexin band due to overloading, in part by distortion of the band because of the overloading, and also by the low contrast of non-denaturing staining with KCl that considerable complicated accurate excision of the band corresponding solely to connexin. A preliminary enrichment of the reaction mixture could improve the electrophoretic purification of connexin.

3.3.10 Enrichment of the *in-vitro* produced Cx26 by phase separation.

A phase segregation of a detergent solution induced by temperature change or by selected additives has been described as an effective approach for isolation of the detergent - solubilized integral membrane proteins from the mixtures with soluble proteins (155, 156). The simplest suggested segregation trigger, a temperature shift, is unlikely applicable in the case of connexin solubilized in Brij78 as it would require a temperature that connexin cannot withstand (157, 158). Among other well-known phase-separation inducers are PEG and ammonium sulfate (AS) (159).

We tested salting out of Brij78-solubilized connexin from the *in-vitro* reaction mixture with ammonium sulfate (AS). Besides 1% Brij78, the reaction mixture contains 2% PEG 8000 that is susceptible to ammonium sulfate too (160) and also can influence the connexin precipitation with AS. Moreover, the ability of polyoxyethylene glycol and detergents of polyoxyethylene alkyl ether series C_xE_y to form complexes (161) can

additionally contribute to the phase separation induced by ammonium sulfate in a cell-free reaction mixture.

Since, as sections 3.3.5-3.3.6 describes, we did not find the expected aid of hexahistidine tag in purification, we have prepared a plasmid for cell-free expression that encodes full length connexin with no any N- or C- terminal extensions, hCx26wt226.

We produced hCx26wt226 as described in section 3.3.4. To find the appropriate ammonium sulfate concentration we took aliquots of reaction mixture and added with a vigorous mixing the saturated at 4°C (or, saturated by 95% at 20°C) solution of ammonium sulfate to make final AS concentration in samples equal to 20%, 25%, 30%, 40%, 50%, 60%, and 70% of saturation. After 1 hour incubation the produced suspensions were clarified by 16 000 *g* centrifugation and pellets were analyzed by SDS-PAGE (Figure 3.3.10, a).
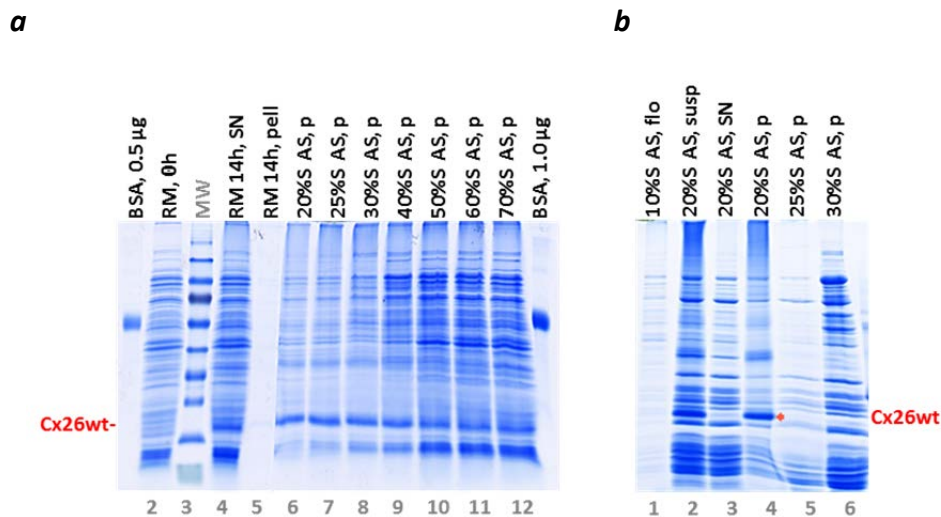

Figure 3.3.10 SDS-PAGE analysis of fractions obtained in course of enrichment by ammonium sulfate precipitation of connexin 26 produced *in vitro* form the reaction mixture. **a.** AS concentration screening. Lanes 1 and 13, staining intensity reference, 0.5µg and 1.0µg BSA, respectively. Lane 2, RM prior to incubation. Lane 3, MWST. Lane 4, RM soluble fraction. Lane 5, RM insoluble fraction. Lanes 6-12, 20%S to 70%S AS precipitates as indicated by the labels above the lanes. **b.** Incremental AS precipitation. 'susp', AS precipitate suspension prior to clarification by centrifugation. 'SN', supernatant for the indicated AS concentrations. 'p', precipitate for the AS concentration indicated by the labels above the lanes. Staining with cbb.

The lowest AS concentration used, 20% of saturation (20%S), had already precipitated significant fraction of the synthesized connexin. The further increase of the AS concentration did not add much connexin to the fraction precipitated at 20%S

and up to ~30%S did not cause considerable precipitation of contaminants. The contaminants started to precipitate significantly at AS concentration above ~40%S.

Since the minimal concentration of AS that is sufficient for quantitative and possible selective connexin precipitation was not found in this experiment, we set up step-wise incremental precipitation as follows. The minimal AS concentration added to the 16 000 g supernatant of a fresh cell-free reaction mixture was 10%S, the obtained suspension with 10%S AS was separated for pellet and supernatant by 20 000 g centrifugation, and the supernatant was supplemented with an additional portion of 95%S AS solution to make a final AS concentration of 20%S. After the appropriate incubation the suspension was separated for 20%S AS pellet and 20%S AS supernatant by centrifugation again. Similarly, 25%S AS and 30%S AS pellets were obtained from 20%S AS and 25%S AS supernatants, respectively. The samples obtained were analyzed in SDS-PAGE (Figure 3.3.10, b). In this test, made in volumes of 0.5 to 1 mL, the segregation of the material salted out with ammonium sulfate for precipitated and floating fractions became clear. Both fractions demonstrated very similar band pattern in SDS-PAGE (not shown) and were combined here.

The incremental ammonium sulfate precipitation clearly demonstrated that a significant enrichment of an *in-vitro* reaction mixture could be readily achieved prior to electrophoretic purification. A minor fraction of contaminants can be precipitated with 10%S AS (Figure 3.3.10, b, Lane 1) and connexin is nearly quantitatively and rather selectively can be precipitated at 20%S AS. This permits discarding with the 20%S AS supernatant most of irrelevant proteins constituting an *in vitro* reaction mixture. What also favors downstream electrophoretic purification is that free of connexin Brij78 from the reaction mixture also remains in the 20%S AS supernatant (Figure 3.3.10, b, Lane 5). Brij78 is responsible for the distortion in SDS-PAGE right in the region where connexin runs (can be seen in Lane 5, 25%S AS) that significantly complicated accurate excision of connexin band from the gel after electrophoresis (section 3.3.8). The connexin band in the 20%S AS precipitate is much sharper comparing to the crude reaction mixture and sufficiently free of closely located contaminants.

3.3.11. Further improvement of the enrichment protocol of the *in-vitro* produced Cx26.

While the achieved enrichment was already significant, we expected somewhat higher selectivity in connexin isolation by induced phase-separation taking into account previously reported results (159). Indeed, since connexin, as an integral membrane protein, was expected to be one of the most hydrophobic proteins in the *in vitro* reaction mixture, it should behave notably different from the other proteins in the considered phase-separation. The contaminants observed in Figure 3.3.10, b, Lane 4 could represent as those proteins that precipitated under direct effect of ammonium sulfate and those trapped by non-specific interactions with Brij-connexin complexes segregating from the reaction mixture. We tried to circumvent both undesired effects by supplementing the reaction mixture with glycerol as follows.
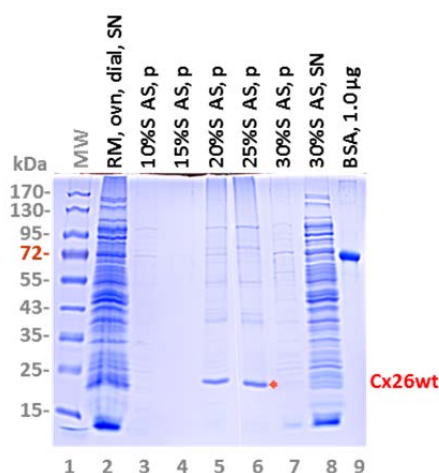

Figure 3.3.11 SDS-PAGE analysis of the fractions obtained in course of improving selectivity of connexin isolation by phase separation. Lane 1, MWST. Lane 2, *in vitro* reaction mixture dialyzed against 20% glycerol and clarified by centrifugation. Lanes 3 to 7, precipitates corresponding to AS concentrations indicated by the labels above the lanes. Lane 8, RM components that remained soluble in 30%S AS in the presence of 20% glycerol. Lane 9, staining intensity reference. Staining with cbb.

After completing the connexin *in vitro* synthesis as described in section 3.3.4 we dialyzed the reaction mixture against buffered 20% glycerol and then against the same glycerol buffer supplemented with 10%S AS and then with 15%S AS. The buffer equilibrated reaction mixture was clarified by 20000 *g* centrifugation and the phase separation was induced by adjusting AS concentration in the supernatant to 20%S and

63

then to 25%S. Centrifugation of the suspensions resulted in precipitated and floating insoluble fractions constituted of essentially all the connexin initially presented in the reaction mixture. Precipitated and floating fractions were combined and dissolved in 0.2% SDS. The samples obtained were analyzed in SDS-PAGE presented in the figure 3.3.11.

The addition of glycerol indeed improved hCx26wt226 purity in the 25%S AS precipitate and what is especially useful for electrophoretic purification it nearly eliminated contaminants that migrate close to connexin in SDS-PAGE.

3.3.12 Electrophoretic purification of the *in-vitro* produced Cx26 and enriched by phase separation.

We produced hCx26wt226 as described in section 3.3.4. After connexin enrichment by phase separation as described in 3.3.11 the precipitated connexin was extensively washed with low-salt 20% glycerol buffer with no detergent and then dissolved in the same buffer supplemented with 0.2% SDS. This sample is further referred to as *crudely purified by phase separation* (Figure 3.3.12, a, Lane 2). The oligomerization state of the sample was analyzed by size-exclusion chromatography and transmission electron microscopy and is described in the sections 3.4.1 and 3.4.4. The sample was loaded onto preparative scale SDS-PAGE. We used modified buffer system as described in methods section 4.3.12 to maximize the yield of the monomeric form by alleviating connexin self-association and aggregation. Specifically, comparing to the typical Laemmli buffer system (72), pH of the concentrating gel was increased to 7.5, the acrylamide concentration in the concentrating gel was raised to 6%, and the SDS concentration was raised to 0.2% in all electrophoretic buffers. After the separation was carried out as it was described in section 3.3.7 (9), the gel was stained with Zn^{++}•imidazole following a modified protocol of (162) and monomeric connexin band was excised (figure 3.3.12.b, Lane 2). The excised peace of the gel was placed in a dialysis tube filled with a 10-fold excess of the dialysis buffer containing 0.2% SDS and connexin diffused out of the gel during 24-hour dialysis against 100-fold excess of the dialysis buffer. The protein was collected from the dialysis tube, concentrated, analyzed in SDS-PAGE (figure 3.3.12.a), and used for reconstitution and oligomerization state analyses (Section 3.4).
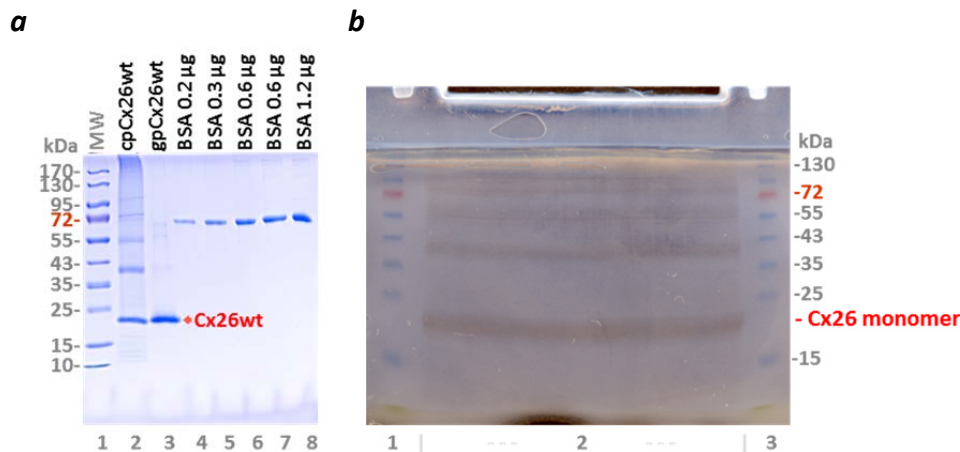
Figure 3.3.12 SDS-PAGE analysis (*a,* cbb staining) of the fractions obtained in result of crude purification of connexin 26 produced *in vitro* and electrophoretic purification (*b*) of its monomeric form. *a*. Lane 1, MWST. Lane 2, crudely purified by phase separation connexin 26. Lane 3, concentrated eluate of the gel-purified monomeric form. *b*. 'Preparative SDS-PAGE' developed with $Zn^{++} \cdot$ imidazole. Lanes 1 and 3, MWST. Lane 2, approximately 1 mg of crudely purified connexin 26.

As figure 3.3.12.b demonstrates, the preliminary cell free reaction mix enrichment (lane 2) indeed significantly reduced in-gel distortions and made possible accurate excision of the gel fragment with connexin monomer what resulted in pure connexin preparation free of contaminants or stable aggregates (figure 3.3.12.a, lane 3). Despite impossible contamination of connexin monomer with a connexin dimer or trimer due to their reliable separation in the preparative gel (figure 3.3.12.b), bands corresponding to connexin dimers and trimer showed up in lane 3 suggesting that the purified monomer underwent minor oligomerization. Concentrating of the monomer preparation solubilized in SDS up to 1 mg/mL did not increase significantly the fraction of oligomers, suggesting that the observed minor oligomerization might be an analysis artefact. In opposite, the fraction of oligomers observed in cell-free reaction mixture (figure 3.3.6) or in the SDS-solubilized 20%S AS precipitate (figure 3.3.12.a, lane 2) is much higher comparing to the purified monomer sample and these oligomers are rather stable; neither 100 mM DTT nor 4% SDS treatment caused these oligomers dissociation. This insensitivity of the oligomerization pattern suggests that the observed oligomerization pattern is unlikely representing any dynamic equilibrium between oligomerization states which turn one into another. Also, high stability of these oligomers is unlike to the stability of hexamers purified from mammalian (163)

or insect (66, 164) cells and raises the doubts on their functional relevance. Therefore, oligomers emerging in cell-free synthesis should be distinguished from those self-assembled from the purified monomers and probably discarded.

The selection of monomers is considered preferable (165) for the downstream crystallization since, comparing to oligomers, monomers provide more interfaces for establishing the intermolecular contacts network that is favored by a particular crystallization conditions thereby increasing the crystallization probability. On the other side, a particular purified oligomers are often more attractive when those are shown to be stable and functional.

Thus, using a detergent-supplemented cell-free expression system we have produced pure connexin 26 free of any N- or C- terminal extensions. The obtained samples showed up as monomers under conditions of SDS-PAGE. However, under these conditions, a stoichiometry of protein complexes could be influenced by the friction caused by the migration through the dense gel matrix. Further, we analyzed a size-homogeneity of the preparations using less destructive techniques.

3.4 Functional analysis

Self-assembling into hexamers is one of the intrinsic connexin properties shared by all the family members and an essential structural feature that is prerequisite for functional channel activity. We have examined an oligomerization state of the connexin samples that we have produced *in vitro* by size-exclusion chromatography (SEC), dynamic light scattering (DLS), and transmission electron microscopy (TEM).

3.4.1 Size exclusion chromatography

In live mammalian cells connexin protomers remain essentially monomeric in Golgi membranes and assemble in hexamers after migration into plasma membrane. However, on overexpression the monomers assemble into hexameric hemichannels or sometimes even into dodecameric full channels already in Golgi network, soon after being synthesized (166, 167). It was therefore interesting to check an oligomerization state of cell-free produced connexin 26 in crude reaction mixture prior to purification.

a.

