universität wien

MASTERARBEIT

Titel der Masterarbeit

"Expression and purification of recombinant surface proteins G and F of Human Respiratory Syncytial Virus (HRSV)"

verfasst von

Kristina Borochova, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 066830
Studienrichtung lt. Studienblatt:	Masterstudium Molekulare Mikrobiologie und Immunbiologie
Betreut von:	Univ. Prof. Dr. Rudolf Valenta

DANKSAGUNG

Manche Menschen sind im Leben wie Rückenwind und helfen einem stets dabei vorwärts zu kommen. Diesen besonderen Menschen möchte ich hiermit herzlich danken.

Im Speziellen gilt mein Dank:

Univ. Prof. Dr. Rudolf Valenta für die mir gegebene Chance, an einem so interessanten Projekt mitarbeiten zu dürfen sowie für seine Betreuung, Beratung und Unterstützung in dieser Zeit.

Weiters bin ich meiner Betreuerin, Katarzyna Niespodziana, PhD, für ihre Geduld, Hilfestellungen und ihr, in mich gesetztes Vertrauen während der praktischen Arbeit, zu großem Dank verpflichtet. Auch danke ich ihr vielmals für die investierte Zeit und die vielen lieben aufbauenden Worte, wenn einmal etwas nicht planmäßig gelaufen ist.

Natürlich sind auch die anderen Damen in Prof. Valentas Team: Clarissa Cabauatan, PhD, Mag. Daniela Gallerano, Carolin Cornelius, MSc, Dr. med. Eva Wollmann und Mag. Yvonne Resch dankend zu erwähnen, die mir ebenfalls oft mit Rat und Tat zur Seite standen, wenn ich einmal nicht weiter wusste.

Selbstverständlich gibt es auch in meinem privaten Umfeld zahlreiche Menschen, ohne deren Hilfe diese Arbeit nie entstanden wäre und welchen an dieser Stelle gedankt werden soll.

Allen voran danke ich herzlich meinen lieben Eltern, die mich häufig mit frisch zubereitetem Essen versorgt haben, wenn es einmal später wurde. ;-).

Auch danke ich sehr Familie List, für Ihre liebevolle Betreuung meines Hundes in den vielen stressigen Zeiten.

Der größte Dank gebührt meinem Freund MMag. Michael List für seine Unterstützung, seine Motivation, seine Geduld sowie den unerschütterlichen Glauben an mich während der letzten Jahre.

Nun danke ich allen anderen Menschen, welche mich unterstützt haben, jedoch aus Platzmangel hier nicht angeführt werden können.

Table of Contents

Abbreviations

aa	Amino acid
ACN	Acetonitrile
AP	Alkaline Phosphatase
APS	Ammonium Persulfate
A2	Human Respiratory Syncytial Virus subtype A2
BA	Human Respiratory Syncytial Virus subtype BA
BCA	bicinchoninic acid *assay*
BCIP	5-bromo-4-chloro-3-indolyl-phosphate
BRSV	Bovine Respiratory Syncytial Virus
BSA	Bovine Serum Albumin
CD	Circular dichroism
Cys	Cystein
CXCR1	Chemokine receptor (Interleukin 8 receptor α)
CX3C	Fractalkine
DCM	Dichlormethane
DIEA	N,N-Diisopropylethylamine
DMF	Dimethylformamide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
F	Fusion protein F
Fmoc	Fluorenylmethyloxycarbonyl
FT	Flow through
F 1/ 2	Fusion protein subunit 1 or 2
G	Attachment protein G
GuHCl	Guanidin hydrochloride
HbTU	Hydroxybenzotriazole
Hfq	Host-factor-I protein
His-tag	Histidine-tag
HOBt	O-Benzotriazole-N,N,N’,N’-tetramethyl-uronium-hexafluoro-phosphate
HPLC	High-performance liquid chromatography

HRSV	Human Respiratory Syncytial Virus
HR 1/2/3	Heptad 1 to 3 repeat
IPTG	Isopropyl-β-thiogalactopyranoside
Kan	Kanamycine
Kb	Kilo-Base
kDa	Kilo Dalton
L	Large polymerase protein L
LB	Luria-Bertani
M	Matrix protein M
mAb	Monoclonal antibodies
MALDI-TOF	Matrix-assisted laser desorption - time of flight
mAU	milli Absorption Unit
MW	Molecular weight
MWCO	Molecular weight cut off
N	Nucleocapsid protein N
NBT	Nitro blue tetrazolium
NCBI	National Center for Biotechnology Information
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
Ni-NTA	Ni-NTA - Nickel-nitrilotriacetic acid
NMP	N-Methyl-2-pyrrolidone
NMR	Nuclear Magnetic Resonance Spectroscopy
NS 1/2	Nonstructural 1 or 2 protein NS1, NS2
ompA	Outer membrane protein A
ORF	Open reading frame
P	Phosphoprotein P
PAA	Polyacrylamide
PBS	Phosphate Buffered Saline
PEG-PS	Polyethylene glycol- Polystyrene
pI	Isoelectric point
PMSF	phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluoride
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis

SH	Small hydrophobic protein SH
TB	Terrific Broth
TFA	Trifluoro acetic acid
TNF	Tumor necrosis factor
TNFr	Tumor necrosis factor receptor
WB	Western-blot
yadF	Carbonic dehydratase

Zusammenfassung

Diese Arbeit beschäftigt sich mit der Herstellung und Charakterisierung rekombinanter Oberflächenproteine G ("Attachment" Protein) und F ("Fusion" Protein), des Humanen Respiratorischen Syncytial Virus (HRSV), welcher der Familie der *Paramyxoviridae* und der Gattung der *Pneumoviren* zuzuordnen ist. Dieser Krankheitserreger spielt bei Infektionen der oberen und unteren Atemwege eine bedeutende Rolle und führt insbesondere bei Säuglingen, Kindern, älteren sowie immungeschwächten Personen oftmals zu lebensbedrohlichen Erkrankungen des respiratorischen Traktes. Im Detail handelt es sich um einen umhüllten Virus mit einzelsträngiger negative-sense RNA als Genom, welcher über eine Genomgröße von ca. 15,2 kB verfügt, die für die Codierung von 11 Proteinen verantwortlich sind. Auf der viralen Hülle befinden sich die Oberflächenproteine F, G und SH ("small hydrophobic" Protein), wobei vor allem das Oberflächen-Glykoprotein G bzw. dessen Sequenzvariabilität für die weitere Klassifizierung in HRSV-Subtype A und Subtype B herangezogen wird. Die beiden anderen Proteine zeigen eine deutlich geringere (SH) bzw. fast keine Sequenzvariabilität (F).

Bis dato existieren nur wenige Studien, welche erfolgreich zeigen konnten, dass HRSV Oberflächenproteine in *E.coli* exprimiert werden können. Aus diesem Grund war das Hauptziel dieser Arbeit die Konstruktion, Expression und Reinigung der HRSV Oberflächenproteine G sowie der Subeinheiten F1 und F2 des F Proteins.

Im ersten Schritt wurde eine genaue Sequenzanalyse der viralen Oberflächenproteine unterschiedlicher HRSV Subtypen durchgeführt. Die Aminosäure Sequenzen wurden aus der NCBI Datenbank bezogen und unter Verwendung der Software Clustal W miteinander verglichen und anschließend anhand ihrer Ähnlichkeiten gruppiert. A2, welcher zur HRSV Subtype A gehört sowie BA welche der Subtype B zuzuordnen ist, wurden ausgewählt und für die Produktion rekombinanter Proteine sowie synthetischer Peptide verwendet.

Für die Herstellung der rekombinanten Proteine wurde zunächst ein Hexahistidin-Tag (6xHis) an den 3` Enden der codon-optimierten cDNAs angebracht, um die spätere Reinigung mittels Ni-NTA Affinitätschromatografie zu erleichtern. Die modifizierten Nukleinsäurestränge, welche für die Protein-Subeinheiten F1 und F2 sowie das Protein G kodieren, wurden anschließend in das pET-27b Plasmid innerhalb der *NdeI/XhoI* Restriktionsseiten eingefügt. Die Expression der rekombinanten Proteine erfolgte in *E. coli* BL 21 (DE3) Zellen.