Superose 6 GL 10/300
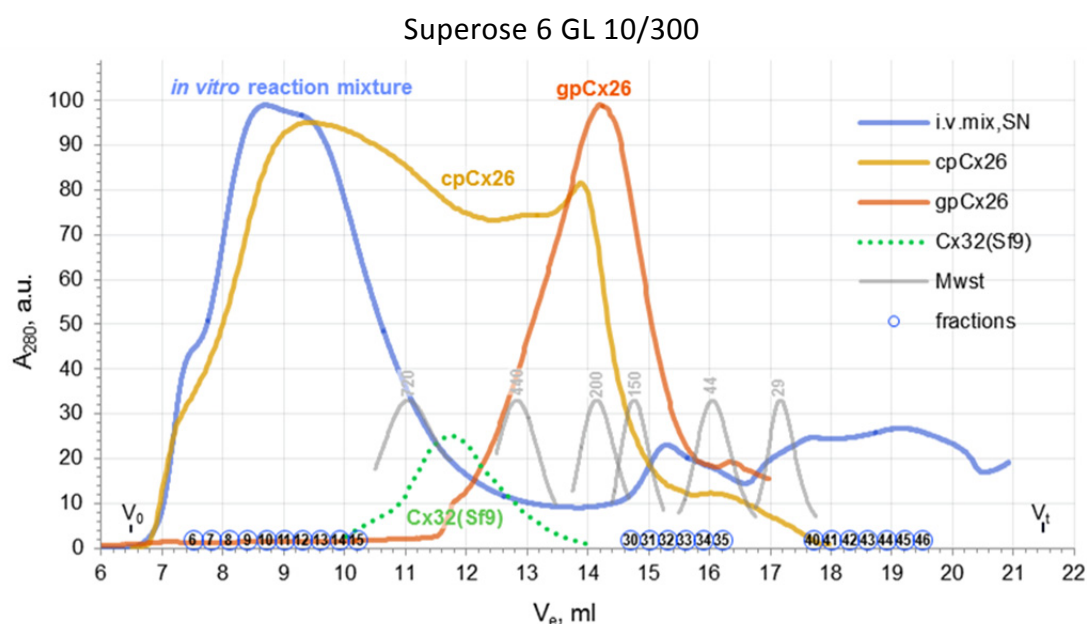
b.
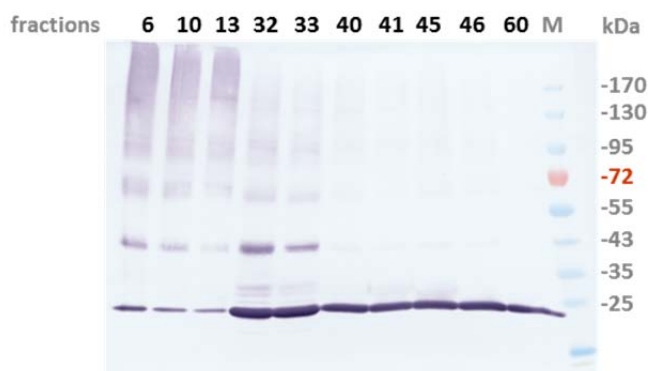
Figure 3.4.1. Size-exclusion chromatography analysis of connexin oligomerization state in non-purified (blue, '*i.v*.mix'), crudely purified by phase separation (orange, 'cpCx26'), and polyacrylamide gel-purified (red, 'gpCx26') samples of connexin 26 produced by *in vitro* expression. (a), elution profiles were recorded with UV-light absorbance flow-through detector at 280 nm and arbitrary scaled for convenience of the comparison. Grey lines represent elution traces of molecular weight standards, green dotted line corresponds to peak II of connexin 32 purified from insect cells as presented in (46). V_0 and V_t point to void and total solvent accessible column volumes, respectively. ①, numbers in blue circles correspond to eluted 0.3 mL fractions. *b.* selected fractions analyzed by anti-His-tag immunostaining.

A fraction of 14-h reaction mixture with connexin solubilized in Brij78 was analyzed by size-exclusion chromatography on Superose-6 column (Figure 3.4.1, *a*, blue line). Selected fractions corresponding to elution peaks were analyzed by western-blotting with anti-His$_6$ antibodies (Figure 3.4.1, *b*). Similarly, a fraction of SDS-solubilized 20%S

AS precipitate (figure 3.3.12.a, lane 2) and polyacrylamide gel-purified 'monomeric' (figure 3.3.12.a, lane 3) preparations were analyzed by SEC (Figure 3.4.1, *a*, orange and red lines correspondingly). An elution trace obtained by Stauffer K. *et al* (46) on Superose-6 resin with connexin 32 connexons purified from insect cells and solubilized in DDM is also given for reference.

In the cell-free reaction mixture large particles above 20 nm dominate and demonstrate faint immunopositivity (see Appendix A6 for the correlation of elution volume Ve and particles size Rs). These particles significantly exceed single connexons in size and presumably consist of aggregated connexins in addition to initial *in vitro* reaction mixture constituents. The highest connexin immunostaining was observed in the overlapping peaks with an elution volume of 14.8 to 16.5 mL approximately corresponding to detergent complexes of connexin 26 monomers or low order oligomers (see appendixes 7 and 8 for calibration curves). Surprisingly, a significant amount of connexin eluted much later (18 to ~24 mL) than expected from its molecular weight indicating nonspecific interaction with the chromatographic media. Together with a very large elution volume it might suggests connexin binding with and dissociation from the resin or partial aggregation and re-solubilization in course of chromatographic separation. We have concluded that Brij alone was not sufficient to maintain connexin in solution as monomers and prevent its aggregation effectively.

In the crudely purified and solubilized in SDS sample of connexin 26 a distinguishable peak at elution volume of 13.88 mL corresponding to 11.9 nm particles appeared. These particles are close in size to connexin 26 hexamers (168) and corresponding fractions were selected for transmission electron microscopy analysis (section 3.4.4). However, higher molecular weight species still dominate in the sample making it barely suitable for crystallization despite the high connexin content.

The polyacrylamide gel-purified sample that appeared as a monomer in the SDS-PAGE showed up as 11.6 nm particles with the elution volume of 14.23 mL in size-exclusion chromatography that is also close to 10.8 nm connexin 26 hexamers solubilized in DDM (168). The 14.23 mL elution peak clearly corresponded to a highly purified protein however it remained much broader at half-height (*hhw* = 1.99 mL) comparing to the globular proteins of similar size, like β-amylase (224 kDa, Rs = 5.9

nm, *hhw* = 0.73 mL) or alcohol dehydrogenase (147 kDa, Rs = 4.5 nm, *hhw* = 0.59 ml). The observed broadening could be explained by a minor protein unfolding. A promising approach to complete an integral membrane protein folding by reconstitution into amphiphilic polymers was developed by J.-L. Popot et.al. (169).

3.4.2 Reconstitution in the amphiphilic polymer A835

One of the initial steps of the detergent-mediated unfolding of alpha-helical membrane multi-pass proteins is assumed to be the detergent intercalation inside the helices bunch and disturbing the network of inter-helices interaction that stabilize the protein as a whole (170). Increasing the size of hydrophobic and hydrophilic moieties of the detergent molecule was suggested to address this harming effect of low molecular weight detergents (67, 171, 172). Amphiphilic polymers, or *amphipols*, represent one of the best developed implementation of the idea (169). Briefly, the use of amphipols for trapping of membrane proteins is presented in figure 3.4.2 below.
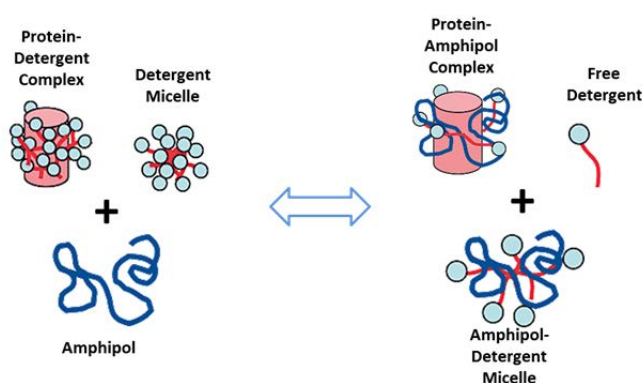
Figure 3.4.2. Stabilization of an integral membrane protein in water solution by substituting the low molecular weight detergent in the protein hydrophobic belt with amphipols. Adapted from (170).

A general applicability of the described approach has been demonstrated with a number of integral membrane proteins including GPCRs (82, 133, 173–178).

We used amphipol A835 to trap gel-purified connexin 26 solubilized in SDS. To favor the detergent replacement dodecyl sulfate was precipitated stepwise with 3-fold potassium excess (Figure 3.4.2.a) following to (179). The sample obtained was further analyzed and purified from the amphipol excess by size-exclusion chromatography (Figure 3.4.2.b, red line). For the comparison, the SDS-solubilized crudely purified by phase separation connexin 26 sample was treated similarly (orange line).

a.

b.

Superose 6 GL 10/300
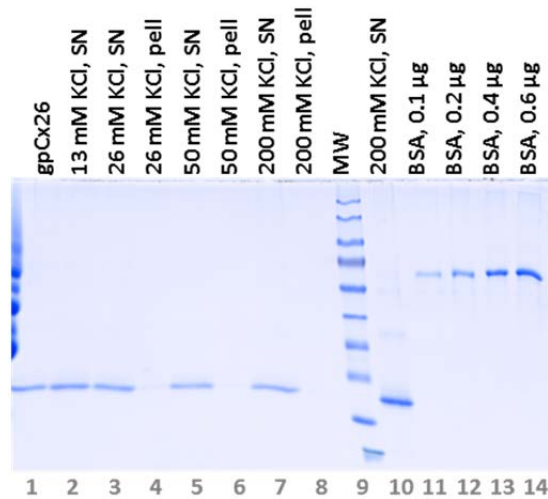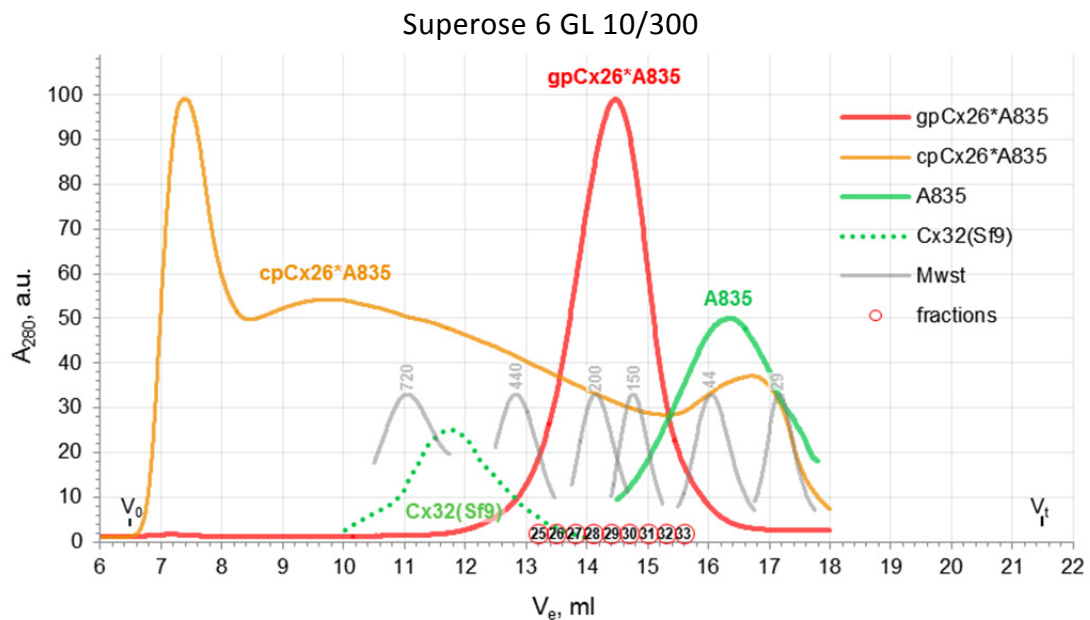
Figure 3.4.2. *a.* Trapping connexin 26 with amphipol A835. Lane 1, connexin-SDS-A835 mixture prior to PDS precipitation. Lanes 2 to 8 and 10, samples taken in course of gradual dodecyl sulfate precipitation by KCl. Lane 9, MWST. Lanes 11-14, cbb staining intensity references used for the connexin 26 concentration determination by gel densitometry with ImageJ. Staining with cbb. ***b.*** Connexin oligomerization state in samples reconstituted in A835 and analyzed by size-exclusion chromatography on Superose6HR. Elution traces were recorded as absorbance at 280 nm for protein samples and at 220 nm for amphipol. UV-light absorption values were arbitrary adjusted for individual curves for the comparison convenience. Connexin 26 trapped by A835 from the sample purified by SDS-PAGE (red line) or from the crudely purified by phase separation (orange line), grey lines represent elution profiles of molecular weight standards (appendix 6), green dotted line corresponds to the peak II of connexin 32 purified from insect cells as presented in (46). V_0 and V_t point to void and total solvent accessible column volumes, respectively.

The dodecyl sulfate depletion in the reconstitution mixture caused no any protein loss by precipitation (Fig. 3.4.2.a, lanes 4,6,8) and the reconstitution procedure resulted in the soluble connexin preparation (Fig. 3.4.2.a, lane 5,7). When amphipol was omitted in a similar test, the dodecyl sulfate depletion caused connexin to precipitate completely (not shown). Size-exclusion chromatography revealed no any aggregates in the amphipol-reconstituted connexin (Fig. 3.4.2.b, red line). The complex of connexin 26 with the amphipol, Cx26*A835, eluted at a volume corresponding to 10.5 nm particles as it was determined by the comparison with a set of samples with hydrodynamic radii measured by dynamic light scattering (appendix 6). The determined average particle size correlates very well with the hydrodynamic diameter of 10.8 nm determined by dynamic light scattering for the hexameric connexin hemichannels purified in DDM from insect cell. The low deviation from the average size of the Cx26*A835 complex can be described by the peak width at a half the height (*hhw*) of 1.3 mL that is only 20% greater than the *hhw* value of the remarkably compact bacteriorhodopsin from *H. salinarum* in complex with A835 (1.1 ml, appendix 8). Although peaks of both amphipol complexes are 50% to 80% broader relative to the peaks of detergent-free and compact b-amylase (0.73 ml) or bacteriorhodopsin in DDM (0.78 ml), the inhomogeneity of the amphipol itself (*hhw* = 2.25 ml) can be used to explain the observed peak broadening.

A fraction with V_e = 14.5 mL representing the Cx26*A835 complex was selected for the analysis by transmission electron microscopy that will be described in section 3.4.4.

Was the high purity of the connexin sample essential for the successful reconstitution described above? Indeed, if all the intermediate oligomeric forms that are observed in the crudely purified by phase separation connexin 26 merely represented a dynamic equilibrium and readily interconverted one into another then trapping with amphipol could drive the equilibrium toward the hemichannels and make the sample size-homogenous, provided that the hexameric assembly is thermodynamically favored. This would be very beneficial in increasing the protein purification yield. Orange line in Figure 3.4.2.b demonstrates that amphipol A835 was ineffective in driving the hemichannels domination in the crudely purified connexin

sample. In opposite, it mostly induced an aggregation (compare orange lines in figures 3.4.2.b and 3.4.1.a) supporting the assumption made in section (3.3.12) that the connexin 26 oligomeric forms that are produced in the cell free expression system consist of inappropriately bound protomers which cannot participate in a productive oligomerization resulting in functional hexameric assemblies.

3.4.3 Connexin assemblies stability.

Until the very recent advance, the crystallization of the amphipol-trapped membrane proteins was a matter of deep concerns (174). The preparations in maltose- and phosphocholine- derived low molecular weight detergents still contribute substantially to the body of high-resolution structure data of integral membrane proteins (80). Therefore, we tried to replace the detergents in the obtained preparations of connexin 26 for DDM or FC12 taking also into account the proven stability of the connexin 26 hemichannels obtained from insect cells in these detergents (157). According to Tribet C. *et al* (170) membrane protein-bound amphipol can be competitively replaced by an excessive low molecular weight detergent present in a solution.

We incubated overnight with 1% DDM the samples of gpCx26*A835 produced as described in section 3.4.2 and analyzed by size-exclusion chromatography on Superose 6 GL 10/300 column equilibrated with 0.1% buffered DDM solution. Alternatively, by dodecyl sulfate precipitation similar to the described above A835 trapping, we replaced SDS with DDM or FC12 in the gel-purified connexin. Because crystallization conditions require highly concentrated protein sample and concentrating of the protein preparations is also associated with a significant elevation of the free detergent micelle concentration, we also examined, after the amphipol substitution, a stability of the connexin assemblies during concentration by ultrafiltration. The chromatography analysis results are presented in Figure 3.4.3.

The elution maximum and hence the average particle size in the major peak of connexin in FC12 (green solid line) were nearly equal to those of connexin in A835, and this peak was a little narrower (*hhw* = 1.2 mL). In opposite, transfer into DDM resulted in the significant right shift of the elution maximum to V_e = 15.20 mL

corresponding to the average particle size of 8.84 nm (or roughly to 96 kDa) and in the significant peak broadening (hhw = 2.04 ml) indicating notable size-inhomogeneity.
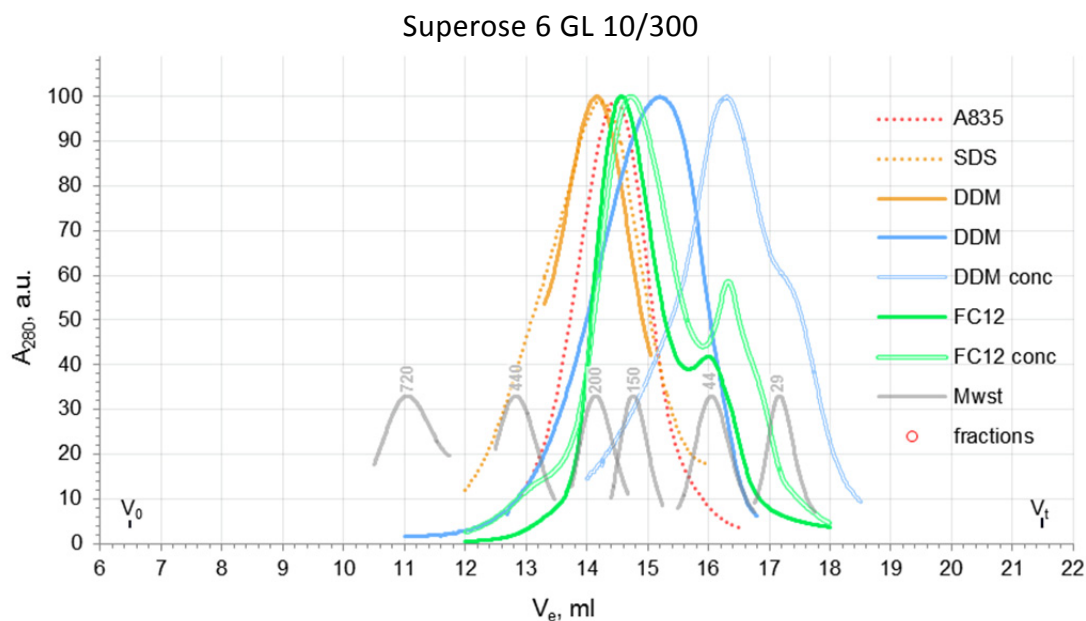
Superose 6 GL 10/300

Figure 3.4.3 Detergent exchange in connexin 26 samples analyzed by size-exclusion chromatography on Superose 6 GL 10/300 column. Dotted lines representing samples of polyacrylamide gel-purified (orange) and amphipol A835-trapped (red) connexin 26 that were used for the detergent replacement has already been given in previous figures and are overlaid here for the reference. Orange solid line, direct replacement of SDS for DDM. Blue solid line, A835 exchange for DDM. Green solid line, SDS exchange for FC12. Blue double line, the sample with A835 substituted for DDM was 10-fold concentrated and examined in DDM-supplemented buffer again. Green double line, the sample with SDS substituted for FC12 was 10-fold concentrated and examined in FC12-supplemented buffer again. V_0 and V_t point to void and total solvent accessible column volumes, respectively. Elution profiles were recorded with UV-light absorbance flow-through detector at 280 nm and arbitrary scaled for the comparison convenience.

The effect of reducing an average particle size developed further in the concentrated samples (double lines) indicating that both tested detergents caused decomposition of the connexin assemblies, and the effect was more prominent in DDM. In opposite, the replacement of A835 for DDM in Cx32 and Cx43t263 (104, 105) as well as ultrafiltration of A835-trapped connexins did not cause a hexamer disassembly. It is also noteworthy that major elution peak of connexin transferred into DDM directly from SDS solution, bypassing the amphipol reconstitution, matched the elution peak of the initial sample in SDS, indicating possibly incomplete detergent substitution. Indeed, in a detergent mixture of SDS and OG with no protein the

potassium dodecyl sulfate solubility was significantly increased (not shown) what can be explained by a formation of mixed micelles (180).

Obtained here connexin samples in DDM and FC12 were selected for TEM observations (section 3.4.4) from the fractions corresponding to the elution maxima.

Thus, trapping with amphipol A835 of the connexin 26 produced in vitro and purified by phase separation followed by the preparative polyacrylamide gel electrophoresis resulted in the preparation closely resembling connexin 26 hemichannels purified from the insect cells producent. The described approach was also shown to be efficient in stabilization of connexin 32 and Q263-truncated fragment of connexin 43 (in collaboration with O. Volkov and M. Silacheva, respectively).

3.4.4 Transmission electron microscopy

The transmission electron microscopy (TEM) made crucial contribution in the connexin structure determination. A large body of TEM data is available that represents how do connexins appear in TEM depending on details of the purification protocols or engineered mutations. That provides an opportunity to compare with the published images the samples that we have produced.

We used samples described in sections 3.4.1-3.4.4 to prepare negatively stained specimens and analyzed them with JEOL 1200-EX II electron microscope. Representative images are given in Figure 3.4.4.

The average particle size of 9 nm observed in the DDM-solubilized sample (*c*) is close to that one observed by EM for connexins purified from insect cells (38, 46). The sample in DDM was also remarkably homogenous. Particles in amphipol (*b*) were a little larger in average and surprisingly less homogenous disagreeing with SEC analysis. While particles of typical for connexin hexamers size were present in SDS-solubilized sample (*a*) much larger clots of various size still dominated. SDS-solubilized and A835-trapped connexin samples resembled connexin preparations isolated from rat liver and solubilized in OG (46). The sample in FC12 (*d*) was heavily fibrillated. Particles, similar in appearance to the specimens obtained from mammalian or insect cells, were found only in the sample solubilized in DDM. Making such a comparison, one should keep in

mind that the recognized doughnut shape of connexin hemichannels is observed in TEM in the very specially prepared samples (see the legend to Figure 2.9 for details). Also, single point mutations significantly impaired the classical doughnut microscopic appearance of DDM-solubilized connexin hemichannels yet those connexins still were able to self-assemble into hexamers and to cluster in plasma membrane into plaques suggesting essentially proper protein folding at least in a lipid environment (38).
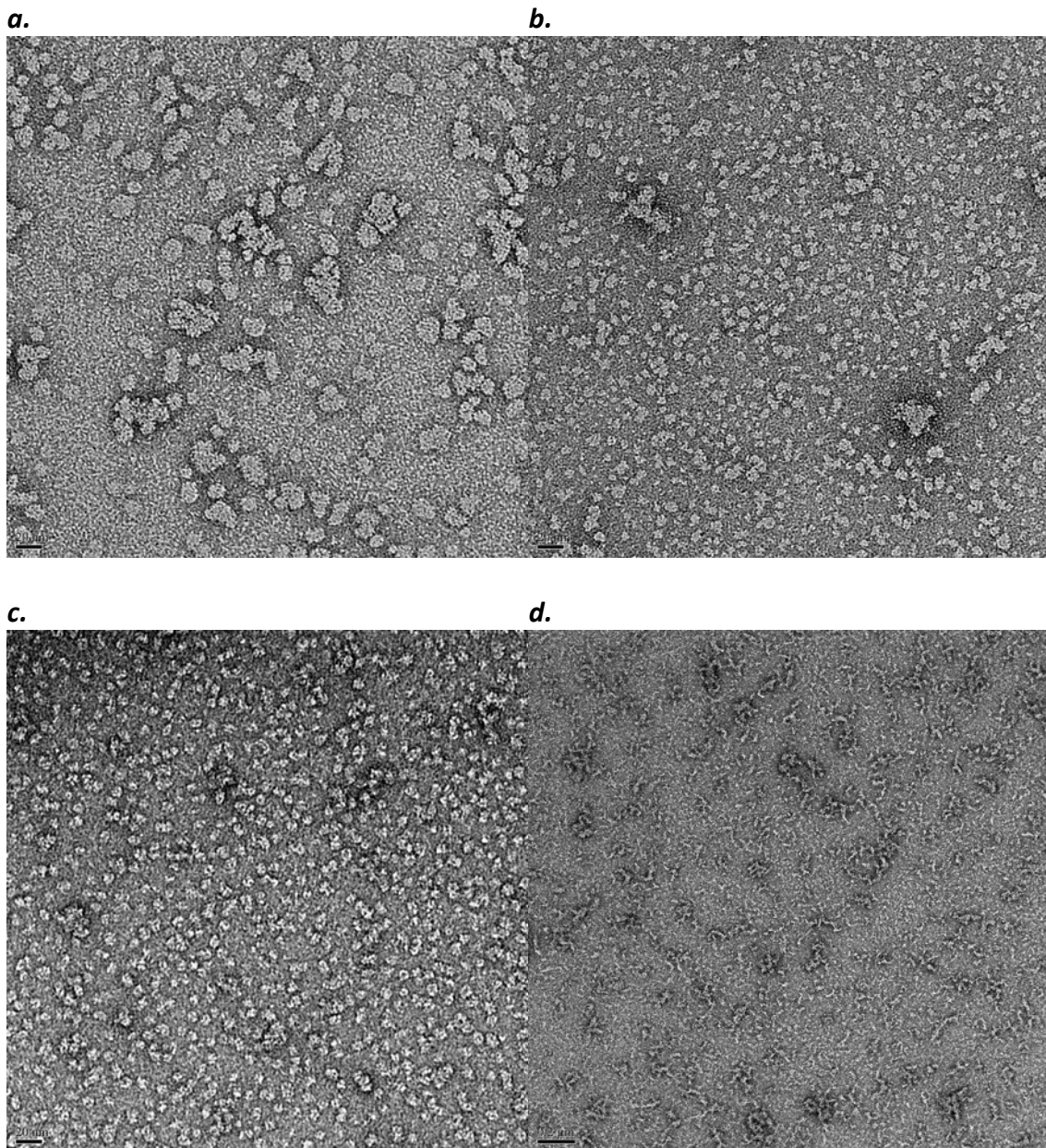

Figure 3.4.4 Transmission electron microscopic analysis of the *in-vitro* produced connexin 26 samples in *a*) SDS; *b*) A835; *c*) DDM; *d*) FC12; bar: 20 nm in *a-c*, 0.2μm in *d*.

3.4.5 Liposome permeability assay

While self-association into hexamer is essential it is not sufficient for connexin protomers to form functional solute-permeable channel. There were several approaches established to test isolated connexin channels permeability (59, 101, 181, 182). The rather convenient one consists in reconstituting of connexin hemichannels into small unilamellar liposomes and observing the exchange of the membrane-impermeable fluorescent dye between the liposomes interior and exterior. When liposomes suspended in a dye-free solution with membrane-incorporated connexin hemichannels in the closed state were preliminary loaded with a dye at a concentration causing the fluorescence quenching then the channel opening caused the dye release, dilution, unquenching, and jump in the integral fluorescence (101). In opposite to the conventional transport specific fractionation (TSF) this approach allows real-time observation of fluorescent intensity changes in response to channel gating.

To implement the described approach we prepared POPC liposomes loaded with Lucifer Yellow by extrusion through the 100 nm CA membrane. The connexin 26 solubilized in SDS was produced *in vitro* as described above and reconstituted into the liposomes by dilution below the detergent CMC. The chosen protein to lipid ratio was estimated to result in one hemichannel per liposome in average. The dye that was not trapped inside the liposomes was removed by dialysis. To maintain the connexin channels in a closed state 0.5 mM Ca^{++} was present in the dialysis buffer. Liposomes prepared in the same manner but with no protein were used as negative control. After recording the quenched fluorescence level samples were supplemented with 100-fold molar excess of EDTA to open connexin hemichannels by Ca^{++} binding and changes in the fluorescence were detected. The obtained results are presented in figure 3.4.5.

The increase in fluorescence in response to Ca^{++} chelating in the sample of liposomes with the reconstituted connexin was 3.5-fold higher comparing to the one with no any protein suggesting Lucifer Yellow efflux through the opened connexin hemichannels and thereby confirming that the hexameric connexin 26 hemichannels that we have produced do respond to the connexin gating trigger.
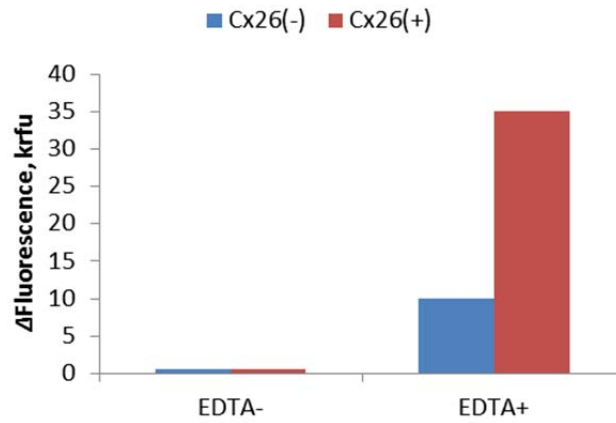
Figure 3.4.5 Liposome permeability assay. Connexin hemichannels were reconstituted in the presence of 0.5 mM Ca^{++} into POPC liposomes loaded with 1% LY-CH (red bars) and quenched LY-CH fluorescence was measured (EDTA-) in the 10-fold diluted liposome suspension. Then, the samples were supplemented with 5 mM EDTA and increase in the fluorescence intensity was detected (EDTA+). *In collaboration with Dr. C. Zeilinger, Leibniz University of Hannover*.

4. Materials and methods

4.1 Instruments and material vendors

Instrument	Vendor
Incubator Incu-Line	VWR (Darmstadt, Germany)
Multitron, Multitron-2, Minitron orbital shakers, Benchtop 10-L bioreactor Labfors 4, 40-L bioreactor Techfors I	Infors HT (Bottmingen, Switzerland)
Mini-Sub Cell GT System	Bio-Rad (Hercules, USA)
Mini Trans-Blot Electrophoretic Transfer Cell	Bio-Rad (Hercules, USA)
Power supply EPS 301	GE Healthcare (Freiburg, Germany)
Transilluminator TFX-20M	Vilber Lourmat (Marne-la-Vallée, France)
InGenius Gel Documentation System	Synoptics Ltd. (Cambridge, UK)
Scanner Perfection V750 Pro	Epson (Tokio, Japan)
Arktik Thermal Cycler	Thermo Fisher Scientific Oy (Vantaa, Finland)
Thermomixer Comfort, Benchtop microcentrifuge 5417R	Eppendorf AG (Hamburg, Germany)
Benchtop multifunctional prep-centrifuge Heraeus Megafuge 16R	Thermo Fisher Scientific (Osterode, Germany)
Prep-centrifuge Avanti J26 XP, Prep-ultracentrifuge Optima L90K, Benchtop micro-ultracentrifuge Optima MAX Benchtop multifunctional prep-centrifuge	Beckman Coulter Inc. (Indianapolis, USA)
French Press	Sim Aminco (Rochester, USA)
Microfluidizer M-110P	Microfluidics (Newton, USA)
Ultrasonic Cell Disruptor 450 Digital	Branson Ultrasonics Co. (Monterrey, Mexico)
SE 250 Mighty Small II vertical electrophoresis unit	Hoefer Inc. (San-Francisco, USA)
Peristaltic pump SciQ 400	Watson-Marlow (Calgary, Canada)
ÄKTA Prime Plus, ÄKTA Pure liquid chromatography systems	GE Healthcare (Freiburg, Germany)
Roller Mixer SRT6D, Gyro-rocker SSL3, magnetic stirrers	Stuart (Stone, UK)

UV-Vis spectrophotometer UV-2450	Shimatzu (Kyoto, Japan)	
UV-Vis spectrophotometer Nanophotometer P300	Implen (Munich, Germany)	
Dyna Pro-E-20-660 Dynamic Light Scattering (DLS) detector	Proterion corp. (USA)	

E.coli strain	Genotype	Source
Top10	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-	Invitrogen
XL10Gold	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 TetR F' (proAB lacIqZΔM15 Tn10 (TetR Amy CmR))	Stratagene
SE1	F-, CmR, ompT, lon, hsdSB (restriction-, modification-), gal, dcm, DE3 (lacI, T7polymerase under the control of the PlacUV5 promoter), ccdB+.	Eurogentec S.A. (Liege, Belgium)
A19	*rna-19, gdhA2, his-95, relA1, spoT1, metB1*	
BL21 Star	F$^-$ *ompT gal dcm lon hsdS$_B$($r_B^-m_B^-$)*	New England BioLabs
BL21(DE3)	F$^-$ ompT gal dcm lon hsdSB(rB$^-$mB$^-$) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)	New England BioLabs

Plasmid DNA	Features	Source
pCR21K287	Synthetic codon-optimized CDS for human connexin 26	Eurofins MWG GmbH
pSCodon1.2	Dual-component stabilized expression plasmid with compensation for human gene codon bias	Eurogentec S.A. (Liege, Belgium)
pLIM14	A template for mistic PCR amplification	(86)
pRK603	Auxiliary plasmid for TEV protease co-expression, middle copy number (Addgene plasmid 8831)	Addgene, Inc. (107)
pKM586	Auxiliary plasmid for TEV protease co-expression, low copy number (Addgene plasmid 19978)	Addgene, Inc. (107)
pRK793	TEV protease expression vector (Addgene plasmid 8827)	Addgene, Inc. (109)
pIVEX2.3d	*In vitro* expression vector	5-Prime
pAR1219	T7 RNA polymerase expression vector	Sigma(183)

Full sequences with feature mappings of the plasmids or of gene of interest are provided in appendix A3.

Chemicals	Vendor
Inorganic compunds	Sigma-Aldrich, Applichem, Merck
Detergents	Affimetrix, Sigma-Aldrich

Lipids	Avanti Polar Lipids, Sigma-Aldrich
Components of cell free expression system	Sigma-Aldrich, Roth, Roche
DNA restriction/modification enzymes	Fermentas
Factor Xa	Qiagen
Enterokinase	New England Biolabs
Thrombin	Sigma-Aldrich

4.2 Solutions and media

Bacterial growth media were prepared according to (58).

Media	Base ingredients
LB	1% tryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.0, 100 mg/L ampicillin and/or 50 mg/L kanamycin when appropriate
LB-agar	1% tryptone, 1% yeast extract, 1.5% bacto-agar, 1% glucose, 0.5% NaCl, pH 7.0, 150 mg/L ampicillin and/or 50 mg/L kanamycin when appropriate
TB	1.2% tryptone, 2.4% yeast extract, 17 mM KH2PO4, 72 mM K2HPO4, 0.5% glycerol, 100 mg/L ampicillin and/or 100 mg/L kanamycin when appropriate
PCM	1% tryptone, 1% yeast extract, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, pH 6.8 2 mM Mg_2SO_4, 0.5 mM $CaCl_2$, 0.5% glucose, 100 mg/L ampicillin and/or 200 mg/L kanamycin when appropriate
AIM	1% tryptone, 1% yeast extract, 50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, pH 6.8 2 mM Mg_2SO_4, 0.5 mM $CaCl_2$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 100 mg/L ampicillin and/or 200 mg/L kanamycin when appropriate

4.3 Experimental procedures

4.3.1 Plasmid DNA manipulations

E. coli strains Top10 or XL10Gold were used for plasmid amplification and storage. Plasmid DNA or DNA fragments were purified on silica membranes with NucleoSpin kits from Macherey-Nagel (Dueren, Germany). Genes of interest were amplified by PCR and/or modified to generate constructs described in chapter 3 and listed in appendix 2 using Phusion Hot Start II DNA Polymerase, DNA restriction/modification

enzymes from Fermentas GmbH and following conventional protocols (184) or enzyme vendor instructions. Synthetic oligonucleotides or genes were purchased from Eurofins MWG GmbH. Nucleotide sequences of the generated constructs were verified by DNA sequencing at Eurofins MWG GmbH.