Jede Kultur wurde von einer einzelnen transformierten *E. coli* Kolonie gewonnen und in 1L LB-Medium mit Zusatz eines passenden Antibiotikums in einem Schüttelinkubator bei 37°C inkubiert. Sobald der OD_{600nm} Wert von 0.5 erreicht wurde, erfolgte die Zugabe von IPTG um die Expression des Zielproteins zu induzieren. Im nächsten Schritt wurden die Zellen durch Zentrifugation vom Medium getrennt und durch chemische sowie mechanische Maßnahmen aufgeschlossen. Die rekombinanten Proteine wurden anschließend von anderen kontaminierenden Bestandteilen unter nativen oder denaturierenden Bedingungen auf einer Ni-NTA Säule gereinigt. Während der Reinheitsgrad durch eine SDS-PAGE Analyse bestimmt wurde, kam für den Nachweis der Proteinidentität ein Western-Blot (Anti-His-Tag-Antikörper) zur Anwendung. Abschließend kam es zu einer Charakterisierung der gewonnenen Proteine anhand ihrer physikochemischen Eigenschaften, welche mittels Zirkulardichroismus-Spektroskopie sowie MALDI TOF-Analyse ermittelt wurden.

Zusätzlich zu den rekombinanten Proteinen wurden synthetische Peptide hergestellt, welche das gesamte Oberflächen-Glykoprotein G überspannen. Dazu wurde die Fluorenylmethoxycarbonyl (Fmoc) Festphasen Peptid Synthese unter Verwendung eines PEG-PS enthaltenden Matrix genutzt. In weiterer Folge wurden die Peptide mit Dichlormethan gewaschen und von der Matrix befreit. Dafür kam ein Gemisch aus 95%iger Trifluoressigsäure (TFA) und 3.5 %igem Silane zum Einsatz. Anschließend erfolgte die Präzipitation in kaltem *tert*-Butylmethylether. Die Entfernung von unerwünschten Nebenprodukten wurde mittels eines 10 -70 %igem Acetonitril Gradienten im Zuge einer Reversed-Phase-HPLC erreicht. Zur Feststellung der korrekten Peptididentität kam erneut die MALDI-TOF Spektrometrie zum Einsatz. Peptidfraktionen mit einem hohen Reinheitsgrad sowie hoher Intensität im Massenspektrum wurden abschließend lyophilisiert und bei -20°C gelagert.

Zusammenfassend wurden während dieser Masterarbeit zwei rekombinante virale Oberflächenproteine und 16 synthetische Peptide des G Proteins hergestellt. Diese Proteine und Peptide können in weiterer Folge für die Entwicklung von serologisch diagnostischen Assays verwendet werden, welche die Antikörperantwort auf eine HRSV Infektion messen. Dadurch können unter Umständen neue Erkenntnisse gewonnen werden, welche als Grundlage für die Entwicklung neuer und innovativer Diagnose und Therapieverfahren dienen.

Summary

The major focus of my master thesis was the production of recombinant proteins and synthetic peptides derived from the Human Respiratory Syncytial Virus (HRSV). This pathogen is an important causal agent of severe respiratory tract infections in infants and children as well as in elderly and immunodeficient persons. The HRSV belongs to the Pneumovirus genus of the *Paramyxoviridae* family. It is an enveloped virus with a negative-sense single-stranded RNA genome of approximately 15,2 kb which encodes for 11 proteins. This pathogen exists as a single serotype with two antigenic subgroups A and B. Genotyping of HRSV-A and HRSV-B viruses is mainly based on the sequence variability of the surface "G" attachment glycoprotein. The other two proteins on the viral surface, "SH" small hydrophobic protein and "F" fusion glycoprotein, show low or no relevant sequence variability, respectively.

Only few studies reported the successful expression of HRSV surface proteins in bacteria. Therefore, the major aim of my master thesis was to construct, express and purify HRSV surface protein G and F protein subunits, F1 and F2.

In a first step, a detailed sequence analysis of viral envelope proteins from different HRSV strains was performed. Amino acid sequences were retrieved from the NCBI database, aligned using Clustal W and grouped according to the percentages (%) of amino acid sequence identities. Two representative HRSV strains were then selected: A2- belonging to the HRSV genotype A and BA- belonging to HRSV group B, respectively. Next, synthetic genes coding for F and G surface proteins were made based on a codon-usage optimized for the expression in *Escherichia coli*. The cDNAs were synthesized with the addition of a sequence encoding a hexahistidine tag at the 3` end and inserted into the NdeI/XhoI site of the pET-27b vector.

The HRSV surface proteins G and both F protein subunits were expressed in *E. coli* BL21 (DE3). Each culture derived from a single *E. coli* colony and the expression of the target protein was induced by adding IPTG. Cells were harvested by centrifugation and lysed by chemical and mechanical homogenization. Recombinant proteins were purified under native or denaturing conditions from the soluble or insoluble fractions using Ni-NTA resin, respectively. The purity and correct identity of the proteins was analyzed by SDS-PAGE and Western-blot, using monoclonal anti-His-tag antibodies, respectively. Proteins were also characterized in relation

to their physicochemical properties, including the measurement of circular dichroism (CD) spectra and the determination of molecular masses using MALDI-TOF analysis.

In addition to recombinant proteins also synthetic peptides spanning the surface glycoprotein G were synthesized by solid phase synthesis with the 9-fluorenyl-methoxy-carbonyl-method, using a PEG-PS preloaded resin. Subsequently, peptides were washed with dichlormethan, cleaved from the resin in 95% TFA, 2.5% Silane and precipitated into *tert*-Butylmethylether. Separation from by-products was achieved by reversed-phase HPLC in an 10-70% acetonitrile gradient and the identity of the peptides was confirmed by mass spectrometry. The mass spectrum of each fraction was analyzed using Flex analysis software. High intensity lines with the right molecular weight of the peptide were identified and fractions of the same composition were pooled together and finally lyophilized.

Taken together, two recombinant HRSV surface proteins and 16 synthetic peptides were produced during the course of my master thesis. These proteins and peptides may subsequently be used to establish a serological diagnostic assay measuring antibody responses to HRSV following experimental human infection. Analysis of HRSV-specific antibody responses may provide new insights for the development of new and innovative diagnostic tools or therapeutic agents.

I Introduction

This work describes the production of recombinant proteins and synthetic peptides of the Human Respiratory Syncytial Virus (HRSV) surface glycoprotein G and F. According to current knowledge these two proteins are the only viral elements that induce HRSV neutralizing antibodies and are therefore considered targets for developing new diagnostic techniques and therapeutic agents. [1] Both envelope glycoproteins will be characterized in detail on the basis of amino acid sequence alignments. Furthermore, methods used to purify and to characterize these proteins and synthetic peptides are described briefly. But first the Human Respiratory Syncytial Virus itself will be presented shortly.

I.1. Prevalence and clinical significance

The Human Respiratory Syncytial Virus (HRSV) is a very common pathogen and an important causative agent of severe respiratory tract infections in infants, elderly and immunocompromised persons.

This pathogen appears to be one of the most infectious and ubiquitous viruses, only affecting humans and all children by age of two years display positive serologic tests. [2, 3] Although most children show no illness or mild symptoms similar to the common cold, some children require hospitalization due to severe lower respiratory tract symptoms ranging from bronchiolitis to pneumonia. The global burden of the disease study 2010 showed that the HRSV is responsible for 6.7% of deaths in children between one month to one year of age and causes also mortality in elderly people in a range similar to influenza virus. [4] Moreover, multiple studies have demonstrated that infection with HRSV in early childhood, leading to hospitalization due to severe respiratory symptoms, might trigger wheezing and allergic asthma in later life. Therefore and because of the immense healthcare costs, approximately $ 300 000 000 per year in the United States due to hospitalization of infants, this viral disease receives high attention. [5, 6, 7, 8]

Transmission and infection of virus takes place directly through close contact with infected people or indirectly by touching contaminated surfaces whereby virus particles come in contact with the mucous membranes of nose, mouth or eyes. The incubation period from

exposure to first observable symptoms is about 2 to 8 days.[9] There are seasonal peaks of disease frequency favouring meteorological conditions such as temperatures between 2-6°C, 24-30°C and 45-65% humidity.[10]

The requirements to gain immunity against HRSV are poorly understood and relapsing infection can occur throughout the life at all ages. Reasons which may contribute to the recurrent infection with this virus include antigenic differences between the two major groups of HRSV (A and B) and genetic variability within a group itself. [11]

This pathogen exists as a single serotype with two antigenic subgroups A and B. The two antigenic groups have been classified on the basis of cross-neutralization studies using monoclonal antibodies against the surface glycoproteins. The main differences between HRSV-A and HRSV-B are located in the surface "G" attachment glycoprotein, with approximately 50% amino acid homology within subgroups.[12] The other two proteins on the viral surface, "SH" small hydrophobic protein and "F" fusion glycoprotein show less or almost no sequence variability between A and B groups, respectively.

Both antigenic groups HRSV-A and HRSV-B can co-circulate within the same time of the year and are responsible for epidemics. However, HRSV-A strain seems to be associated with severe disease and is more common than group B infections.[13, 14] HRSV-A and HRSV-B viruses have been further classified into subgroup types. Eleven HRSV-A subtypes have been reported from different geographical regions, and designated as GA1 to GA7, SAA1 (South Africa, A1), NA1, NA2 and ON1. HRSV-B subtypes include 20 different genotypes denoted as GB1 to GB4, SAB1 to SAB4, URU1 to URU2 and BA1 to BA10 (Buenos Aires).[15, 16]

At present no specific anti-HRSV medication or efficient and safe vaccines have been developed. An exception is the therapeutical drug, Palivizumab (Synagis), a humanized F-directed monoclonal antibody, which is used as a passive prophylaxis in high-risk infants to prevent HRSV infection. Ribavirin is the only drug currently available for HRSV infection therapy (non-specific) but its clinical use is limited due to ongoing concerns about its effectiveness, side effects and high costs. [17] In the majority of cases supportive care as mechanical removal of secretions, proper positioning of the infant as well as the administration of anti-inflammatory agents, bronchodilators and humidified oxygen are used for treatment to ease symptoms.[18]

I.2. The HRSV genome

The Human Respiratory Syncytial Virus belongs to the *pneumovirus* genus of the *Paramyxoviridae* family of order *Mononegavirales*. It is an enveloped virus with a negative-sense single-stranded nonsegmented RNA genome of approximately 15,2 kb which encodes for 11 proteins. [19]

NS1 and NS2 are the non-structural proteins-encoding part on the 3`end of the negative sense genome, known to inhibit the release of interferon-α/ß in response to viral infection by the host.[20] HRSV N (nucleocapsid) associates to genomic RNA as well as to the replicative intermediate antigenomic RNA to create RNAse-resistant nucleocapsids. The next following part on the genome is coding for the P (phosphoprotein) protein which has a function as a polymerase cofactor and helps to maintain soluble nucleocapsid by binding this protein, allowing the assemble process. After this the information for the M (matrix) protein is positioned which forms the inner surface of the envelope and is therefore important for virion morphogenesis. Then, the three surface envelope glycoproteins encoding parts for SH (small hydrophobic) protein of unknown function, G (attachment) and F (fusion) proteins are designated on genome followed by the M2-encoding part which form two proteins M2-1 and M2-2 due to two overlapping open reading frames (ORF`s).[21] M2-1 associates with the nucleocapsid and functions as a transcription factor whereas M2-2 appears to have a regulatory role in shifting the balance between RNA transcription and replication.[22, 23] The last information is located at the 5` end of the genome, it is the L (large polymerase) protein which has polymerase catalytic domains and associates together with P and M2-1 with the nucleocapsid.[24]

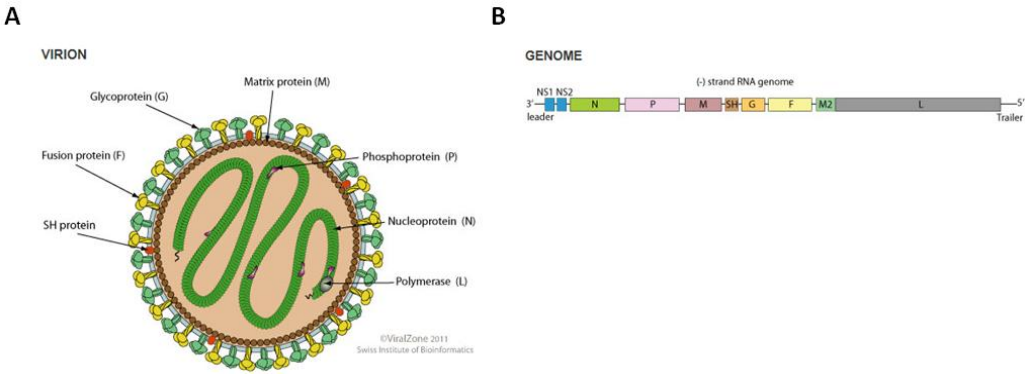


Figure 1: A) Enveloped human respiratory syncytial virion structure and **B)** schematic negative-stranded, linear viral genome are shown. Protein encoding parts are indicated (**B**): **NS1/2** (non-structural protein), **N** (nucleocapsid), **P** (phosphoprotein), **M** (matrix), **SH** (small hydrophobic protein), **G** (attachment-glycoprotein), **F** (fusion protein), **M2-1/2** (transcription factor and regulatory protein). (http://viralzone.expasy.org/all_by_protein/90.html)

I.3. Characterization of the envelope proteins G and F

I.3.1. G-protein amino acid sequence analysis

Genotyping of HRSV-A and HRSV-B viruses is mainly based on the sequence variability of the surface G attachment glycoprotein. For the multiple sequence alignment, 86 different G protein aa sequences of viruses from various geographical regions were retrieved from the NCBI database and aligned in order to select the most varying HRSV strains (data not shown). Sequences showing the lowest sequence similarity [%] were selected and aligned using the A2 as a reference strain, presented in Figure 2. The aa amino acid sequence homology between the G proteins of the prototypes A2 (RSV-A) and BA (RSV-B) is only 49.3 %.

Depending on the strain, the G protein has a lenghth of 292 to 315 aa. The G protein is a highly glycosilated type II transmembrane protein and has a single hydrophobic region near the N- terminal end (aa 38-63) that serves as an uncleaved signal peptide and a membrane anchor, leaving the C-terminal two-thirds of the molecule oriented externally.[25] The G protein is also expressed as a secreted form (sG) which helps HRSV to escape antibody-dependent restriction of replication by acting as an antigen decoy and by modulating the activity of leukocytes bearing Fcγ receptors. The difference between the full-length HRSV membrane bound G protein (mG), which is anchored by a transmembrane domain near the N terminus, and the secreted version (sG) is that the latter one lacks the transmembrane domain. This is due to the alternative initiation of translation at the second methionine (aa 48) in the open reading frame, followed by proteolytic processing to form a new N-terminus at amino acid 66.[26] Several studies indicate that approximately 80% of the total released G protein in cell culture belongs to the sG version and the remaining 20% is present as mG and is associated with the virion particles.[27]

The polypeptide backbone of G protein has a molecular weight of approximately 33 kDa. However, due to the extensive glycosylation of the mature protein, it migrates as a broad band between 40 to 90 kDa in SDS-PAGE.[28] Most of the carbohydrate moieties on the G protein are *O*-linked sugars at serine and threonine residues. For example, the A2 strain shows

25 positions of the serine and threonine *O*-linked side chains and 4 *N*-linked residues as demonstrated for the engineered F-G chimeric protein expressed in insect cells.[29]

Besides a very high amount of variable regions within the G sequence, a conserved part from 164 to 176 aa can also be found in both A and B genotypes. In a certain HRSV group, this conserved region can even be 10 aa longer (aa 164 to 186). Given its conserved nature, this central domain was initially assumed to play a role in the presumed major function of G protein, namely in the virus attachment. However, it was later found out that the deletion of this region did not lead to changes in the replication of a recombinant HRSV.[30] The conserved domain comprises a CX3C motif containing Cys-182 and Cys-186 that has a limited sequence relatedness with the CX3C domain of the chemokine fractalkine at amino acid positions 182-186. Fractalkine plays an important role in immunity. The interaction of this chemokine with its receptor CXCR1 affects cell trafficking and activation of T cells as well as cells of the innate immune system. It was reported that the G glycoprotein CX3C-motif imitates the chemokine fractalkine as it interacts with the specific CXCR1 and therefore competes with fractalkine for binding to its receptor. This G protein mimicry of chemokine fractalkine facilitates infection most likely due to the alteration of fractalkine-mediated immune response.[31] In an another study, synthetic peptides containing the sequence of the central conserved domain were shown to inhibit activation of the NF-κB transcription factor and secretion of inflammatory cytokines by human monocytes, indicating that this domain inhibits the host innate immune response to HRSV.[32] Within the C-terminal part of the conserved domain, sequencial and structural homology analysis showed close resemblance between this part of G protein and the fourth sub-domain of the 55-kDa type I tumor necrosis factor receptor (TNFr).[33] TNF appears to play a protective role in HRSV infection in culture and in mice.[34] Therefore, it is tempting to speculate that soluble G protein could function as a TNF antagonist but Teng *et al.* could show that this function does not have a significant effect on RSV replication in the case of its deletion.[35]

The next interesting region within the protein sequence is the heparan binding domain which partially overlaps with the indicated conserved region (Figure 2). HRSV-G protein has been proven to function as an attachment protein by S. Levine *et al.* who showed that binding of