4.3.2 *E.coli* transformation and selection

Competent cells BL21 were prepared and transformed according to (185), XL10Gold and Top10 were done according to (186) and PIPES was replaced for HEPES. Positive clones were selected on LB-agar with appropriate antibiotic. When double transformation of SE1 cells was required for TEV protease co-expression it was done in two steps. First, commercial chemically competent SE1 cells were transformed with plasmids bearing connexin gene according to pStaby system manual (Eurogentec S.A., Liege, Belgium) and selected for ampicillin resistance on LB-agar plates. Second, the selected clones were transformed with auxiliary plasmid according to (185) and selected for double-resistance to ampicillin and kanamycin.

4.3.3 *E.coli* culture propagation

Baffled-bottom glass 0.15 to 2L Erlenmeyer flasks with liquid bacterial cultures were incubated in 25 or 50 mm orbital shaker adjusted between 120 to 180 rpm to provide the best aeration. XL10Gold and Top10 cultures were maintained in LB media at 37°C. Cultures of SE1 cells were propagated in PCM at 37°C when non-inducing condition were required and in AIM for induction. For autoinduction, AIM was inoculated with fresh late log-phase PCM cultures with OD_{600} of 4 to 6 AU_{600} to have initial OD_{600} of 0.2 to 0.4 AU_{600}. Cultures in AIM were incubated at 37°C until reaching OD_{600} of 1.0 AU_{600}, and then incubation temperature was reduced to 20°C unless different condition is specified. To measure optical densities of bacterial cultures tungsten-lamp photometer was used. Glucose consumption was monitored with colorimetric glucose test stripes (Merck, 1.17866.0001). In selected cases, to propagate the bacterial cultures 10-L or 40-L bioreactors were used. After 12 to 20 hour cultivation, AIM cultures were harvested at 4°C 5000 *g* for 30 min. Collected cells were frozen in liquid nitrogen or used immediately for isolation of cell lipid membranes.

4.3.4 Subcellular fractionation

All procedures were carried out on ice-water bath or in a cold room and using preliminary refrigerated at +4°C solutions and equipment. For cell lysis 7% glycerol supplemented with 20 mM TrisHCl pH 7.5, 2 mM 6-AHA, 1 mM PMSF, 2 mM TCEP, 1 mM EGTA, protease inhibitors cocktail Complete® was used. Cell pellets were resuspended immediately after harvesting or thawed with a vigorous stirring in a lysis buffer in a 1:4 (w/v) ratio, 80 mL of the buffer per 20 g of cell pellet was typically used. The suspension was further homogenized with Dounce homogenizer. Lyophilized lysozyme from chicken egg white, 1/1000 of wet cell pellet weight, and bovine pancreatic DNase I, 1/10000 of wet cell pellet weight, were dissolved in fresh buffer supplemented with 2 mM $MgCl_2$, 1/100 of the cell suspension volume, and added to the suspension with a vigorous stirring. After 20 min, cells were disrupted by 2 passages through the ice-cold French Pressure Cell at a pressure of 90 MPa and the suspension was supplemented with 0.1 M NaCl, 10 mM EDTA pH 7.5. Unbroken cells and other microscopic contaminants were removed by 10 min centrifugation at 1000 g. The supernatant was distributed in ultracentrifuge tubes, underlaid with 1-2 mL of buffered 40% glycerol cushion and insoluble cell components were isolated by centrifugation for 2 hours at 100 000 g. Supernatant was carefully aspirated and the precipitated slurry was resuspended in TBS buffer for the next fractionation step of isopycnic density gradient ultracentrifugation. Depending on their capacity the ultracentrifuge tubes were filled with 7 to 10 mm layers in 5% decrementing steps of sucrose or 20% steps of glycerol up to minimal concentration of 20-30% then the isolated insoluble cell fraction sample was overlaid. The solutions making up layers were buffered with 10 mM TrisHCl, 10 mM EDTA, and 2 mM 6-AHA at pH 7.5. The separation was continued for 14 to 18 hours at 100 000 g. The separate layers were collected by aspiration generally in five serial fractions: washing buffer, inner membranes, interphase, outer membranes, and heavy aggregates. When reasonable separation was not achieved the inner and outer membrane fractions were diluted to reduce glycerol or sucrose content below 20% and independently separated once again as described above. In the last case, the corresponding fractions from independent samples of inner and outer membranes were combined. Resulting inner

and outer membrane fractions were diluted when it was necessary with TBS buffer to adjust glycerol concentration to 40% and kept at -20°C or processed further.

4.3.5 Membrane solubilization

All procedures were carried out in a cold room and using preliminary refrigerated at +4°C solutions and equipment. For solubilization 10 mM TrisHCl, 0.1 M NaCl, 1 mM TCEP, 2 mM 6-AHA, pH 8.0 was used routinely. Prior to solubilization cell membrane fractions were equilibrated with the solubilization buffer by dialysis or dilution followed by ultracentrifugation at 100 000 g for 1 hour and resuspension. Final membrane suspension density was adjusted referring to the amount of cells yielding the membrane sample. Membrane sample obtained from 10 g of cells were homogenized in 90 mL of buffer. Then, 10- or 20- fold solution of a detergent of choice was added with vigorous stirring that was continued for 16 hours more. Insoluble material was separated by 1 hour 100 000 g ultracentrifugation and supernatant was purified further by chromatography.

4.3.6 Co- and Ni-NTA chromatography

Immobilized metal ion affinity chromatography was carried out following the manufacturer recommendations (187) in 10 mM TrisHCl, 0.1 M NaCl, 1 mM TCEP, 2 mM 6-AHA, pH 7.5 was used as a basic buffer with detergent and imidazole or EDTA adjusted as particular experiment required. Typically, 25 mM imidazole was used to wash out contaminants and 0.25 M one for elution. After elution, protein fractions were supplemented with 5 mM EDTA and 2 mM DTT end equilibrated with a downstream analysis buffer by gel-filtration.

4.3.7 Size-exclusion chromatography

Superose 6 GL 10/300 column was developed in 10 mM TrisHCl, 0.1 M NaCl, 1 mM DTT, 2 mM 6-AHA, 2 mM EDTA, pH 7.5 at 0.1 or 0.2 ml/min with Äkta Prime Plus chromatographic system. When appropriate, the running buffer was supplemented with a detergent. UV absorbance traces were recorded at 280 nm using 20 nm band-pass filter and 2 mm path length flow-through cell. Protein molecular weight standards were prepared as 10 mg/mL stock solutions in buffered 50% glycerol and stored at -20°C. The column was calibrated regularly. The positions of maxima for the elution

peaks were determined by fitting the elution trace with Gaussian curve (188). The calibration data are given in appendixes 6 to 8. Samples for a size-distribution analysis were injected in the same sample volume as samples of MW standards.

4.3.8 TEV protease expression and purification

The autoinactivation-resistant mutant S219V of TEV protease (109) for the *in-vitro* cleavage experiments was expressed in *E.coli* BL21(DE3) harboring pRK793 and purified on NiNTA essentially as discussed in (189). The purified enzyme was stored at -20°C in 10 mM TrisHCl, 0.1 M NaCl, 2 mM DTT, 2 mM EDTA, 50% glycerol, 0.1% Triton X-100. Activity of the produced enzyme was verified in comparison with commercial ProTEV Plus Protease (Promega).

4.3.9 *In-vitro* proteolysis

Purified protein samples in detergent micellar solution or reconstituted into lipid vesicles were equilibrated with a cleavage buffer by dialysis in 14 kDa MWCO dialysis tubing or 12.5 kDa MWCO micro-dialysis devices (Scinova GmbH, Germany) before the cleavage tests. For TEV protease, 20 mM HEPES, 1 mM EDTA, 1 mM DTT was used as a cleavage buffer base and 20 mM TrisHCl, 50 mM NaCl, 1 mM $CaCl_2$ was the buffer base for enterokinase. Incubation time, reaction temperature, pH level, detergent supplement, and enzyme concentration varied as described in the results section 3.1.1.10.

4.3.10 *In vitro* expression

Plasmid DNA for *in-vitro* expression was purified using Plasmid Plus Midi Kit from Qiagen (Hilden, Germany). T7 RNA polymerase was expressed and purified according to (139). Continuous exchange protocol had been employed as explained in (132, 139). Analytical scale reactions for detergent screening and Mg^{++} adjustments were set up in 0.05 mL custom dialysis devises and preparative reactions were done in 2 mL scale as described in (135). Pipetting protocols are summarized in appendix 5. After the *in-vitro* synthesis reaction mixtures were kept in the same dialysis device that was used for the synthesis and equilibrated with 10 mM TrisHCl, 0.1 M NaCl, 1 mM DTT, 2 mM 6-AHA, 2 mM EDTA, 20% glycerol, 0.5% brij78, pH 7.5 by shaking at + 10°C with two dialysis buffer exchanges every 30 min and then with 14%S AS in the same buffer by the same

way. The mixtures were clarified by 30 min centrifugation at 70 000 *g* and crude connexin preparation was isolated from the supernatant by phase separation induced by 24%S AS.

4.3.11 Crude connexin purification by phase separation

Samples were incubated with ammonium sulfate with gentle rocking for 1 hour minimum or overnight at 4°C when AS concentration was below 20%S, otherwise they were incubated at 20°C. The pH level of the saturated solution of ammonium sulfate was adjusted at 7.2. Phases were separated by 1 hour centrifugation at 70 000 *g*. Liquid phase was aspirated via ⌀0.26 mm Hamilton needle.

4.3.12 Electrophoretic purification

A 'preparative scale' SDS-PAGE was casted using Hoefer gel caster, 8 × 10 cm rectangular glass plates with notched aluminum oxide plates and custom 5 mm spacers and comb resulting in one 5 × 50 × 10 mm sample pocket and two 1.5 × 1.5 × 10 mm pockets on flanks for MW standards. Stacking 6% acrylamide gel was made 5-7 mm long and buffered with 0.2 M TrisHCl pH 7.5. Apart from the elevated to 0.2% SDS concentration, electrophoretic buffers corresponded to Laemmli system (72).

4.3.13 Extraction of the purified protein from PAGE by diffusion

After electrophoretic purification, excised gel pieces were placed into 14MWCO dialysis tube filled with 10 mM TrisHCl pH 7.5, 100 mM NaCl, 2 mM DTT, 2 mM 6-AHA, 1 mM EDTA and dialyzed overnight against the same buffer. Volume ratio of the excised gel pieces and the buffer inside the dialysis tube was 1:10 and the ratio of the buffer inside and outside the dialysis tube was 1:100.

4.3.14 Electroelution

Electroelution of the gel-purified protein from the excised gel pieces was done essentially similar to (190). The gel pieces were placed into and fixed on one side of a dialysis tube with cut off size of 3.5 kDa filled with Tris-glycine electrophoresis running buffer. The dialysis tube was immersed in the electrode buffer in a chamber for horizontal agarose electrophoresis and electric field of 10 V/cm was applied to elute

protein into dialysis bag from the gel for 30 min. The connexin solution was then collected from the dialysis bag.

4.3.15 Protein sample preparation for analytical SDS-PAGE and Western blotting.

We used modified loading buffer containing 10% glycerol instead of sugar as a thickener and 4% SDS instead of 2% reported by Laemmli (72). For SDS-PAGE or western blotting analysis proteins were precipitated from samples containing high level of detergent, lipids, or salts. For the precipitation samples were mixed with equal volume of acetone, then 100% w/v TCA was added to the mixture to make up 10% final TCA concentration. Samples were incubated 1 hour to overnight at -20°C and centrifuged 10 min at 16 000 g and 0°C. Supernatant was carefully removed and pellets were resuspended in 0.5-1 mL of pure acetone with gentle sonication with ⌀3 mm microtip (5-10 seconds at 10% power settings in 450W sonicator, with the actual delivered power of 11W). Washed thereby protein precipitate was collected by 10 min centrifugation at 16 000 g and 0°C. The obtained pellet was washed once again with aceton and dissolved in 1x loading buffer avoiding complete acetone evaporation and with gentle sonication when it was necessary.

Samples of bacterial cultures were prepared as follows. Bacteria were precipitated for 3 min at 8000 g from aliquots of culture with integral OD_{600} of 2 AU*ml, i.e. a sample of 0.667 mL was taken from a culture with OD_{600} of 3 AU to prepare the sample for SDS-PAGE. Supernatants were carefully removed and pellets were resuspended in 0.16 mL of water. Then, 0.04 mL of the 5-fold loading buffer were added, samples were briefly sonicated as above and kept frozen till analysis. Final density of the samples prepared this way was 10 AU*mL referring to the bacterial culture originating the samples and this value was used in densitometric calculations of protein yield.

4.3.16 Transmission electron microscopy

Transmission electron microscopy measurements were carried out in the electron microscopy laboratory of Institute of Structural Biology, Grenoble, France with the kind assistance of Dr. Evelina Edelweiss. Samples were adsorbed on the clean side of a carbon film formed on mica (the carbon-mica interface) and negatively stained in 1%

uranyl acetate. Images were captured under low-dose conditions with a JEOL 1200 EXII electron microscope operating at 100 kV at a nominal magnification of 40000×.

4.3.17 Liposome permeability assay

POPC liposomes were prepared in 10 mM TrisHCl pH 7.5 100 mM NaCl. A dialysis for the removal of excess of Lucifer Yellow after the connexin reconstitution was continued for 24 hours with three buffer refreshments against the same buffer supplemented with 0.5 mM Ca^{++}. Fluorescence measurements were done in duplicates with the kind assistance of Dr. Carsten Zeilinger in 0.050 mL aliquots using Berthold Technologies Mithras plate reader with excitation wavelength set to 435 nm and fluorescence emission recorded at 535 nm.

5. Conclusions

- Human connexin 26 in fusion with *B.Subtilis* Mistic protein have been overexpressed in *E.coli* with over 1 mg per 1L culture yields.

- The mistic-connexin fusion proteins separated by different proteases recognition sites have been purified to homogeneity and different cleavage conditions for mistic removal have been explored. The purified detergent-solubilized mistic-connexin fusion proteins demonstrated resistance to the specific proteolytic cleavage.

- An *in vivo* mistic cleavage approach employing co-expression of TEV protease encoded by helper plasmids has been evaluated. The TEV protease activity was controlled by the helper plasmid copy number. The helper plasmid with the *p15* replication origin provided sufficient protease for the complete mistic removal in small-scale but not large scale cultures.

- A new alternative expression tag, ESR, and its fragments have been evaluated to address possible topological limitations for cleavage tag accessibility.

- An *in-vitro E.coli* expression of Cx26 have been tested and yielded 1 mg of Cx26 from 1 mL of the reaction mixture. An original approach for purification of Cx26 produced *in vitro* with no any tag has been developed. The purified protein self-assembled into hexamers. The assembling was facilitated in polymeric detergent, amphipol A8-35.

- After reconstitution into small unilamelar liposomes the purified protein responded to Ca^{++} level by switching between closed and opened state.

Acknowledgements

Bibliography

1. N. S. McNutt, R. S. Weinstein, Carcinoma of the cervix: deficiency of nexus intercellular junctions., *Science* **165**, 597–9 (1969).

2. J. P. Revel, M. J. Karnovsky, Hexagonal array of subunits in intercellular junctions of the mouse heart and liver., *J. Cell Biol.* **33**, C7–C12 (1967).

3. R. Devis, D. W. James, Close Association Between Adult Guinea-Pig Fibroblasts In Tissue Culture, Studied With The Electron Microscope., *J. Anat.* **98**, 63–8 (1964).

4. Y. Kanno, W. R. Loewenstein, Intercellular diffusion., *Science* **143**, 959–60 (1964).

5. A. Jamakosmanović, W. R. Loewenstein, Intercellular communication and tissue growth. III. Thyroid cancer., *J. Cell Biol.* **38**, 556–61 (1968).

6. W. R. Loewenstein, R. D. Penn, Intercellular communication and tissue growth. II. Tissue regeneration., *J. Cell Biol.* **33**, 235–42 (1967).

7. W. R. Loewenstein, Y. Kanno, Intercellular communication and tissue growth. I. Cancerous growth., *J. Cell Biol.* **33**, 225–34 (1967).

8. W. R. Loewenstein, M. Nakas, S. J. Socolar, Junctional membrane uncoupling. Permeability transformations at a cell membrane junction., *J. Gen. Physiol.* **50**, 1865–91 (1967).

9. Y. Kanno, W. R. Loewenstein, Low-resistance Coupling between Gland Cells. Some Observations on Intercellular Contact Membranes and Intercellular Space, *Nature* **201**, 194–195 (1964).

10. R. Ashman, Y. Kanno, W. Loewenstein, Intercellular electrical coupling at a forming membrane junction in a dividing cell, *Science* **145**, 604–605 (1964).

11. J. D. Sheridan, Electrophysiological study of special connections between cells in the early chick embryo., *J. Cell Biol.* **31**, C1–5 (1966).

12. D. D. Potter, E. J. Furshpan, E. S. Lennox, Connections between cells of the developing squid as revealed by electrophysiological methods., *Proc. Natl. Acad. Sci. U.S.A.* **55**, 328–36 (1966).