HRSV particles to cultured cells could be inhibited using specific anti-G antibodies.[36.] However, no specific receptor that recognizes the G glycoprotein has been described yet. Recent studies revealed that HRSV could bind to immobilized heparin.[37, a] *In vivo*, heparin is primarily located in the granules of mast cells and basophils. Heparan sulfate, a related compound, is found on the surface of most mammalian cell types and in the extracellular matrix.[38] Many viruses, including herpesviruses, human immunodeficiency viruses, flaviviruses, picornaviruses and alphaviruses use heparan sulfate to mediate attachment and infection of target cells. Certain synthetic peptides, derived from consensus sequences of the G protein ectodomain from both RSV genotypes A and B were shown to bind heparin by Feldman *et al.* using heparin-agarose affinity chromatography. This study identified a single linear heparin binding domain (HBD) for RSV genotype A (184A-T198) and B (183K-K197).[39]

The last highlighted region shown in the protein alignment is a duplicated part of 20 aa in length (Figure 2). This region is encoded by a 60 nucleotide duplication region on genome which occurs only in BA subtypes (BA 1-10) of the B-group genotype. The BA subtype was firstly identified in Buenos Aires in 1999 and soon became the predominant strain in every country analyzed so far. It has now advanced to the point where almost 100% of the group B viruses identified after 2005-2006 compromise this 60 nucleotide duplication in the G protein gene. The process by which this duplicated region provides a selective advantage for the BA subtypes over the rest of the B genotype group viruses is unknown. It is possible that the additional residues altered significantly the antigenic characteristics of the G protein, allowing the virus to evade the immune system. An additional possibility is that the duplication has affected G protein-mediated attachment of the virus to the host cell and thereby enhanced the virus fitness.[40]

[a] Heparin belongs to the glycosaminglycans (GAGs). GAGs are the most abundant heteropolysaccharides in the body. They consist of long unbranched polysaccharides containing a repeating disaccharide unit. The disaccharide units contain one of the two modified sugars, N-acetylgalactosamine (GalNAc) or N-acetylglucosamine (GlcNAc), and an uronic acid such as glucuronate or iduronate. GAGs are highly negatively charged molecules, with extended conformation that impart high viscosity to the solution. These molecules are located primarily on the surface of cells or in the extracellular matrix (ECM). (http://themedicalbiochemistrypage.org/glycans.php).

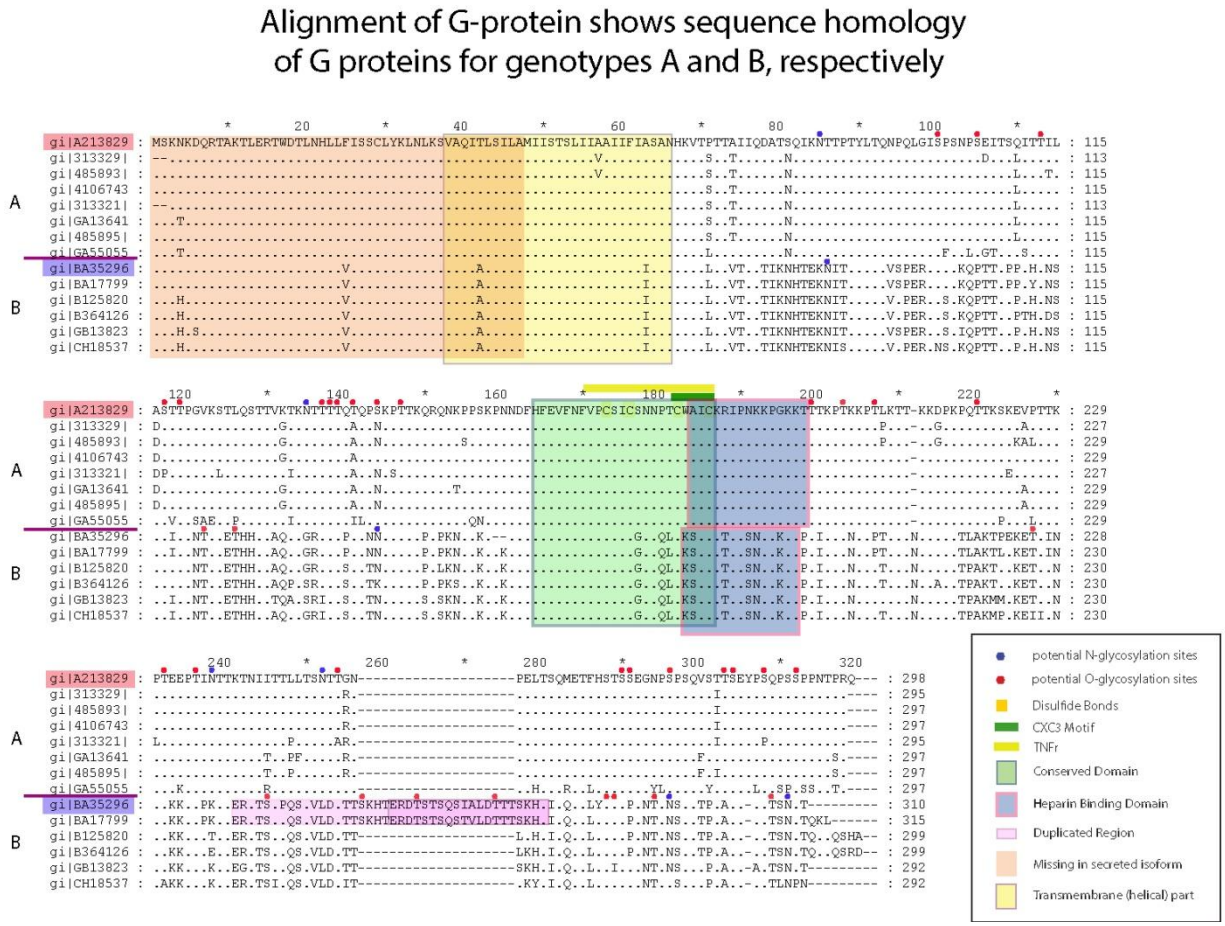


Figure 2: Amino acid sequence alignments of the G proteins from the most variable strains representing the two major antigenic HRSV subgroups, A and B. The A2 strain (subgroup A) was used as a master sequence and is placed on the top of the alignment. Accession numbers of the protein sequences are indicated on the *left side*. Genotypes A and B are separated with the *dark red* line (*above the line:* genotypes belonging to subgroup A; *below the line:* genotypes belonging to subgroup B), genotype accentuated *in red* indicates A2 strain; whereas genotype accentuated *in blue* indicates the BA strain. Dots show identical amino acids, letters display the amino acid differences and dashes indicate gaps. Conserved motifs, potential glycosylation sites and other specific regions are also shown in the alignment.

I.3.2. F-protein amino acid sequence analysis

HRSV type I transmembrane surface protein F is highly conserved among A and B genotypes with calculated amino acid homology of 89% between the selected prototype strains A2 (RSV-A) and BA (RSV-B) used in the sequence alignment shown in Figure 3 and 4.

The F protein is synthesized as an inactive polypeptide precursor of 574 aa and approximately 70 kDa molecular weight (F0) (Figure 3 and 4). This precursor F0 undergoes glycosylation on asparagine residues, homotrimerization and proteolytic processing in the *trans*-Golgi complex to yield two disulfide-linked subunits F1 and F2 (NH_2-F2-F1-COOH), before reaching the cell

surface as mature and active form of the fusion protein.[41, 42] Both protein subunits are framed in the alignment (Figure 3 and 4) and the predicted *N*-linked glycosylation sites for the A2 strain are also indicated. In contrary to HRSV-A, all HRSV-B strains have an asparagine residue at the aa position 120, which is also predicted to form a possible *N*-glycosylation site.[43]

The F1 subunit comprised two hydrophobic domains, one at the N-terminus, named fusion peptide, and the second near the C-terminus of the F1 chain called transmembrane region. Fusion peptide at amino acid positions 137-157 is known to be inserted directly into the host membrane, initiating the fusion process. The transmembrane segment found in the C-terminal region at the position 530-550 is followed by a 24 aa cytoplasmic tail acting as an anchor holding the peptide in the viral membrane. Furthermore, the F1 subunit contains two heptad repeat regions (HR-1 and HR-2) which associate during fusion, driving a conformational change to form a six-helix bundle structure that brings the viral and the host cell membranes into close proximity facilitating membrane fusion and allowing viral entry.[44, 45, 46]

A third hydrophobic domain within the F polypeptide, called signal peptide, is located at the N-terminus of the F2 chain within the amino acid positions 1-22. This signal peptide is needed for translocation into the endoplasmatic reticulum and is removed from the mature protein.[47] In the F2 subunit an additional heptad repeat, between residues 53 and 100 is shown (Figure 3). This heptad repeat region (HR-3) is known to form weak α-helical homodimers that do not appear to interact with either one of the two heptad repeat regions located in F1 protein subunit nor with the HR-1-HR-2-six-helix bundle complex. So the role of HR-3 in the F2 protein subunit has not yet been identified.[48] Two furin cleavage sites between the amino acid positions 109-110 and 136-137 are also indicated and those allow the formation of a short peptide of 27 aa, which function is unknown. The first cleavage recognition site consists of six tandem arginine and lysine residues. HRSV and BRSV are unique in *Paramyxoviridae* in possessing the second cleavage site that is further upstream and that also contains the consensus furin motif. Mutation analysis where no cleavages were processed at both sites revealed that this change seems to remove the capacity of the F protein to promote membrane fusion.[49]

Summarizing the functions of both protein subunits, it can be stated that F1 contains certain regions that mediate fusion with target cell membrane but only little is known about the F2

subunit function. Mutation studies of F2 subunit in other *Paramyxoviridae* family viruses resulted in altered fusion promotion activity indicating a possible stabilizing function of F2 subunit during membrane fusion.[50] Moreover F2 seems to be the key factor responsible for HRSV host specificity. This was shown by Schlender *et al*. based on chimeric F constructs in which the two F subunits were derived from different viruses (HRSV and BRSV) and demonstrated that the species specific entry was dependent on the F2 subunit alone.[51] Despite decades of research the specific receptor for HRSV is still not known. However, cell surface glycosaminoglycans (GAGs) and in particular heparan sulfates have been shown to be important for virus attachment to the host cell. Both G and F protein are able to bind GAGs but neither the surface protein G nor the SH protein are required for HRSV infectivity. On the other hand F protein alone is able to mediate attachment as well as fusion to the host cells as shown by recombinant HRSV lacking these genes. Although the F protein is necessary and sufficient for virus attachment and cell fusion, HRSV lacking the G protein is highly attenuated *in vivo as* those viruses without the G protein attached to cells with one third of the efficiency of those containing the G protein. Therefore, virions lacking the G protein show lower infectivity due to less efficient attachment to cells. All in all, the fusion F protein is independent of the G protein although its presence enhances binding and fusion activity of the F protein.[52, 53] Due to the critical role of the fusion protein in virus entry, along with the fact that F is highly conserved among HRSV groups and further on is one of the major antigens for neutralizing antibodies this protein seems to be the ideal target for development of therapeutic agents as well as of diagnostic tools.

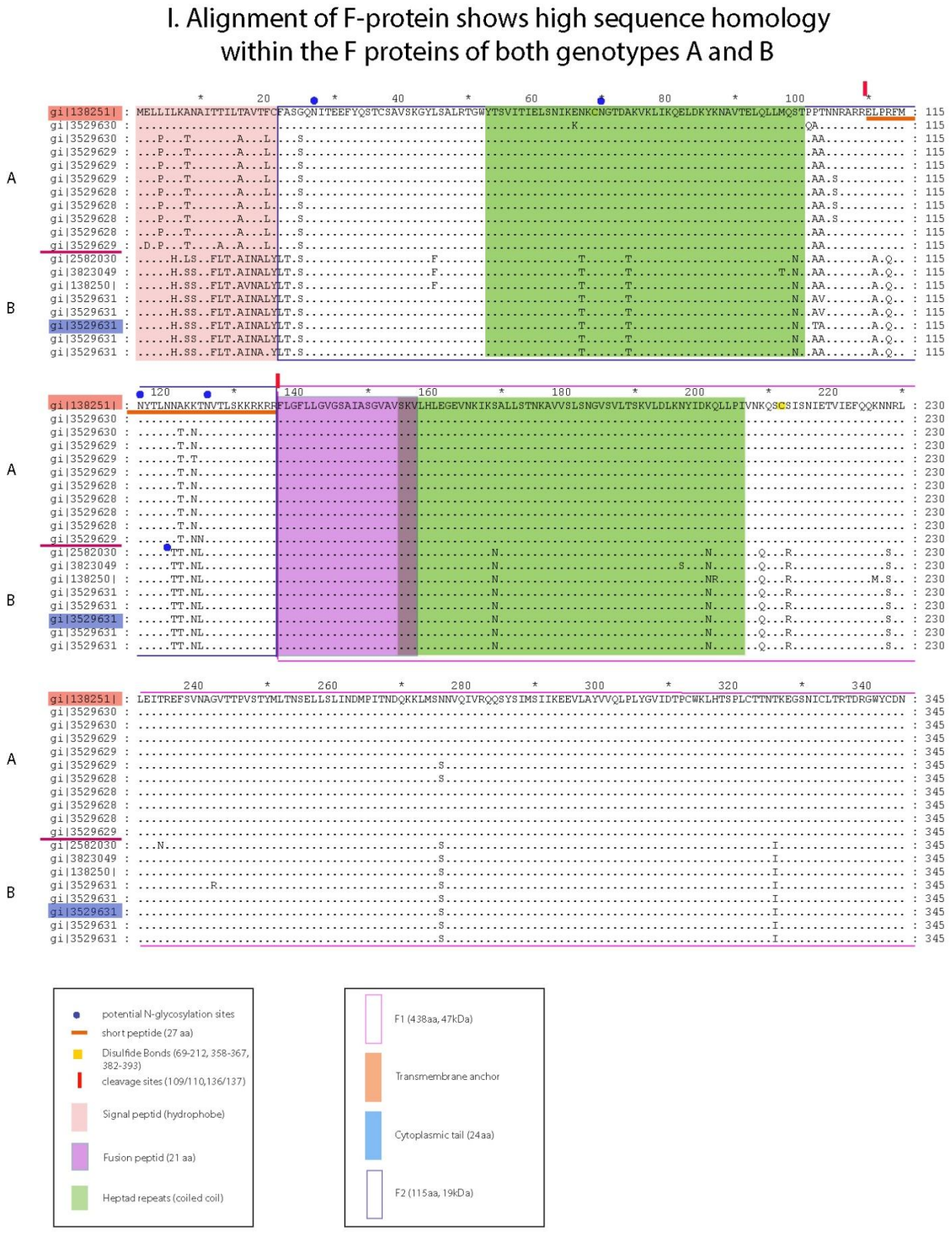


Figure 3: Amino acid sequence alignment of the F proteins of the most variable strains from the two antigenic HRSV subgroups, A and B. The A2 strain (subgroup A) was used as a master sequence and is placed on the top of the alignment. Accession numbers are indicated on the *left side*. Genotypes A and B are separated with the *dark red* line (*above the line:* genotypes belonging to subgroup A; *below the line:* genotypes belonging to subgroup B), genotype accentuated *in red* indicates A2 strain; whereas genotype accentuated *in blue* indicates the BA strain. Dots show identical amino acids, letters display the amino acid differences and dashes indicate gaps. Protein subunits, potential glycosylation sites and other specific regions are also shown in the alignment.