13. B. W. Payton, M. V. Bennett, G. D. Pappas, Permeability and structure of junctional membranes at an electrotonic synapse., *Science* **166**, 1641–3 (1969).

14. M. S. Nielsen *et al.*, Gap junctions., *Compr Physiol* **2**, 1981–2035 (2012).

15. D. A. Goodenough, J. A. Goliger, D. L. Paul, Connexins, connexons, and intercellular communication., *Annu. Rev. Biochem.* **65**, 475–502 (1996).

16. A. L. Harris, Emerging issues of connexin channels: biophysics fills the gap., *Q. Rev. Biophys.* **34**, 325–472 (2001).

17. G. E. Sosinsky, G. M. Gaietta, B. N. Giepmans, in *Connexins: A Guide*, Harris A., Locke D., Ed. (Springer, 2009), pp. 241–261.

18. L. S. Musil, in *Connexins: A Guide*, Harris A., Locke D., Ed. (Springer, 2009), pp. 225–240.

19. G. E. Sosinsky, B. J. Nicholson, Structural organization of gap junction channels., *Biochim. Biophys. Acta* **1711**, 99–125 (2005).

20. D. A. Goodenough, D. L. Paul, Gap junctions., *Cold Spring Harb Perspect Biol* **1**, a002576 (2009).

21. G. Söhl, K. Willecke, An update on connexin genes and their nomenclature in mouse and man., *Cell Commun. Adhes.* **10**, 173–80.

22. T. U. Consortium, Activities at the Universal Protein Resource (UniProt)., *Nucleic Acids Res.* **42**, D191–8 (2014).

23. S. Maeda *et al.*, Structure of the connexin 26 gap junction channel at 3.5 Å resolution, *Nature* **458**, 597–602 (2009).

24. S. Maeda, T. Tsukihara, Structure of the gap junction channel and its implications for its biological functions., *Cell. Mol. Life Sci.* **68**, 1115–29 (2011).

25. A. Oshima, Structure and closure of connexin gap junction channels., *FEBS Lett.* **588**, 1230–7 (2014).

26. O. Jara *et al.*, Critical role of the first transmembrane domain of Cx26 in regulating oligomerization and function., *Mol. Biol. Cell* **23**, 3299–311 (2012).

27. J. C. Sáez, M. A. Retamal, D. Basilio, F. F. Bukauskas, M. V. L. Bennett, Connexin-based gap junction hemichannels: gating mechanisms., *Biochim. Biophys. Acta* **1711**, 215–24 (2005).

28. F. F. Bukauskas, V. K. Verselis, Gap junction channel gating., *Biochim. Biophys. Acta* **1662**, 42–60 (2004).

29. C. I. Foote, L. Zhou, X. Zhu, B. J. Nicholson, The pattern of disulfide linkages in the extracellular loop regions of connexin 32 suggests a model for the docking interface of gap junctions., *J. Cell Biol.* **140**, 1187–97 (1998).

30. P. L. Sorgen *et al.*, Structural changes in the carboxyl terminus of the gap junction protein connexin43 indicates signaling between binding domains for c-Src and zonula occludens-1., *J. Biol. Chem.* **279**, 54695–701 (2004).

31. P. L. Sorgen, H. S. Duffy, D. C. Spray, M. Delmar, pH-dependent dimerization of the carboxyl terminal domain of Cx43., *Biophys. J.* **87**, 574–81 (2004).

32. L. N. Axelsen, K. Calloe, N.-H. Holstein-Rathlou, M. S. Nielsen, Managing the complexity of communication: regulation of gap junctions by post-translational modification., *Front Pharmacol* **4**, 130 (2013).

33. A. Saidi Brikci-Nigassa *et al.*, Phosphorylation controls the interaction of the connexin43 C-terminal domain with tubulin and microtubules., *Biochemistry* **51**, 4331–42 (2012).

34. A. Oshima, K. Tani, Y. Hiroaki, Y. Fujiyoshi, G. E. Sosinsky, Projection structure of a N-terminal deletion mutant of connexin 26 channel with decreased central pore density., *Cell Commun. Adhes.* **15**, 85–93 (2008).

35. D. Henderson, H. Eibl, K. Weber, Structure and biochemistry of mouse hepatic gap junctions., *J. Mol. Biol.* **132**, 193–218 (1979).

36. A. Oshima, K. Tani, Y. Hiroaki, Y. Fujiyoshi, G. E. Sosinsky, Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10034–9 (2007).

37. J. J. Lacapère, Membrane Protein Structure Determination: Methods and Protocols, *Methods in Molecular Biology* **654**, 1–459 (2010).

38. A. Oshima, T. Doi, K. Mitsuoka, S. Maeda, Y. Fujiyoshi, Roles of Met-34, Cys-64, and Arg-75 in the assembly of human connexin 26. Implication for key amino acid residues for channel formation and function., *J. Biol. Chem.* **278**, 1807–16 (2003).

39. C. Ambrosi *et al.*, Analysis of four connexin26 mutant gap junctions and hemichannels reveals variations in hexamer stability., *Biophys. J.* **98**, 1809–19 (2010).

40. A. Oshima *et al.*, Asymmetric configurations and N-terminal rearrangements in connexin26 gap junction channels., *J. Mol. Biol.* **405**, 724–35 (2011).

41. J. C. Saez, V. M. Berthoud, M. C. Branes, A. D. Martinez, E. C. Beyer, Plasma membrane channels formed by connexins: their regulation and functions., *Physiol. Rev.* **83**, 1359–400 (2003).

42. E. L. Hertzberg, A detergent-independent procedure for the isolation of gap junctions from rat liver., *J. Biol. Chem.* **259**, 9936–43 (1984).

43. G. E. Sosinsky, G. A. Perkins, Purification of gap junctions., *Methods Mol. Biol.* **154**, 57–75 (2001).

44. E. Gogol, N. Unwin, Organization of connexons in isolated rat liver gap junctions., *Biophys. J.* **54**, 105–12 (1988).

45. S. Ghoshroy, D. A. Goodenough, G. E. Sosinsky, Preparation, characterization, and structure of half gap junctional layers split with urea and EGTA., *J. Membr. Biol.* **146**, 15–28 (1995).

46. K. A. Stauffer, N. M. Kumar, N. B. Gilula, N. Unwin, Isolation and purification of gap junction channels., *J. Cell Biol.* **115**, 141–50 (1991).

47. T. Roosild, J. Greenwald, S. Choe, Compositions and methods for producing recombinant proteins, (2013).

48. M. E. Lundberg, E. C. Becker, S. Choe, MstX and a putative potassium channel facilitate biofilm formation in Bacillus subtilis., *PLoS ONE* **8**, e60993 (2013).

49. T. P. Roosild *et al.*, NMR structure of Mistic, a membrane-integrating protein for membrane protein expression., *Science* **307**, 1317–21 (2005).

50. H. Dvir, S. Choe, Bacterial expression of a eukaryotic membrane protein in fusion to various Mistic orthologs., *Protein Expr. Purif.* **68**, 28–33 (2009).

51. T. P. Roosild, M. Vega, S. Castronovo, S. Choe, Characterization of the family of Mistic homologues., *BMC Struct. Biol.* **6**, 10 (2006).

52. G. Kefala, W. Kwiatkowski, L. Esquivies, I. Maslennikov, S. Choe, Application of Mistic to improving the expression and membrane integration of histidine kinase receptors from Escherichia coli., *J. Struct. Funct. Genomics* **8**, 167–72 (2007).

53. A. Deniaud *et al.*, Expression of a chloroplast ATP/ADP transporter in E. coli membranes: behind the Mistic strategy., *Biochim. Biophys. Acta* **1808**, 2059–66 (2011).

54. O. V. Nekrasova *et al.*, A new hybrid protein for production of recombinant bacteriorhodopsin in Escherichia coli., *J. Biotechnol.* **147**, 145–50 (2010).

55. E. N. Lyukmanova *et al.*, N-terminal fusion tags for effective production of g-protein-coupled receptors in bacterial cell-free systems., *Acta Naturae* **4**, 58–64 (2012).

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

57. C. Y. Szpirer, M. C. Milinkovitch, Separate-component-stabilization system for protein and DNA production without the use of antibiotics., *BioTechniques* **38**, 775–81 (2005).

58. F. W. Studier, Protein production by auto-induction in high density shaking cultures., *Protein Expr. Purif.* **41**, 207–34 (2005).

59. M. C. Fiori *et al.*, Permeation of calcium through purified connexin 26 hemichannels., *J. Biol. Chem.* (2012), doi:10.1074/jbc.M112.383281.

60. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis., *Nat. Methods* **9**, 671–5 (2012).

61. A. Rath, F. Cunningham, C. M. Deber, Acrylamide concentration determines the direction and magnitude of helical membrane protein gel shifts., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15668–73 (2013).

62. X. Bao, L. Reuss, G. A. Altenberg, Regulation of purified and reconstituted connexin 43 hemichannels by protein kinase C-mediated phosphorylation of Serine 368., *J. Biol. Chem.* **279**, 20058–66 (2004).

63. E. L. Hertzberg, N. B. Gilula, Isolation and characterization of gap junctions from rat liver., *J. Biol. Chem.* **254**, 2138–47 (1979).

64. A. L. Harris, A. Walter, D. Paul, D. A. Goodenough, J. Zimmerberg, Ion channels in single bilayers induced by rat connexin32., *Brain Res. Mol. Brain Res.* **15**, 269–80 (1992).

65. I. V. Koreen, W. A. Elsayed, Y. J. Liu, A. L. Harris, Tetracycline-regulated expression enables purification and functional analysis of recombinant connexin channels from mammalian cells., *Biochem. J.* **383**, 111–9 (2004).

66. O. Gassmann *et al.*, The M34A mutant of Connexin26 reveals active conductance states in pore-suspending membranes., *J. Struct. Biol.* **168**, 168–76 (2009).

67. S. C. Lee *et al.*, Steroid-based facial amphiphiles for stabilization and crystallization of membrane proteins., *Proc. Natl. Acad. Sci. U.S.A.* **110**, E1203–11 (2013).

68. C. K. Manjunath, G. E. Goings, E. Page, Detergent sensitivity and splitting of isolated liver gap junctions., *J. Membr. Biol.* **78**, 147–55 (1984).

69. J. A. Diez, S. Ahmad, W. H. Evans, Assembly of heteromeric connexons in guinea-pig liver en route to the Golgi apparatus, plasma membrane and gap junctions., *Eur. J. Biochem.* **262**, 142–8 (1999).

70. C. Zeilinger, M. Steffens, H.-A. Kolb, Length of C-terminus of rCx46 influences oligomerization and hemichannel properties., *Biochim. Biophys. Acta* **1720**, 35–43 (2005).

71. H. Dvir, M. E. Lundberg, S. K. Maji, R. Riek, S. Choe, Mistic: cellular localization, solution behavior, polymerization, and fibril formation., *Protein Sci.* **18**, 1564–70 (2009).

72. U. K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4., *Nature* **227**, 680–5 (1970).

73. M. Thein, G. Sauer, N. Paramasivam, I. Grin, D. Linke, Efficient subfractionation of gram-negative bacteria for proteomics studies., *J. Proteome Res.* **9**, 6135–47 (2010).

74. J. H. Weiner, L. Li, Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis., *Biochim. Biophys. Acta* **1778**, 1698–713 (2008).

75. Z. Yu *et al.*, Role for Escherichia coli YidD in membrane protein insertion., *J. Bacteriol.* **193**, 5242–51 (2011).

76. C. Filip, G. Fletcher, J. L. Wulff, C. F. Earhart, Solubilization of the cytoplasmic membrane of Escherichia coli by the ionic detergent sodium-lauryl sarcosinate., *J. Bacteriol.* **115**, 717–22 (1973).

77. R. I. Hobb, J. A. Fields, C. M. Burns, S. A. Thompson, Evaluation of procedures for outer membrane isolation from Campylobacter jejuni., *Microbiology (Reading, Engl.)* **155**, 979–88 (2009).

78. I. Chopra, S. W. Shales, Comparison of the polypeptide composition of Escherichia coli outer membranes prepared by two methods., *J. Bacteriol.* **144**, 425–7 (1980).

79. J. M. Vergis, M. C. Wiener, *Detergent Compatibility Assay: A Simple Method for Exploring Detergent Space* (Charlottesville, VA 22908-0736, 2009).

80. P. Raman, V. Cherezov, M. Caffrey, The Membrane Protein Data Bank., *Cell. Mol. Life Sci.* **63**, 36–51 (2006).

81. M. Suga, S. Maeda, S. Nakagawa, E. Yamashita, T. Tsukihara, A description of the structural determination procedures of a gap junction channel at 3.5 A resolution., *Acta Crystallogr. D Biol. Crystallogr.* **65**, 758–66 (2009).

82. I. Mus-Veteau, Membrane Proteins Production for Structural Analysis, (2014).

83. V. Beswick *et al.*, Dodecylphosphocholine micelles as a membrane-like environment: new results from NMR relaxation and paramagnetic relaxation enhancement analysis., *Eur. Biophys. J.* **28**, 48–58 (1999).

84. H. Xia, L. Liu, C. Reinhart, H. Michel, Heterologous expression of human Neuromedin U receptor 1 and its subsequent solubilization and purification., *Biochim. Biophys. Acta* **1778**, 2203–9 (2008).

85. X. Wu *et al.*, A novel method for high-level production of TEV protease by superfolder GFP tag., *J. Biomed. Biotechnol.* **2009**, 591923 (2009).

86. M. Noirclerc-Savoye, B. Gallet, F. Bernaudat, T. Vernet, Large scale purification of linear plasmid DNA for efficient high throughput cloning., *Biotechnol J* **5**, 978–85 (2010).

87. A. Skerra, I. Pfitzinger, A. Plueckthun, The functional expression of antibody Fv fragments in Escherichia coli: Improved vectors and a generally applicable purification technique, *Biotechnology* **9**, 273–278 (1991).

88. E. Hochuli, W. Bannwarth, H. Döbeli, R. Gentz, D. Stüber, Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent, *Nature Biotechnology* **6**, 1321–1325 (1988).

89. M. R. Mehlenbacher, F. Bou-Abdallah, X. X. Liu, A. Melman, Calorimetric studies of ternary complexes of Ni (II) and Cu (II) nitrilotriacetic acid and N-acetyloligohistidines, *Inorganica Chimica Acta* **437**, 152–158 (2015).

90. F. Angius, O. Ilioaia, M. Uzan, B. Miroux, in *Heterologous Expression of Membrane Proteins: Methods and Protocols*, Mus-Veteau, Isabelle, Ed. (Springer New York, New York, NY, 2016), pp. 37–52.

91. E. J. Lugtenberg, R. Peters, Distribution of lipids in cytoplasmic and outer membranes of Escherichia coli K12., *Biochim. Biophys. Acta* **441**, 38–47 (1976).

92. R. P. Huijbregts, de K. AI, de K. B, Topology and transport of membrane lipids in bacteria., *Biochim. Biophys. Acta* **1469**, 43–61 (2000).

93. I. Shibuya, Metabolic regulations and biological functions of phospholipids in Escherichia coli., *Prog. Lipid Res.* **31**, 245–99 (1992).

94. D. P. Ricci, T. J. Silhavy, The Bam machine: a molecular cooper., *Biochim. Biophys. Acta* **1818**, 1067–84 (2012).

95. I. Yamato, Y. Anraku, K. Hirosawa, Cytoplasmic membrane vesicles of Escherichia coli. A simple method for preparing the cytoplasmic and outer membranes., *J. Biochem.* **77**, 705–18 (1975).

96. P. Overath, M. Brenner, T. Gulik-Krzywicki, E. Shechter, L. Letellier, Lipid phase transitions in cytoplasmic and outer membranes of Escherichia coli., *Biochim. Biophys. Acta* **389**, 358–69 (1975).

97. G. Taylor, M. Hoare, D. Gray, F. Marston, Size and density of protein inclusion bodies, *Nature Biotechnology* **4**, 553–557 (1986).

98. N.-S. Cheng, Formula for the viscosity of a glycerol-water mixture, *Industrial \& engineering chemistry research* **47**, 3285–3288 (2008).

99. N. Pandey *et al.*, Screening and identification of genetic loci involved in producing more/denser inclusion bodies in Escherichia coli., *Microb. Cell Fact.* **12**, 43 (2013).

100. J. M. Vergis, M. C. Wiener, The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal., *Protein Expr. Purif.* **78**, 139–42 (2011).

101. M. Steffens *et al.*, Regulation of connexons composed of human connexin26 (hCx26) by temperature., *Biochim. Biophys. Acta* **1778**, 1206–12 (2008).

102. A. K. Mohanty, C. R. Simmons, M. C. Wiener, Inhibition of tobacco etch virus protease activity by detergents., *Protein Expr. Purif.* **27**, 109–14 (2003).

103. A.-K. Lundbäck, S. van den Berg, H. Hebert, H. Berglund, S. Eshaghi, Exploring the activity of tobacco etch virus protease in detergent solutions., *Anal. Biochem.* **382**, 69–71 (2008).

104. V. Oleksandr, Human connexin 32 overexpression for 3D crystallization, *RWTH Aachen University* (2013).

105. S. Maria, Expression hétérologue de la connexine humaine 43 dans Escherichia coli, *École doctorale physique, Grenoble* (2014).

106. J. D. Fox, D. S. Waugh, Maltose-binding protein as a solubility enhancer., *Methods Mol. Biol.* **205**, 99–117 (2003).