II. Alignment of F-protein shows high sequence homology within the F proteins of both genotypes A and B

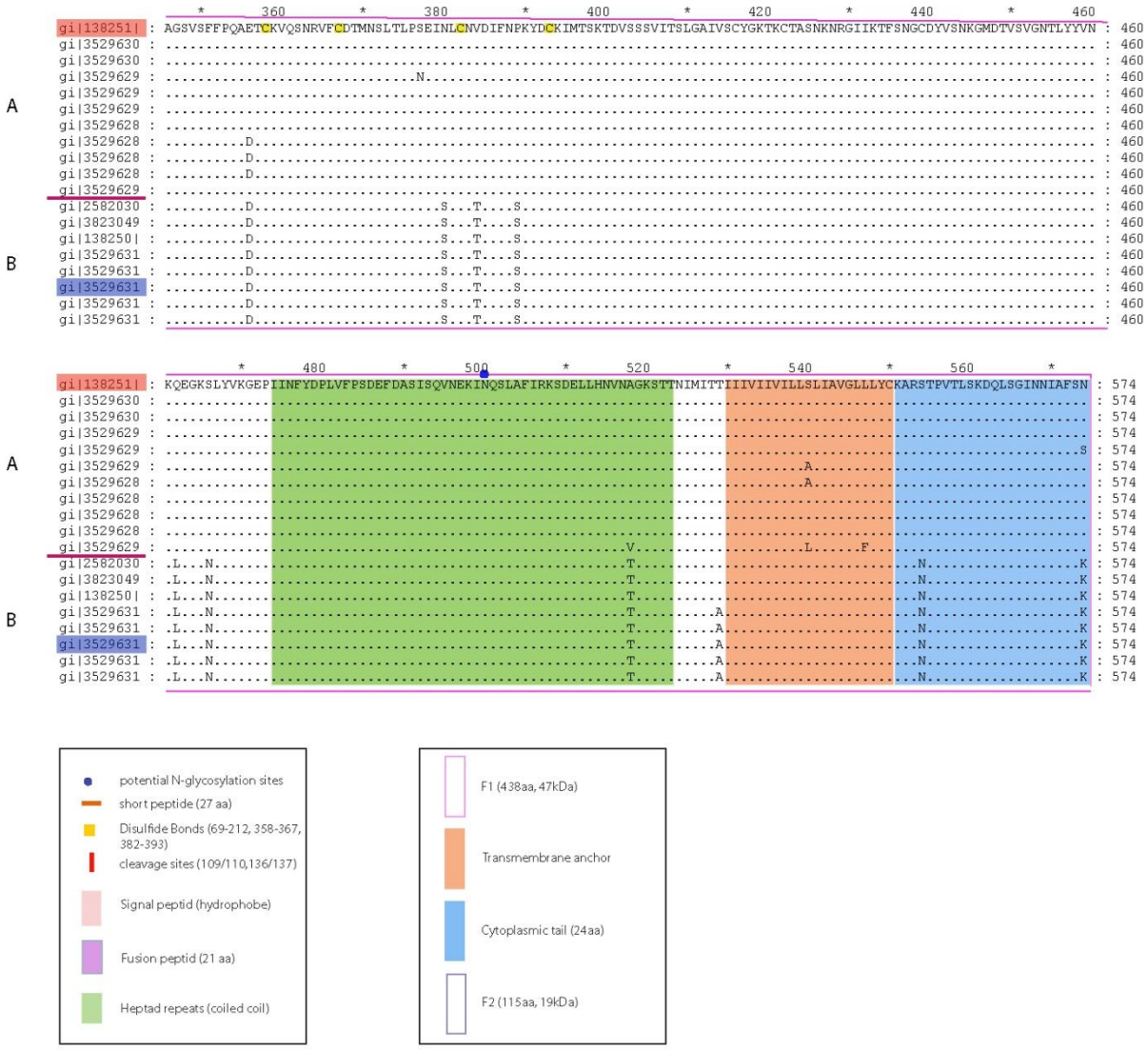


Figure 4: Continued amino acid sequence alignment of the F proteins of the most variable strains from the two antigenic HRSV subgroups, A and B. Complete figure legend is on previous page.

I.4. Methods of protein purification and biophysical characterization

Here the theoretical background of the techniques used in this work for producing, purifying, identifying and characterization of recombinant proteins as well as synthetic peptides will be described.

I.4.1. Ni-NTA affinity chromatography

Ni-NTA affinity chromatography relies on the affinity of histidine's imidazole ring for nickel ions. These are immobilized on a resin with Nitrilotriacetic acid (NTA) and bind proteins containing more than 3 consecutive histidines. The purification of His-tagged proteins consists of 4 stages, i. e. cell lysis, binding, washing and elution which will be described in the method section. Due to the fact that histidine residues in the 6x-His tag have a pK_a of ~ 6.0 they protonate if the pH become acidic. Under these conditions the 6xHis-tagged protein loses its affinity to the nickel ions and can therefore be eluted from the Ni-NTA resin. Analogically, if the imidazole concentration is raised up to 100-250 mM, the 6xHis-tagged proteins are no longer competitive for binding sites on the Ni-NTA ions and will therefore also dissociate from the resin.[54]

A very common problem that occurs during Ni-NTA affinity chromatography is the co-purification of *E. coli* proteins along with His-tagged recombinant proteins. Contaminants include proteins that contain multiple histidine residues or molecular chaperones that may bind to the resin directly or to the recombinant protein.[55] Most of these co-purified proteins are present in *E. coli* strains of different genetic backgrounds, also in BL21 (DE3) strain used in this work. The majority of *E. coli* contaminants are stress-responsive proteins such as Fur, Crp, ArgE, SlyD, GlmS, GlgA, ODO1, ODO2, YadF and Yf bG. The binding of these proteins to Ni-NTA resin can mostly be explained by their native metal-binding functions or by the presence of surface clusters of histidine residues.[56] A table containing proteins which are commonly co-purified during protein purification by Ni-NTA affinity chromatography, is shown below (Table I).

Native proteins from *E. coli* commonly co-purified during IMAC					
Protein	SwissProt access code	Molecular Mass (kDa)	% Histidine residues	Isoelectric point (pI)	Metal requirement
Fur	P06975	16.7	8.1	5.6	Fe^{2+}, Zn^{2+} [a]
YodA	P76344	22.3	5.2	5.6	Cd^{2+}, Zn^{2+} [a]
Cu-Zn-SODM	AAC74718	17.6	4.0	5.9	Cu^{2+} [a], Zn^{2+} [a]
ArgE	P23908	42.3	4.4	5.5	Fe^{2+}, Ni^{2+}
YadF	P36857	25.0	5.5	6.1	Zn^{2+}, Hg^{2+} [a,b]
GlgA	P08323	51.7	3.4	6.0	Mg^{2+} [a,c]
GlmS	P17169	66.8	3.9	5.5	
CAT	AAA57080	25.5	5.5	5.9	Co^{2+}
Crp	P03020	23.6	2.9	8.3	
Hfq	P25521	11.1	4.9	6.9	
SlyD	P30856	20.8	10.2	4.8	Zn^{2+}, Ni^{2+}
S15	P02371	10.2	5.6	10.4	
YfbG	P77398	74.2	4.1	6.3	
Hsp60	AAC77103	57.0	0.2	4.8	
ODO1	P07015	10.5	3.6	6.0	
ODO2	P07016	44.0	1.7	5.5	
G6PD	P22992	55.7	1.2	5.5	

Table I: Commonly co-purified native *E. coli* proteins during Ni-NTA immobilised metal affinity chromatography. Protein names, molecular masses, isoelectric points and the histidine residues percentages are indicated.[57]

I.4.2. Peptide synthesis

In the case of solid-phase peptide synthesis the C-terminus of the first amino acid, instead of C-terminal protection with a chemical group, is coupled to a polystyrene (PS) resin onto which long chains of polyethylene glycol (PEG) are attached. The resin acts as the C-terminal protecting group and provides a rapid method to separate the growing peptide product from the different reaction mixtures during synthesis.

Peptide synthesis ensues by coupling the carboxyl group of the incoming amino acid to the N-terminus of the growing peptide chain. This C-to-N synthesis is opposite from protein biosynthesis in which N-to-C peptide elongation takes place. The addition of amino acids to the growing peptide chain occurs in a precise, step-wise and cyclic manner until full-length peptide is formed. All in all the peptide synthesis consists of three following consecutive steps: peptide deprotection, amino acid coupling and peptide cleavage. Because amino acids have multiple reactive groups, peptide synthesis must be carefully performed to avoid side reactions that can reduce the length and cause branching of the peptide chain. To facilitate peptide formation with minimal side reactions the amino acid N-termini are protected by Fmoc. The N-terminal protecting group has than to be removed by deprotector solution in order to permit peptide bond formation between the exposed N-terminus of the first amino acid and the activated C-terminus of the incoming amino acid. Synthetic peptide coupling requires after a deprotection step, the activation of the C-terminal carboxylic acid on the incoming amino acid using activating group Triazoles (HOBt).

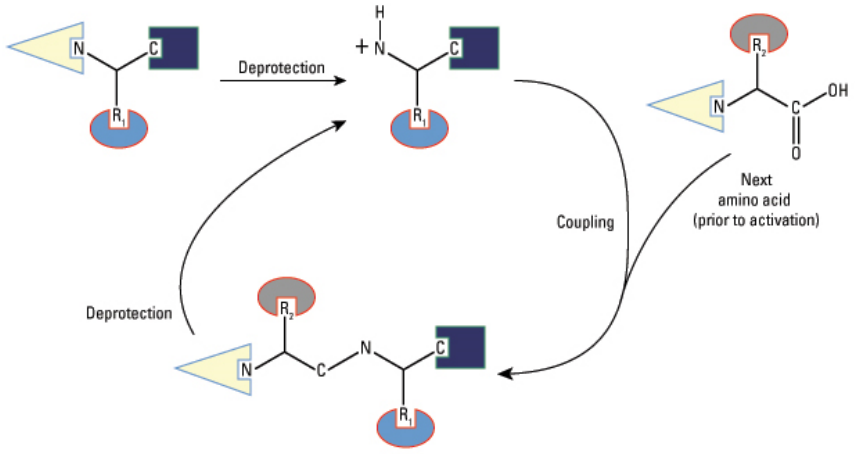


Figure 5: Diagram of peptide synthesis. The N-terminal protecting group on the C-terminal amino acid of the peptide to be synthesized is first deprotected. After removing of unbound protecting groups, the next amino acid is activated at the C-terminal end by a coupling agent, which facilitates peptide bond formation between the deprotected N-terminus of the first amino acid and the activated C-terminus of the incoming amino acid. The newly synthesized N-terminus of the growing peptide is then deprotected and coupled to the next amino acid. This cycle of deprotection and coupling is repeated until the full-length peptide is formed. (http://www.piercenet.com/method/peptide-synthesis)