107. R. B. Kapust, D. S. Waugh, Controlled intracellular processing of fusion proteins by TEV protease., *Protein Expr. Purif.* **19**, 312–8 (2000).

108. R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements., *Nucleic Acids Res.* **25**, 1203–10 (1997).

109. R. B. Kapust *et al.*, Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency., *Protein Eng.* **14**, 993–1000 (2001).

110. T. G. M. Schmidt, A. Skerra, The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins., *Nat Protoc* **2**, 1528–35 (2007).

111. R. B. Kapust, J. Tözsér, T. D. Copeland, D. S. Waugh, The P1' specificity of tobacco etch virus protease., *Biochem. Biophys. Res. Commun.* **294**, 949–55 (2002).

112. J. Phan *et al.*, Structural basis for the substrate specificity of tobacco etch virus protease., *J. Biol. Chem.* **277**, 50564–72 (2002).

113. C. Renicke, R. Spadaccini, C. Taxis, A tobacco etch virus protease with increased substrate tolerance at the P1' position., *PLoS ONE* **8**, e67915 (2013).

114. D. Locke, F. Kieken, L. Tao, P. L. Sorgen, A. L. Harris, Mechanism for modulation of gating of connexin26-containing channels by taurine., *J. Gen.*

Physiol. **138**, 321–39 (2011).

115. R. J. Jenny, K. G. Mann, R. L. Lundblad, A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa., *Protein Expr. Purif.* **31**, 1–11 (2003).

116. M. Kurz *et al.*, Incorporating a TEV cleavage site reduces the solubility of nine recombinant mouse proteins., *Protein Expr. Purif.* **50**, 68–73 (2006).

117. D. F. Rodrigues *et al.*, Architecture of thermal adaptation in an Exiguobacterium sibiricum strain isolated from 3 million year old permafrost: a genome and transcriptome approach., *BMC Genomics* **9**, 547 (2008).

118. L. E. Petrovskaya *et al.*, Predicted bacteriorhodopsin from Exiguobacterium sibiricum is a functional proton pump., *FEBS Lett.* **584**, 4193–6 (2010).

119. D. Bratanov, Expression, Purification, and Crystallization of Bacteriorhodopsin and Its Derivatives, *RWTH Aachen University* (2014).

120. S. Marullo, C. Delavier-Klutchko, Y. Eshdat, A. D. Strosberg, L. Emorine, Human beta 2-adrenergic receptors expressed in Escherichia coli membranes retain their pharmacological properties., *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7551–5 (1988).

121. M. P. Chapot *et al.*, Localization and characterization of three different beta-adrenergic receptors expressed in Escherichia coli., *Eur. J. Biochem.* **187**, 137–44 (1990).

122. R. M. Lacatena, A. Cellini, F. Scavizzi, G. P. Tocchini-Valentini, Topological analysis of the human beta 2-adrenergic receptor expressed in Escherichia coli., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10521–5 (1994).

123. S. Danyi *et al.*, Solubilisation and binding characteristics of a recombinant beta2-adrenergic receptor expressed in the membrane of Escherichia coli for the multianalyte detection of beta-agonists and antagonists residues in food-producing animals., *Anal. Chim. Acta* **589**, 159–65 (2007).

124. A. J. Link, G. Skretas, E.-M. Strauch, N. S. Chari, G. Georgiou, Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli., *Protein Sci.* **17**, 1857–63 (2008).

125. I. Gushchin *et al.*, Structural insights into the proton pumping by unusual proteorhodopsin from nonmarine bacteria., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12631–6 (2013).

126. M.-F. Hsu *et al.*, Using Haloarcula marismortui bacteriorhodopsin as a fusion tag for enhancing and visible expression of integral membrane proteins in Escherichia coli., *PLoS ONE* **8**, e56363 (2013).

127. M. M. Falk, L. K. Buehler, N. M. Kumar, N. B. Gilula, Cell-free synthesis and assembly of connexins into functional gap junction membrane channels., *EMBO J.* **16**, 2703–16 (1997).

128. M. M. Falk, Connexins/connexons. Cell-free expression., *Methods Mol. Biol.* **154**, 91–116 (2001).

129. T. A. Nguyen, S. S. Lieu, G. Chang, An Escherichia coli-based cell-free system for large-scale production of functional mammalian membrane proteins suitable for X-ray crystallography., *J. Mol. Microbiol. Biotechnol.* **18**, 85–91 (2010).

130. Y. Moritani, S. M. Nomura, I. Morita, K. Akiyoshi, Direct integration of cell-free-synthesized connexin-43 into liposomes and hemichannel formation., *FEBS J.* **277**, 3343–52 (2010).

131. F. Junge *et al.*, Advances in cell-free protein synthesis for the functional and structural analysis of membrane proteins., *N Biotechnol* **28**, 262–71 (2011).

132. B. Schneider *et al.*, Membrane protein expression in cell-free systems., *Methods Mol. Biol.* **601**, 165–86 (2010).

133. R. Sachse, S. K. Dondapati, S. F. Fenz, T. Schmidt, S. Kubick, Membrane protein synthesis in cell-free systems: from bio-mimetic systems to bio-membranes., *FEBS Lett.* **588**, 2774–81 (2014).

134. D. Schwarz, D. Daley, T. Beckhaus, V. Dötsch, F. Bernhard, Cell-free expression profiling of E. coli inner membrane proteins., *Proteomics* **10**, 1762–79 (2010).

135. S. Reckel *et al.*, Strategies for the cell-free expression of membrane proteins., *Methods Mol. Biol.* **607**, 187–212 (2010).

136. E. D. Carlson, R. Gan, C. E. Hodgman, M. C. Jewett, Cell-free protein synthesis: applications come of age., *Biotechnol. Adv.* **30**, 1185–94.

137. D. F. Savage, Cell-free protein synthesis: Methods and protocols, edited by Alexander S. Spirin and James R. Swartz., *Protein Science* **17**, 962–963 (2008).

138. A. S. Spirin, V. I. Baranov, L. A. Ryabova, S. Y. Ovodov, Y. B. Alakhov, A continuous cell-free translation system capable of producing polypeptides in high yield., *Science* **242**, 1162–4 (1988).

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

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

141. J. J. Wuu, J. R. Swartz, High yield cell-free production of integral membrane proteins without refolding or detergents., *Biochim. Biophys. Acta* **1778**, 1237–50 (2008).

142. A. Pedersen, K. Hellberg, J. Enberg, B. G. Karlsson, Rational improvement of cell-free protein synthesis., *N Biotechnol* **28**, 218–24 (2011).

143. F. Katzen *et al.*, Insertion of membrane proteins into discoidal membranes using a cell-free protein expression approach., *J. Proteome Res.* **7**, 3535–42 (2008).

144. K.-H. Park *et al.*, In the cauldron of cell-free synthesis of membrane proteins: playing with new surfactants., *N Biotechnol* **28**, 255–61 (2011).

145. C. Roos *et al.*, Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers., *Mol. Membr. Biol.* **30**, 75–89 (2013).

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

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

148. K. Corin, B. Cook, S. Zhang, A robust, rapid, and simple method of producing olfactory receptors using commercial *E. coli* cell-free systems., *Methods Mol. Biol.* **1003**, 101–8 (2013).

149. F. A. Junge, Biochemical and functional characterization of cell-free expressed G-protein coupled receptors of the human endothelin system, *der Goethe-Universität in Frankfurt am Main* (2010).

150. D. A. P. Gutmann *et al.*, A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay., *Protein Sci.* **16**, 1422–8 (2007).

151. B. Becher, J. Cassim, Improved isolation procedures for the purple membrane of Halobacterium halobium., *Preparative biochemistry* **5**, 161 (1975).

152. H. Seelert, F. Krause, Preparative isolation of protein complexes and other bioparticles by elution from polyacrylamide gels., *Electrophoresis* **29**, 2617–36 (2008).

153. I. Maslennikov *et al.*, Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10902–7 (2010).

154. R. Kalmbach *et al.*, Functional cell-free synthesis of a seven helix membrane protein: in situ insertion of bacteriorhodopsin into liposomes., *J. Mol. Biol.* **371**,

639–48 (2007).

155. U. Sivars, F. Tjerneld, Mechanisms of phase behaviour and protein partitioning in detergent/polymer aqueous two-phase systems for purification of integral membrane proteins., *Biochim. Biophys. Acta* **1474**, 133–46 (2000).

156. A. Sánchez-Ferrer, R. Bru, F. García-Carmona, Phase separation of biomolecules in polyoxyethylene glycol nonionic detergents., *Crit. Rev. Biochem. Mol. Biol.* **29**, 275–313 (1994).

157. K. A. Baker, E. Y. T. Chien, Y. Hua, M. Yeager, R. C. Stevens, *Expression Screening and Thermal Stability of Human Connexin 26* (2007).

158. W.-X. Hong *et al.*, Design, synthesis, and properties of branch-chained maltoside detergents for stabilization and crystallization of integral membrane proteins: human connexin 26., *Langmuir* **26**, 8690–6 (2010).

159. T. Arnold, D. Linke, Phase separation in the isolation and purification of membrane proteins., *BioTechniques* **43**, 427–30, 432, 434 passim (2007).

160. Y.-T. Wu, D.-Q. Lin, Z.-Q. Zhu, Thermodynamics of aqueous two-phase systems—the effect of polymer molecular weight on liquid-liquid equilibrium phase diagrams by the modified NRTL model, *Fluid phase equilibria* **147**, 25–43 (1998).

161. H. Yuan *et al.*, Aggregation of sodium dodecyl sulfate in poly (ethylene glycol) aqueous solution studied by 1H NMR spectroscopy, *Colloid and Polymer Science* **280**, 479–484 (2002).

162. C. Fernandez-Patron, in *The Protein Protocols Handbook*, (Springer, 2002), pp. 251–258.

163. C. G. Bevans, M. Kordel, S. K. Rhee, A. L. Harris, Isoform composition of connexin channels determines selectivity among second messengers and uncharged molecules., *J. Biol. Chem.* **273**, 2808–16 (1998).

164. L. K. Buehler, K. A. Stauffer, N. B. Gilula, N. M. Kumar, Single channel behavior of recombinant beta 2 gap junction connexons reconstituted into planar lipid bilayers., *Biophys. J.* **68**, 1767–75 (1995).

165. I. Kufareva *et al.*, A novel approach to quantify G-protein-coupled receptor dimerization equilibrium using bioluminescence resonance energy transfer., *Methods Mol. Biol.* **1013**, 93–127 (2013).

166. N. M. Kumar, D. S. Friend, N. B. Gilula, Synthesis and assembly of human beta 1 gap junctions in BHK cells by DNA transfection with the human beta 1 cDNA., *J. Cell. Sci.* **108 (Pt 12)**, 3725–34 (1995).

167. D. Segretain, M. M. Falk, Regulation of connexin biosynthesis, assembly, gap junction formation, and removal., *Biochim. Biophys. Acta* **1662**, 3–21 (2004).

168. M. C. Fiori *et al.*, Permeation of calcium through purified connexin 26 hemichannels., *J. Biol. Chem.* **287**, 40826–34 (2012).

169. J.-L. Popot *et al.*, Amphipols from A to Z., *Annu Rev Biophys* **40**, 379–408 (2011).

170. C. Tribet *et al.*, Thermodynamic characterization of the exchange of detergents and amphipols at the surfaces of integral membrane proteins., *Langmuir* **25**, 12623–34 (2009).

171. C. Tribet, R. Audebert, J. L. Popot, Amphipols: polymers that keep membrane proteins soluble in aqueous solutions., *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15047–50 (1996).

172. P. S. Chae *et al.*, Tandem facial amphiphiles for membrane protein stabilization., *J. Am. Chem. Soc.* **132**, 16750–2 (2010).

173. P. Bazzacco *et al.*, Nonionic homopolymeric amphipols: application to membrane protein folding, cell-free synthesis, and solution nuclear magnetic resonance., *Biochemistry* **51**, 1416–30 (2012).

174. V. Polovinkin *et al.*, High-resolution structure of a membrane protein transferred from amphipol to a lipidic mesophase., *J. Membr. Biol.* **247**, 997–1004 (2014).

175. N. Planchard *et al.*, The Use of Amphipols for Solution NMR Studies of Membrane Proteins: Advantages and Constraints as Compared to Other Solubilizing Media., *J. Membr. Biol.* (2014), doi:10.1007/s00232-014-9654-z.

176. A. Sverzhinsky *et al.*, Amphipol-Trapped ExbB-ExbD Membrane Protein Complex from Escherichia coli: A Biochemical and Structural Case Study., *J. Membr. Biol.* (2014), doi:10.1007/s00232-014-9678-4.

177. M. Etzkorn, M. Zoonens, L. J. Catoire, J.-L. Popot, S. Hiller, How amphipols embed membrane proteins: global solvent accessibility and interaction with a flexible protein terminus., *J. Membr. Biol.* **247**, 965–70 (2014).

178. S. Elter *et al.*, The Use of Amphipols for NMR Structural Characterization of 7-TM Proteins., *J. Membr. Biol.* **247**, 957–64 (2014).

179. T. Dahmane, M. Damian, S. Mary, J.-L. Popot, J.-L. Banères, Amphipol-assisted in vitro folding of G protein-coupled receptors., *Biochemistry* **48**, 6516–21 (2009).

180. H.-C. Gao *et al.*, Mixed micelles of polyethylene glycol (23) lauryl ether with ionic surfactants studied by proton 1D and 2D NMR., *J Colloid Interface Sci* **249**, 200–8 (2002).

181. D. Locke, L.-X. Wang, C. G. Bevans, Y. C. Lee, A. L. Harris, Open pore block of connexin26 and connexin32 hemichannels by neutral, acidic and basic glycoconjugates., *Cell Commun. Adhes.* **10**, 239–44.

182. A. L. Harris, Connexin channel permeability to cytoplasmic molecules., *Prog. Biophys. Mol. Biol.* **94**, 120–43.

183. P. Davanloo, A. H. Rosenberg, J. J. Dunn, F. W. Studier, Cloning and expression of the gene for bacteriophage T7 RNA polymerase., *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2035–9 (1984).

184. M. A. Frederick *et al.*, *Current Protocols in Molecular Biology* (John Wiley & Sons, http://onlinelibrary.wiley.com/book/10.1002/0471142727, 2014).

185. C. T. Chung, S. L. Niemela, R. H. Miller, One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2172–5 (1989).

186. H. Inoue, H. Nojima, H. Okayama, High efficiency transformation of Escherichia coli with plasmids., *Gene* **96**, 23–8 (1990).

187. The QIAexpressionist: A handbook for high-level expression and purification of 6xHis-tagged proteins., *Qiagen Inc.* (2003).

188. Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, P. Tikhonov, Reconstruction of chromatographic peaks using the exponentially modified Gaussian function, *Journal of Chemometrics* **25**, 352–356 (2011).

189. P. G. Blommel, B. G. Fox, A combined approach to improving large-scale production of tobacco etch virus protease., *Protein Expr. Purif.* **55**, 53–68 (2007).

190. R. E. Stephens, High-resolution preparative SDS-polyacrylamide gel electrophoresis: fluorescent visualization and electrophoretic elution-concentration of protein bands., *Anal. Biochem.* **65**, 369–79 (1975).

Appendix

A1. Sequence alignment of human connexins.

```
                            440       450       460       470       480       490       500       510       520       530       540       550       560       570
SP|A6NN92|CXE1_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|Q6PEY0|CXB7_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|P29033|CXB2_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|O95452|CXB6_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|Q9NTQ9|CXB4_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|O75712|CXB3_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|O95377|CXB5_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|Q8NFK1|CXG3_HUMAN    -----------------------KKATDSLPVVET---KEQ------FQE-AVP-GRSLAQE------------------------KQRPVGPRDA----------------------------------------
SP|P08034|CXB1_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|Q8N144|CXD3_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|Q9UKL4|CXD2_HUMAN    -------------------------------------------------------------------------------------------------------------------------------------------------
SP|P35212|CXA4_HUMAN    -------------------SEQNWANLTTE-----E---RLASS-RPP-LFLDPPPQN-----GQKP-PSRPSSS------------AS------K------------------KQYV------------------
SP|P36382|CXA5_HUMAN    -------------------QNTDNLVTE-----Q---VRGQE-QTP-G-EGFIQVR----YGQKPEVPNGV-S----------------PGHRLPH-----------------------------------------
SP|Q96KN9|CXD4_HUMAN    --ASEKL-------------GRQ-----PRGRPHREAAQDPRGSGSEEQPSAAPSRLAAPPSCS----SLQPPDPPASS----------------SG----------APHLRARKSEWV----------------
SP|P17302|CXA1_HUMAN    ----------------SCRNYN-KQASEQNWANYSAE-----Q---NRMGQ-AGS-TISN--------SHAQP--F----------------DFPDDNQ-------------------NSKKLAA---------
SP|P36383|CXG1_HUMAN    -------------------YTE---LSNAKIAYKQNKANTAQEQQY-GSH-EENLPADLEALQREIRMAQERLDLAV-QAYS--HQNNPHGPREKKAKVGSKAGSNKSTASSKSGDGKTSVWI----------
SP|P48165|CXA8_HUMAN    ----DLS---R---------GYQ-ETLPSYAQVGAQ---------EVEGE-----GPPAEE-----GAEP-EVGEKKE-------------EAERLTTEE----------------QEKVAVPEGEKVETPGVDKEGE
SP|Q9Y6H8|CXA3_HUMAN    -----KLYNGHH----------HLL-MTEQNWANQAAE-----R---Q------PP-ALKAYPAAS----TPAAPSPVGSSSP--------------PLAHEAE--A-------------GAAPLLLDGSGSSLEGSALAGT
SP|Q5T442|CXG2_HUMAN    ----------------RAHDQNLANLALQALRDGAAAGDRDRD-SSP-CVGLPAASRQPPRAGA-PASRTGSAT-SA----GTVGEQGRPGTHERPGAKPRAGSEKGSASSRDGKTTVWI--------------
SP|P57773|CXA9_HUMAN    CSHFQHISSN-NNKDTHKIFGKELNGNQLMEKRETEGKDSKRNYYSRGHR-SIP-GVAIDGENN----MRQSPQTVFSLPANCDWKPRWLRATWGSSTEHENRGSPPK----------GNLKGQFRKGTVRTLPPSQGDSQ
SP|Q969M2|CXA10_HUMAN   CPWAGSAGNQHLGQQSDHSSFGLQNTMSQSWLGTTTA-----P---RN-----CP-SFAV-GTWE----QSQDPEPSGEPLT------------DLHSHCRDSEGSM--------------RESGVWID---RSRPGSRKASF
```

jnetpred

JNETCONF

JNETHMM

Conservation

Quality

Consensus

Figure A1.

a. List of the protein constructs and plasmids prepared and tested in course of the study.