After successive cycles of amino acid deprotection and coupling which are performed by the machine automatically, all remaining protecting groups (side chain protecting group as tBut and resin) had to be removed from the created peptide in solution. Therefore, peptides should be washed with dichlormethan and cleaved from the resin in 95% TFA, 2.5% Silane and precipitated into *tert*-Butylmethylether. Unfortunately, events such as incomplete deprotection or reaction with free protecting groups can cause truncated sequences, isomers or other side products. These events can occur at any step during peptide synthesis and lead to impurity of the peptide precipitate. Therefore, peptide purification by reverse-phase HPLC in an acetonitrile gradient can be performed to separate by-products.[58]

During reverse-phase HPLC in an acetonitrile gradient continuous partitioning of the molecules between the mobile phase and the hydrophobic stationary phase takes place. Hydrophobic molecules adsorb onto a hydrophobic solid support in a polar mobile phase (mixture of 10% pure ACN and 90% of 0.1% TFA diluted in ddH_2O). Decreasing the mobile phase polarity by adding more organic solvent (linear ACN gradient 10-70%) reduces the hydrophobic interaction between the analyte molecule and the solid support resulting in desorption. A peptide with a larger hydrophobic surface is retained longer on the solid support and will need higher concentration of ACN that is necessary to cause desorption. On

the contrary peptides with higher polar surface area are less retained as they are better integrated into water and desorbs at lower ACN concentration. Therefore different fractions can be collected and separated by their different milli absorption units (mAU).[59]

I.4.3. MALDI-TOF mass spectrometry

MALDI is a gentle ionization technique in which a short laser pulse is used to ionize molecules. MALDI is attached to a time of flight (TOF) analyzer which detects the time it takes for created ions to travel a fixed distance. A protein or peptide sample is placed on a target plate and mixed with an appropriate matrix directly on the plate. The matrix absorbs the ultraviolet light and converts it to heat energy, causing that the sample molecules and the matrix enter gas phase. This leads to release of samples molecules, matrix and ions from the target plate. Following acceleration through the electric field, the generated ions drift through a field-free path and finally reach the detector. Ion masses (mass-to-charge ratios [m/z]) are typically calculated by measuring their TOF, which is longer for larger molecules than for smaller ones. Hence, the detected time of flight of a molecule is proportional to its mass/charge ratio.[60]

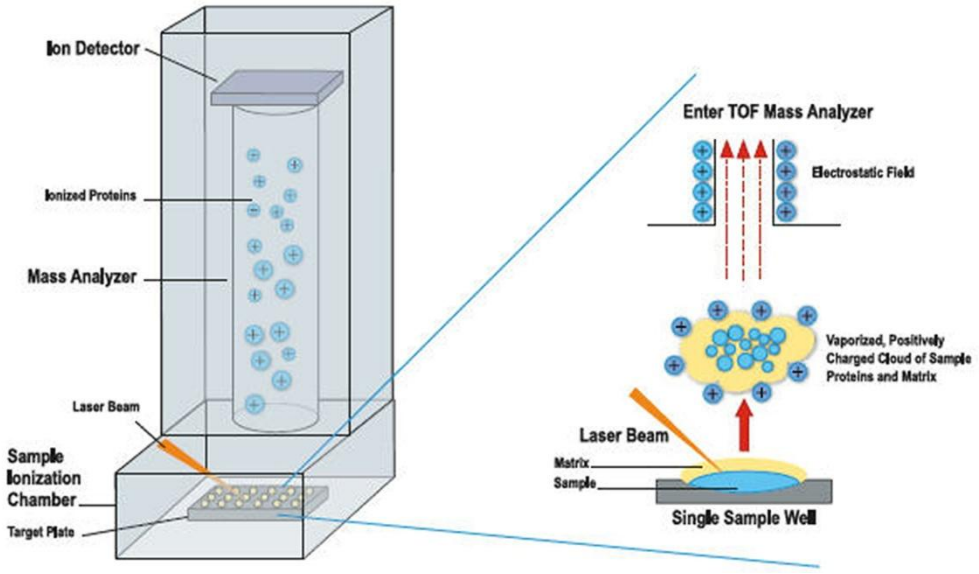


Figure 6: Illustration of MALDI-TOF mass spectrometer. A protein sample is embedded in a matrix on a target plate which gets ionized by a short laser beam. An electrical field accelerates the ions formed through the flight tube towards the detector. Ions with the lowest mass/charge ratio will have the highest velocity and arrive first at the detector. (http://www.mayomedicallaboratories.com/articles/communique/2013/01-maldi-tof-mass-spectrometry/index.html)

I.4.4. Circular dichroism (CD)

Briefly, circular dichroism is defined as the unequal absorption of left-handed (E_L) and right-handed (E_R) circularly polarized light. When asymmetric molecules interact with light, they may absorb right and left handed circularly polarized light to different extents (hence the term circular dichroism) and also have different indices of refraction for the two waves. The result is that the plane of the light wave is rotated and that the addition of the E_R and E_L vectors results in a vector that traces out an ellipse and the light is said to be elliptically polarized. CD is reported either in units of ΔE, the difference in absorbance of E_R and E_L by an asymmetric molecule, or in degrees of ellipticity, which is defined as the angle whose tangent is the ratio of the minor to the major axis of the ellipse.[61]

CD spectra in the far-UV (190-260nm) can be used to predict the percentages of each secondary structural element in the structure of a protein. Some of the common protein secondary structural elements and the CD spectra associated with them are shown in Figure 7.

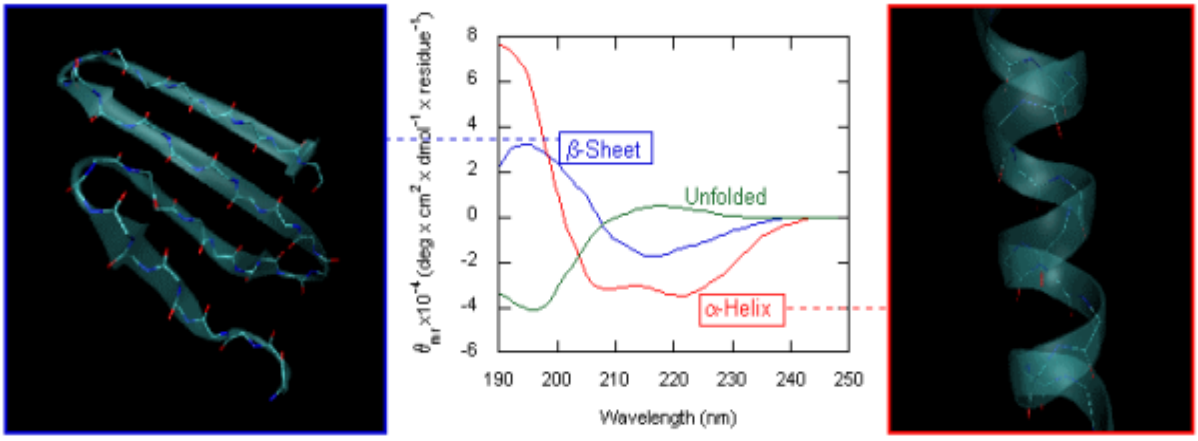


Figure 7: The secondary structure conformation of a protein/peptide and the CD spectra of its structural elements. Right is an example of the backbone conformation of a protein in its α -helix structure and left is the conformation of a protein in a ß-sheet. In the centre are the corresponding CD spectra for these different conformations.
(University of California, Irvine; Department of Chemistry; www.chem.uci.edu/.../CD%20spectroscopy.pdf)

I.4.5. SDS-PAGE

SDS-PAGE is one of the most widely used methods for analyzing proteins. This technique separates proteins according to their size and therefore is used to determine the molecular masses of proteins, to monitor protein purity and in conjunction with Western blot analysis, to identify the presence of proteins in complex mixtures.[62]