#	N_t tag	N_t linker	N_t prosite	Cx species	C_t prosite	C_t tag	Vector
1	H_6	Lx..Xpress	EkFx	hCx26wt			pTrc
	MggsHHHHHH	gmasmtggqqmgrdly	DDDDK'drwgseleIDGR'				
2	H_6		Tr	hCx26wt			pET15b
	HHHHHH		LVPR'GS				
3				hCx26wt		H_6	pStaby
4				hCx26$_{M34A}$		H_6	pStaby
5	Cherry		Ek	hCx26wt		H_6	pStaby
			DDDDK'				
6		HAT	Ek	ohCx26$_{M34A}$			pSCodon
		KDHLIHNVHKEEHAHAHNKI		optimized code			
7	Cherry	HAT	Ek	ohCx26$_{M34A}$			pSCodon
		KDHLIHNVHKEEHAHAHNKI					
8				ohCx26$_{M34A}$		H_6	pSCodon
9	Cherry		Ek	ohCx26$_{M34A}$		H_6	pSCodon
10				ohCx26$_{M34A}$			pSCodon

#							
11	H6-Mistic	L5	Ek	ohCx26M34A			pSCodon
		gsggs					
10	H6-Mistic	L5^P	Ek	ohCx26M34A			pSCodon
		gpggs					
11	H6-Mistic	L7	Ek	ohCx26M34A			pSCodon
		gsgsggs					
12	H6-Mistic	L8	Ek	ohCx26M34A			pSCodon
		gpgssgpg					
13	H6-ssg-Mistic	L7	Ek	ohCx26M34A			pSCodon
14	H6-ssg-Mistic	L8	Ek	ohCx26M34A			pSCodon
15	H6-ssg-Mistic	L7	Ek	ohCx26wt			pSCodon
16	H6-ssg-Mistic	L8	Ek	ohCx26wt			pSCodon
17	Mistic	L7TrStr	Ek	ohCx26wt	Fx	H8	pSCodon
		gpggssgLVPR'GSgasWSHPQFEKgaapv			gsgIEGR'sgap		
18	H6-Tr-Mistic	L9Tr	Fx	ohCx26M34A			pStaby
		gsgsggssgLVPR'GSgsg	IEGR'				
19	H6-Tr-Mistic	L9Tr	Fx	ohCx26wt	Fx	H8	pSCodon
20	Mistic	L9Tr	Fx	ohCx26wt	Fx	H8	pSCodon
21	Mistic	L7TrStr	Tem	ohCx26wt	Fx	H8	pSCodon

22	Mistic	L7TrStr	Tev	ohCx26wt	Fx	H8	pSCodon
23	Mistic	L7TrStr	Teg	ohCx26wt	Fx	H8	pSCodon
24				ohCx26wt	Fx	H8	pIVEX2.3d
25				ohCx26wt			pIVEX2.3d

A3. Nucleotide sequences and features of basic vectors and supplementary plasmids used in the study.

a) pSCodon 1.2

pSCodon 1.2 vector map:
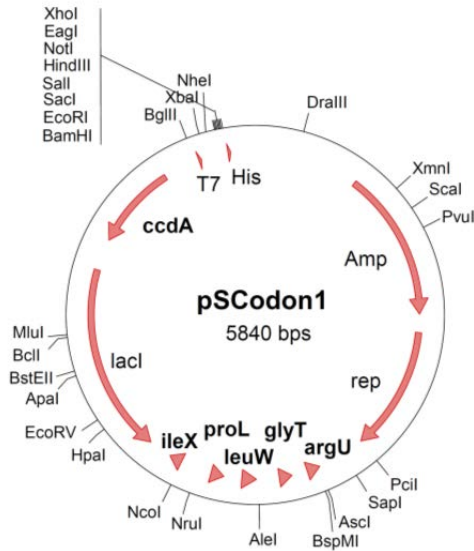
pSCodon 1.2 vector sequence:
```
tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgacc
gctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccg
gctttccccgtcaagctctaaatcggggggctccctttagggttccgatttagtgctttacggcacctcga
ccccaaaaaacttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccct
ttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctatct
cggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgattta
acaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtggcacttttcggggaaat
gtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataac
cctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttat
tccctttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgct
gaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagtt
ttcgccccgaagaacgtttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccg
tattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactca
ccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatga
gtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttttgca
caacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgac
gagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaactactta
ctctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctc
ggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcatt
gcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaacta
tggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagacca
agtttactcatatatactttagattgatttaaaacttcattttttaatttaaaaggatctaggtgaagatc
ctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagacccgtag
aaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaacc
accgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttc
agcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctg
tagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtg
tcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcg
tgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaa
gcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg
cacgagggagcttccaggggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt
```

```
gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttt
tacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtgga
taaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtca
gtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacacc
gcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgc
tacgtgactggggaattgtaatggcgcgccctgcaggattcgaacctgcggcccacgacttagaaggtcg
ttgctctatccaactgagctaagggcgcgttgataccgcaatgcggtgtaatcgcgtgaattatacggtc
aacccttgctgagtcaatggctttttgatcgcggtggctccccggtggccacggccacgcgatggcgtagc
ccgagacgataagttcgcttaccggctcgaataaagagagcttctctcgatattcagtgcagaatgaaaa
tcaggtagccgagttccaggatgcgggcatcgtataatggctattacctcagccttccaagctgatgatg
cgggttcgattcccgctgcccgctccaagatggggtgaatggtgcgggaggcgagacttgaactcgcaca
ccttgcggcgccagaacctaaatctggtgcgtctaccaatttcgccactcccgcaaaaaaagatggtggc
tacgacgggattcgaacctgtgaccccatcattatgagtgatgtgctctaaccaactgagctacgtagcc
atcttttttttcgtttggtcggcacgagaggatttgaacctccgacccccgacaccccatgacggtgcgc
taccaggctgcgctacgtgccgactcgtggctgctaatactaccgttttccacaccgattgcaagtaaga
tatttcgctaactgatttataattaatcgcgaggccgataccttatcggcgttgcgccatttataaaaac
agcaggcgcgcggtaatggctggattgcgacacggagttactttataatccgctaccatggccccttagc
tcagtggttagagcaggcgactcataatcgcttggtcgctggttcaagtccagcaggggccaccagatat
agccgaggctagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcgg
tcgagatcccggtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccag
tcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattg
ggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccct
gagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaa
cggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgc
agcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgg
gaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccg
ttccgctatcggctgaatttgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgag
acagaacttaatgggcccgctaacagcgcgatttgctggtgacccaatgcgaccagatgctccacgccca
gtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataa
cgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatc
agcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgccgcttcgttcta
ccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgacgg
cgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgcc
acgcggttgggaatgtaattcagctccgccatcgccgcttccacttttttcccgcgttttcgcagaaacgt
ggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataa
cgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaag
gttttgcgccattcgatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcag
cccagtagtaggttgaggccgttgagcaccgccgccgcaaggaatggtgcatgctcaccagtccctgttc
tcgtcagcaaaagagccgttcatttcaataaaccgggcgacctcagccatcccttcctgatttttccgctt
tccagcgttcggcacgcagacgacgggcttcattctgcatggttgtgcttaccagaccggagatattgac
atcgtatgccttgagcaactgatagctgtcgctgtcaactgtcactgtaatacgctgcttcatgcctgcc
cctccctttggtgaccaaccggctcgacgggggcagcgcaaggcggtgcctccggcgggccactcaatg
cttgagtatactcactagactttgcttcgcaaagtcgtgaccgcctacggcggctgcggcgccctacggg
cttgctctccgggcttcgccctgcgcggtcgctgcgctcccttgccagcccgtggatatgtggacgatgg
ccgcgagcggccaccggctggctcgcttcgctcggcccgtggacaacgcatgcaaggagatggcgcccaa
cagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggcga
gcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgc
cggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactataggg
gaattgtgagcggataacaattcccctctagaaataattttgtttaactttaagaaggagatatacatat
ggctagcatgactggtggacagcaaatgggtcgcggatccgaattcgagctccgtcgacaagcttgcggc
cgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaaggaagctgagttg
gctgctgccaccgctgagcaataactagcataaccccttggggcctctaaacgggtcttgaggggttttt
tgctgaaaggaggaactatatccggat
```

Features List:

Name	type	location
Amp	CDS	599..1456
rep	CDS	1560..2273
argU tRNA	CDS	rev:2542..2618
glyT	CDS	2823..2897
tRNA leuW	CDS	rev:2910..2994

```
tRNA proL              CDS                    rev:3096..3172
tRNA IleX              CDS                    3350..3425
lacI                   CDS                    rev:3556..4635
ccdA                   signal                 rev:4815..5033
ccdA (start: 5128) +PCDS                      rev:4815..5301
promoter T7            CDS                    5511..5527
His-Tag                CDS                    5681..5698
```

b) pRK603

pRK603 plasmid map:

pRK603 plasmid sequence:
gacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc
ttcacctcgagtccctatcagtgatagagattgacatccctatcagtgatagagatactgagcacatcag
caggacgcactgaccgaattcattaaagaggagaaaggtaccatgggagaaagcttgtttaaggggccgc
gtgattacaacccgatatcgagcaccatttgtcatttgacgaatgaatctgatgggcacacaacatcgtt
gtatggtattggatttgggccctcatcattacaaacaagcacttgtttagaagaaataatggaacactg
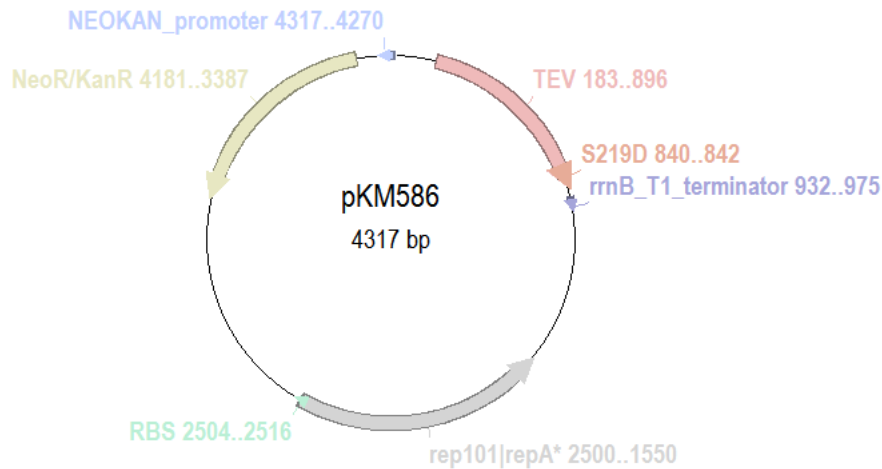ttggtccaatcactacatggtgtattcaaggtcaagaacaccacgactttgcaacaacacctcatcgatg
ggcgggacatgatcattattcgaatgcctaaggatttcccaccatttcctcaaaagctgaaatttcgcga
gccacaacgggaagagcgcatttgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtg
tcagacactagttgcacattcccttcatctgatggcatattctggaagcattggattcaaaccaaggatg
ggcagtgtggcagtccattagtatcaactcgagatgggttcattgttggtatacactcagcatcgaattt
caccaacacaaacaattatttcacaagcgtgccgaaaaacttcatggaattgttgacaaatcaggaggcg
cagcagtgggttagtggttggcgattaaatgctgactcagtattgtgggggggccataaagttttcatgg
acaaacctgaagagcctttcagccagttaaggaagcgactcaactcatgaattaataaggatcccatgg
tacgcgtgctagaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgtt
gtttgtcggtgaacgctctcctgagtaggacaaatccgccgccctagacctaggggatatattccgcttc
ctcgctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatggcttacgaacggggcgga
gatttcctggaagatgccaggaagatacttaacagggaagtgagagggccgcggcaaagccgttttttcca
taggctccgcccccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaacccgacagga
ctataaagataccaggcgtttcccccctggcggctccctcgtgcgctctcctgttcctgcctttcggttta
ccggtgtcattccgctgttatggccgcgtttgtctcattccacgcctgacactcagttccgggtaggcag
ttcgctccaagctggactgtatgcacgaacccccgttcagtccgaccgctgcgccttatccggtaacta
tcgtcttgagtccaacccggaaagacatgcaaaagcaccactggcagcagccactggtaattgatttaga
ggagttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagttttggtgactgcgctcct
ccaagccagttacctcggttcaaagagttggtagctcagagaaccttcgaaaaaccgccctgcaaggcgg
ttttttcgttttcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatcttattaatca
gataaaatatttctagatttcagtgcaatttatctcttcaaatgtagcacctgaagtcagccccatacga
tataagttgttactagtgcttggattctcaccaataaaaaacgcccggcggcaaccgagcgttctgaaca
aatccagatggagttctgaggtcattactggatctatcaacaggagtccaagcgagctctcgaacccag
agtcccgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggagcggcgatac
cgtaaagcacgaggaagcggtcagcccattcgccgccaagctcttcagcaatatcacgggtagccaacgc
tatgtcctgatagcggtccgccacacccagccggccacagtcgatgaatccagaaaagcggccattttcc
accatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgccgtcgggcatgcgcgcct
tgagcctggcgaacagttcggctggcgcgagcccctgatgctcttcgtccagatcatcctgatcgacaag
accggcttccatccgagtacgtgctcgctcgatgcgatgtttcgcttggtggtcgaatgggcaggtagcc
ggatcaagcgtatgcagccgccgcattgcatcagccatgatggatactttctcggcaggagcaaggtgag
atgacaggagatcctgccccggcacttcgcccaatagcagccagtcccttcccgcttcagtgacaacgtc

```
gagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcctgcagttca
ttcagggcaccggacaggtcggtcttgacaaaaagaaccgggcgcccctgcgctgacagccggaacacgg
cggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcctctccacccaagcggccgg
agaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatcctgtctcttgatcagatctt
gatcccctgcgccatcagatccttggcggcaagaaagccatccagtttactttgcagggcttcccaacct
taccagagggcgccccagctggcaattcc
```

Features List:

Name	type	location
rrnB_T1_terminator	terminator	932..975
NEOKAN_promoter	promoter	rev:2852..2899
TevS219D	CDS	186..893
Kan/neoR	CDS	rev:1969..2760
RBS_pZE21	RBS	165..176
PL_tetO	promoter	82..155
tetO	protein_bind	82..100
tetO	misc_feature	107..125
frw_AatII_prom	primer_bind	1..36
rev_Aat_term	primer_bind	1000..1028
t0	misc_feature	1838..1943
p15A	rep_origin	1035..1831

c) pKM586

pKM586 plasmid sequence:

```
gacgtctaagaaaccattattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc
ttcacctcgagtccctatcagtgatagagattgacatccctatcagtgatagagatactgagcacatcag
caggacgcactgaccgaattcattaaagaggagaaaggtaccatgggagaaagcttgtttaaggggccgc
gtgattacaacccgatatcgagcaccatttgtcatttgacgaatgaatctgatgggcacacaacatcgtt
gtatggtattggatttgggcccttcatcattacaaacaagcacttgtttagaagaaataatggaacactg
ttggtccaatcactacatggtgtattcaaggtcaagaacaccacgactttgcaacaacacctcatcgatg
ggcgggacatgatcattattcgaatgcctaaggatttcccaccatttcctcaaaagctgaaatttcgcga
gccacaacgggaagagcgcatttgtcttgtgacaaccaacttccaaactaagagcatgtctagcatggtg
tcagacactagttgcacattcccttcatctgatggcatattctggaagcattggattcaaaccaaggatg
ggcagtgtggcagtccattagtatcaactcgagatgggttcattgttggtatacactcagcatcgaattt
caccaacacaaacaattatttcacaagcgtgccgaaaaacttcatggaattgttgacaaatcaggaggcg
cagcagtgggttagtggttggcgattaaatgctgactcagtattgtggggggggccataaagttttcatgg
acaaacctgaagagcctttcagccagttaaggaagcgactcaactcatgaattaataaggatcccatgg
tacgcgtgctagaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgtt
gtttgtcggtgaacgctctcctgagtaggacaaatccgccgccctagacctagggtacgggttttgctgc
ccgcaaacgggctgttctggtgttgctagtttgttatcagaatcgcagatccggcttcaggtttgccggc
tgaaagcgctatttcttccagaattgccatgattttttccccacgggaggcgtcactggctcccgtgttg
tcggcagctttgattcgataagcagcatcgcctgtttcaggctgtctatgtgtgactgttgagctgtaac
aagttgtctcaggtgttcaatttcatgttctagttgctttgttttactggtttcacctgttctattaggt
gttacatgctgttcatctgttacattgtcgatctgttcatggtgaacagctttaaatgcaccaaaaactc
gtaaaagctctgatgtatctatctttttacaccgttttcatctgtgcatatggacagttttccctttga
tatctaacggtgaacagttgttctacttttgtttgttagtcttgatgcttcactgatagatacaagagcc
ataagaacctcagatccttccgtatttagccagtatgttctctagtgtggttcgttgttttttgcgtgagc
catgagaacgaaccattgagatcatgcttactttgcatgtcactcaaaaattttgcctcaaaactggtga
gctgaattttgcagttaaagcatcgtgtagtgttttttcttagtccgttacgtaggtaggaatctgatgt
aatggttgttggtattttgtcaccattcatttttatctggttgttctcaagttcggttacgagatccatt
tgtctatctagttcaacttggaaaatcaacgtatcagtcgggcggcctcgcttatcaaccaccaatttca
tattgctgtaagtgtttaaatctttacttattggtttcaaaacccattggttaagcctttaaaactcatg
gtagttattttcaagcattaacatgaacttaaattcatcaaggctaatctctatatttgccttgtgagtt
ttcttttgtgttagttcttttaataaccactcataaatcctcatagagtatttgttttcaaaagacttaa
catgttccagattatatttttatgaatttttttaactggaaaagataaggcaatatctcttcactaaaaac
taattctaattttttcgcttgagaacttggcatagtttgtccactggaaaatctcaaagcctttaaccaaa
ggattcctgatttccacagttctcgtcatcagctctctggttgctttagctaatacaccataagcatttt
ccctactgatgttcatcatctgagcgtattggttataagtgaacgataccgtccgttctttccttgtagg
gttttcaatcgtggggttgagtagtgccacacagcataaaattagcttggtttcatgctccgttaagtca
tagcgactaatcgctagttcatttgctttgaaaacaactaattcagacatacatctcaattggtctaggt
gattttaatcactataccaattgagatgggctagtcaatgataattactagtccttttcccgggagatct
gggtatctgtaaattctgctagacctttgctggaaaacttgtaaattctgctagaccctctgtaaattcc
gctagacctttgtgtgtttttttttgtttatattcaagtggttataatttatagaataaagaaagaataaa
aaaagataaaaagaatagatcccagccctgtgtataactcactactttagtcagttccgcagtattacaa
```

```
aaggatgtcgcaaacgctgtttgctcctctacaaaacagaccttaaaaccctaaaggcttaagtagcacc
ctcgcaagctcgggcaaatcgctgaatattccttttgtctccgaccatcaggcacctgagtcgctgtctt
tttcgtgacattcagttcgctgcgctcacggctctggcagtgaatgggggtaaatggcactacaggcgcc
ttttatggattcatgcaaggaaactacccataatacaagaaaagcccgtcacgggcttctcagggcgttt
tatggcgggtctgctatgtggtgctatctgacttttttgctgttcagcagttcctgccctctgattttcca
gtctgaccacttcggattatcccgtgacaggtcattcagactggctaatgcacccagtaaggcagcggta
tcatcaacaggcttacccgtcttactgtccctagtgcttggattctcaccaataaaaaacgcccggcggc
aaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatctatcaacaggagtccaag
cgagctctcgaaccccagagtcccgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcg
aatcgggagcggcgataccgtaaagcacgaggaagcggtcagcccattcgccgccaagctcttcagcaat
atcacgggtagccaacgctatgtcctgatagcggtccgccacacccagccggccacagtcgatgaatcca
gaaaagcggccattttccaccatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgc
cgtcgggcatgcgcgccttgagcctggcgaacagttcggctggcgcgagcccctgatgctcttcgtccag
atcatcctgatcgacaagaccggcttccatccgagtacgtgctcgctcgatgcgatgtttcgcttggtgg
tcgaatgggcaggtagccggatcaagcgtatgcagccgccgcattgcatcagccatgatggatactttct
cggcaggagcaaggtgagatgacaggagatcctgccccggcacttcgcccaatagcagccagtcccttcc
cgcttcagtgacaacgtcgagcacagctgcgcaaggaacgcccgtcgtggccagccacgatagccgcgct
gcctcgtcctgcagttcattcagggcaccggacaggtcggtcttgacaaaaagaaccgggcgcccctgcg
ctgacagccggaacacggcggcatcagagcagccgattgtctgttgtgcccagtcatagccgaatagcct
ctccacccaagcggccggagaacctgcgtgcaatcatcttgttcaatcatgcgaaacgatcctcatcct
gtctcttgatcagatcttgatcccctgcgccatcagatccttggcggcaagaaagccatccagtttactt
tgcagggcttcccaaccttaccagagggcgccccagctggcaattcc
```

Features List:

Name	type	location
source:pKM586	source	1..4317
TEV	CDS	183..896
rrnB_T1_terminator	terminator	932..975
rep101|repA*	CDS	rev:1550..2500
NeoR/KanR	CDS	rev:3387..4181
NEOKAN_promoter	promoter	rev:4270..4317
RBS	RBS	2504..2516
S219D	CDS	840..842

A5. *in vitro* reaction set up*

a. Detergent supplemented *in vitro* reaction set up, "50µl" scale, detergent selection

Einheiten	Konz. der Stock-Lösung	Kompon. des RM	Zugabe zum RM [µl]	zugegebe Endkonz.	Mg²⁺ Anteil [mM]	K⁺ Anteil [mM]		Endkonz.	Kompon. des FM	Zugabe zum RM/FM [µl]	zusätzliche Zugabe zum FM [µl]
µg/ml	10	*E.coli* lipids	0.0	0.00							
x	1	S30-Extrakt	294.0	0.35	4.9	21		0.35	S30-Puffer		3920
mg/ml	2.07	Plasmid	4.1	0.010							54
U/µl	40	RNasin (Amersham)	6.30	0.30							84
mg/ml	3.2	T7-RNA Pol.	10.50	0.04							140
mg/ml	40	tRNA *E.coli*	10.50	0.50							140
mg/ml	10	Pyruvat Kinase	3.36	0.04							45
mM	16.57	RCWMDE	30.42	0.60					RCWMDE	436.0	406
mM	4.0	AA-Mix	126.0	0.60				0.65	AA-Mix	1806.0	1820
mM	1000	AcP	16.8	20.00		22			AcP	240.8	
mM	1500	PEP	16.8	30.00		67			PEP	240.8	
x	75	NTP	11.2	1.0					NTP	160.5	
mM	500	DTT	3.4	2.0					DTT	48.2	
mg/ml	10	Folsäure	8.4	0.10					Folsäure	120.4	
x	50	Complete	16.8	1					Complete	240.8	
x	24	Puffer (H+E)	30.8	1.0		50			Buffer (H+E)	441.5	
mM	2000	Mg(OAc)₂	5.5	13.1	13.1				Mg(OAc)₂	78.9	
mM	4000	KOAc	27.3	130.0		130.0			KOAc	391.3	
%	40	PEG8000	42.0	2					PEG8000	602.0	
%	10	NaN₃	4.2	0.05					NaN₃	60.2	
		H₂O	171.7						H₂O	2461.1	902
			668.3							9579	10298
		MIX I Vol.	668.3							4867	10298
		Volumen RM	840	TOTAL	18	290.0			Vol RM+FM	12040	11200
		aus RM/FM	309.2		mM Mg²⁺	mM K⁺				4867	Volumen
					(+/- 2mM)	(+/- 20mM)			für RM	-309.2	FM
									für FM	4558.1	

*) The spreadsheet for calculation of the pipetting volumes was kindly provided by Frank Bernhard Lab, University of Frankfurt/Main, Max-von-LaueStrasse 9, D-60438 Frankfurt/Main, Germany

b. Detergent supplemented *in vitro* reaction set up, "preparative" scale

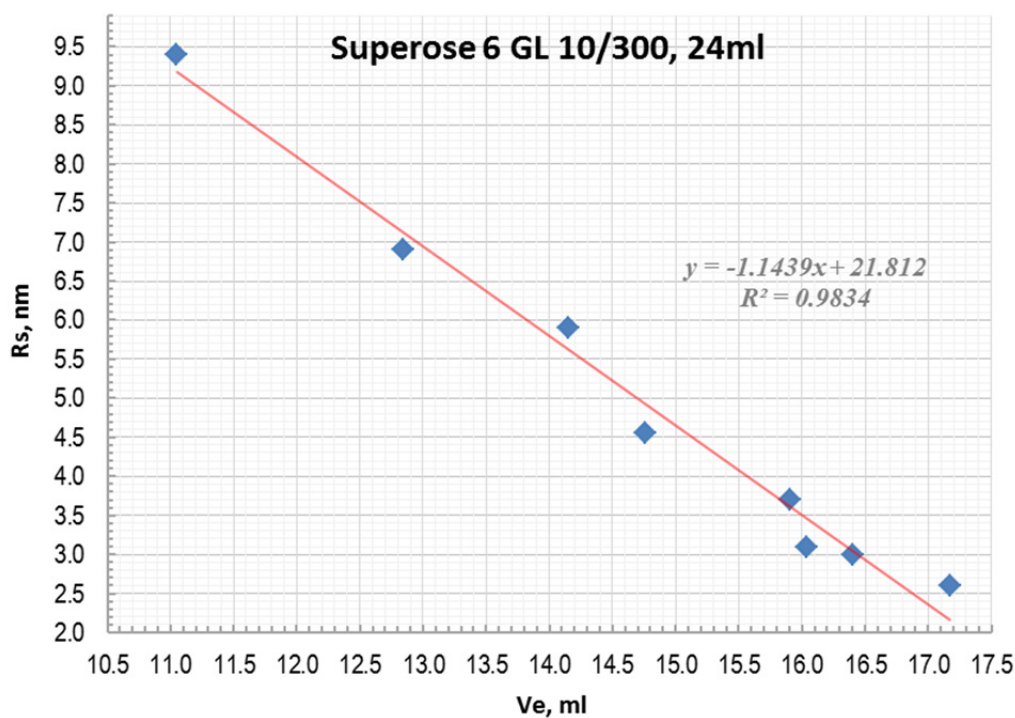| Einheiten	Konz. der Stock-Lösung	Kompon. des RM	Zugabe zum RM [µl]	zugegebe Endkonz	Mg²⁺ Anteil [mM]	K⁺ Anteil [mM]		Endkonz.	Kompon. des FM	Zugabe zum RM/FM [µl]	zusätzliche Zugabe zum FM [µl]
mM	200	retinal	0.0	0.00			7				
x	1	S30-Extrakt	542.5	0.35	4.9	21	8	0.35	S30-Puffer		7595
mg/ml	0.25	Plasmid	62.0	0.010			9				868
U/µl	40	RNasin (Amer	11.63	0.30			10				163
mg/ml	3.2	T7-RNA Pol.	19.38	0.04			11				271
mg/ml	40	tRNA *E.coli*	19.38	0.50			12				271
mg/ml	10	Pyruvat Kinase	6.20	0.04			13				87
%	10	Brij78	155.0	1.00			14		Brij78	2325.0	
mM	16.57	RCWMDE	56.1	0.60			15		RCWMDE	841.9	
mM	4.0	AA-Mix	232.5	0.60			16	0.4	AA-Mix	3487.5	2170
mM	1000	AcP	31.0	20.00		22	17		AcP	465.0	
mM	1500	PEP	31.0	30.00		67	18		PEP	465.0	
x	75	NTP	20.7	1.0			19		NTP	310.0	
mM	500	DTT	6.2	2.0			20		DTT	93.0	
mg/ml	10	Folsäure	15.5	0.10			21		Folsäure	232.5	
x	50	Complete	31.0	1			22		Complete	465.0	
x	24	Puffer (H+E)	56.8	1.0		50	23		Buffer (H+E)	852.5	
mM	2000	Mg(OAc)₂	10.2	13.1	13.1		24		Mg(OAc)₂	152.3	
mM	4000	KOAc	50.4	130.0		130.0	25		KOAc	755.6	
%	40	PEG8000	77.5	2			26		PEG 8000	1162.5	
%	10	NaN₃	7.8	0.05			27		NaN₃	116.3	
		H₂O	107.3				28		H₂O	1609.8	993
			1442.7				29			21640	20707
		MIX I Vol.	1550.0				30			11724	20707
		Volumen RM	1550	TOTAL	18	290.0	31		Vol RM+FM	23250	21700
		aus RM/FM	781.6		mM Mg²⁺	mM K⁺	32			11724	Volumen
					(+/- 2mM)	(+/- 20mM)	33		für RM	-781.6	FM
									für FM	10942.4	
			E							L	M

119

A6.

Figure A6. Calibration plot for the GE Healthcare Superose 6 GL 10/300 size-exclusion chromatography column. The elution volume dependence on the particle hydrodynamic radius determined by quasy-elastic light scattering for globular protein standards: 29 kDa carbonic anhydrase 2 from bovine erythrocytes, 44 kDa ovalbumin from chicken egg white, 67 kDa bovin serum albumin, 150 kDa tetrameric alcohol dehydrogenase from *S.cerevisiae*, 200 kDa tetrameric β-amylase from sweet potato, 440 kDa apo-ferritin from horse spleen (major peak I), 660 kDa dimeric thyroglobulin, 773 kDa Ferritin, apo-ferritin from horse spleen (minor peak II). Column was developed at 0.2 ml/min in a 100 mM NaCl solution buffered with 10 mM Tris-HCl pH 7.5. Protein samples were prepared in concentration of 0.1-0.2 mg/mL in the running buffer and injected in a volume of 0.5 ml. The column was developed at 0.2 ml/min. Measured values are represented by blue dots. Red line represents a trend obtained by the simple linear regression analysis. The regression parameters are given in the plot by a grey script.
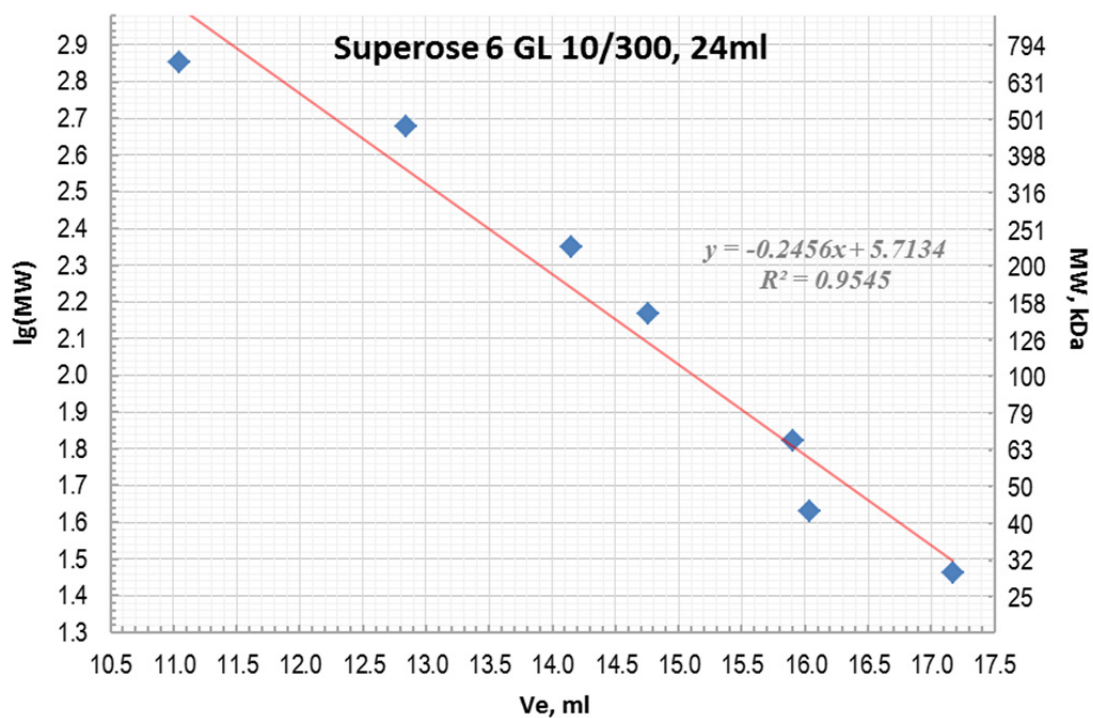
A7.

Figure A7. Calibration plot for the GE Healthcare Superose 6 GL 10/300 size-exclusion chromatography column. The elution volume dependence on the protein molecular weight for the globular protein standards. The detailed description is given in the legend to figure A.6.

A8.

Name	Ve, ml	Rs, nm	hhw, ml	hhw_r	hhw/Ve
Aceton	21.30	0.25	0.60	1.00	0.028
CA	17.17	2.60	0.57	0.95	0.033
A835	16.40	3.00	2.25	3.75	0.137
OVA	16.04	3.10	0.74	1.23	0.046
BSA	15.91	3.70			
ADH	14.76	4.55	0.59	0.98	0.040
bAmylase	14.15	5.90	0.73	1.22	0.052
Ferritin, pI	12.84	6.90			
Ferritin, pII	11.05	9.40			

Figure A8. Calibration data for Superose 6 GL 10/300. . More details are given in the legend to Figure A.6