Sodium dodecyl sulfate (SDS) is an anionic detergent which denatures secondary and non-disulfide-linked tertiary protein folding states, and imparts a negative charge to each protein in proportion to its molecular weight. Without SDS addition, diverse proteins with similar masses would migrate differently due to conformational variability, as differences in folding patterns allow some proteins to better fit through the gel matrix than others. Treatment with SDS solves this problem, as it linearizes the proteins. Thus, proteins are separated according to their mass-charge ratio alone. Besides the addition of SDS, proteins are heated up to 95°C in the presence of a reducing agent, such as ß–Mercaptoethanol or DTT.[63] Those reducing agents promote the protein denaturation by reducing disulfide linkages and ensuring that no tertiary or quaternary protein folding remains (oligomeric subunits). This procedure is known as "reducing SDS-PAGE", which is used most commonly.[64] Most SDS-PAGE gels in the presented master thesis have been run under reducing conditions, meaning that ß–Mercaptoethanol has been incorporated into the SDS-containing loading buffer. The loading of isolated proteins on

a non-reducing SDS-PAGE (SDS-containing loading buffer without addition of ß–Mercaptoethanol and boiling) provides an easy way to determine if a protein of interest might be covalently associated with another protein within a cell. [65]

Regardless of the used system, the SDS-PAGE gel preparation requires casting two different layers of polyacrylamide between two glass plates. Therefore, a porous "stacking gel" with lower polyacrylamide content above the primary "separating gel" is prepared. The upper layer (stacking gel) includes the sample wells. It is designed to help focus the different proteins into sharp bands before they reach the separating gel. The separating gel, also called resolving gel, is used to separate the proteins based on their mass weight. It has a higher polyacrylamide concentration, making the gel`s pores more narrower. This classic Laemmli system needs further on a discontinuous buffer system where the pH and ionic strength of the buffer used for running the gel (Tris pH 8.8) is different from the buffers used in the stacking gel (Tris, pH 6.8) and separating gel (Tris, pH 8.3).[66]

I.4.6. Western-Blot

Western blot is often used in research to separate and identify proteins. In this technique a mixture of proteins is separated based on molecular weight through gel electrophoresis. Gels are then transferred to an adsorbent membrane, producing a band for each protein, under the influence of an electric current. The membrane is then incubated with labelled antibodies specific for the protein of interest (in this work antibodies specific for His-tag were used).

The unbound antibody is washed off leaving only the bound antibody to the protein of interest. The bound antibodies are then detected by secondary antibodies labelled with an enzyme such as alkaline phosphatase. This chromogenic reporter enzyme reacts with a substrate to yield an intensely coloured visible product. As the antibodies only bind to the protein of interest, only one band should be visible. The intensity of the band corresponds to the amount of protein present; thus doing a standard can indicate the amount of protein present.[67]

II Aims of the thesis

The main focus of the present study was to construct, express and purify recombinant F and G proteins of Human Respiratory Syncytial Virus (HRSV) using technologies of gene- and peptide-synthesis. The mentioned proteins are the only viral elements that are known to induce HRSV neutralizing antibodies and are therefore, attractive targets for developing new diagnostic tools and therapeutic agents.

Most of the work focused on the achievement of a controlled and high-level expression of the His-tagged proteins in *Escherichia coli*; the identification of the optimal time of induction and the optimization of both the most efficient solubilization procedure and Ni-NTA affinity chromatography purification. For the physicochemical characterization, the objective was to evaluate the molecular mass, fold and thermal stability. An additional part of the present thesis was focused on the amino acid sequence analysis of the F and G protein derived from two antigenic HRSV groups, HRSV-A and HRSV-B, with the aim to identify the conserved structural motifs and domains as well as potential sites for antibody binding.

III Materials and Methods

III.1. Materials

Gold buffer (5 L):
37,5 g (0,05 M x 5) Na_2HPO_4
5 g (0,006 M x 5) NaH_2PO_4
25 g BSA (0.5%)
2,5 g NaN_3 (0.05%)
25 ml Tween 20 (0.5%)
→ add 5 L ddH_2O

Blotting buffer (5 L):
15,13 g (25 mM x 5) Tris
72,06 g (192 mM x 5) Glycin
1L Methanol (20%)
→ add 5 L ddH_2O

Alkaline phosphatase staining solution (AP) (1 L):
12,1 g (100 mM) Tris/HCl
5,8 g (100 mM) NaCl
1 g (5 mM) $MgCl_2$
→ add 1 L ddH_2O; pH 9.5

Electrophoresis buffer (Running Buffer) (10x):
60,6 g (0,25 M) Tris
292 g (2,5 M) Glycin
20 g SDS (1%)
→ add 2 L ddH_2O

Laemmli sample buffer (4x):
10 mM Tris/HCl; pH 6.8
10% ß-Mercaptoethanol
4% SDS 1,96 g
0.2% Bromphenol bue 100 µl
20% Glycerin 11,5 ml 85% Glycerine
→ add 50 ml ddH_2O

50% Acrylamid solution:
370 g Acrylamid
10 g Bisacrylamid
→ add 760 ml ddH_2O; mixed until solved and filtered through a paper filter

4x Lower Tris (1 L):
363,3 g Tris (3M)
4 g SDS (0.4%)
→ add 1 L ddH_2O; pH 8.85

4x Upper Tris (1 L):
60,6 g (0,5M) Tris
4 g SDS (0.4%)
→ add 1 L ddH_2O; pH 6.8

10% APS (Amonium- peroxidsulfat):
5 g of APS
→ add 50 ml ddH_2O

Stacking gel (5%, 4 gels)
1 ml 50% PAA
2,5 ml Upper Tris, pH 6.8
6,5 ml ddH_2O
20 µl TEMED, 200 µl APS

Separating/ Resolving gel (12.5%, 4gels)
6,25 ml 50% PAA
6,25 ml Lower Tris, pH 8.85
12,5 ml ddH_2O
30 µl TEMED, 300 µl APS

Coomassie Staining Solution (1 L):
1 g Coomassie brilliant blue
0,5 L Methanol
0,1 L Acetic acid
→ add 1 L ddH_2O

Coomassie Destainer (1 L):
0,3 L Methanol
0,1 L Acetic acid
→ add 1 L ddH_2O

Imidazol Solution (0,5 L):
0,85 g (25 mM) Imidazol
0.1% Triton
pH 7.4

Triton X-100 Solution (100 ml):
6% Triton X-100
2,23 g (60 mM) EDTA
8,766 g (1,5 M) NaCl
pH 7.0

Lysis buffer under denaturing conditions 6 M Guanidin HCl (1 L):
15,6 g (100 mM) NaH_2PO_4
1,2 g (10 mM) Tris
573 g (6 M) GuHCl
pH 8.0

Lysis/Wash or Elution buffer under denaturing conditions 8 M Urea (1 L):
15,6 g (100 mM) NaH_2PO_4
1,2 g (10 mM) Tris
480,5 g (8 M) Urea
pH 8.0 (Elution buffer pH 6.5/ 5.5/ 5.0/ 4.5/ 4.0 and 3.5)

Lysis buffer under native conditions (1 L):
7,8 g (50 mM) NaH_2PO_4
17,54 g (300 mM) NaCl
0,68 g (10 mM) imidazole
pH 8.0

Wash buffer for protein purification under native conditions (1 L):
7,8 g (50 mM) NaH_2PO_4
17,54 g (300 mM) NaCl
1,36 g (20 mM) imidazole
pH 8.0

Elution buffer for protein purification under native conditions (1 L):
7,8 g (50 mM) NaH_2PO_4
17,54 g (300 mM) NaCl
X g (50 mM-250 mM) imidazole
pH 8.0

Protein storage buffers:

- **A.** 10 mM NaH_2PO_4, pH 5.5
- **B.** 100 mM NaH_2PO_4, 10 mM Tris, pH 11.2
- **C.** 100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl; pH 6.5
- **D.** 100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl; 10% Sucrose, pH 6.5
- **E.** 100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl; 10% Glycerol, pH 6.5
- **F.** 100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl; 5 mM ß-Mercaptoethanol, pH 6.5
- **G.** 1x PBS: 10 mM NaH_2PO_4, 10 mM Tris, 140 mM NaCl; 2,7 mM KCl, 1,8 mM KH_2PO_4, pH 7.3

LB-Medium (1 L):
10 g NaCL
10 g Pepton
5 g Yeast extract
→ add 1 L ddH_2O
pH 7.2-7.4 ; after autoclaving 500 µl Kanamycin of 100 mg/ml stock solution (final conc. 50 µg/ml)

Kan. Agar-Plates (1 L):
10 g NaCL
10 g Pepton
5 g Yeast extract
→ add 1 L ddH_2O and adjust pH 7.2-7.4;
15 g Agar/Agar
After autoclaving add 500 µl Kanamycin of 100 mg/ml stock solution (final conc. 50 µg/ml)

Thermo Scientific PageRuler Plus Prestained Protein Ladder (10-250 kDa)

Antibodies
Mouse mAb IgG1 Anti 6-His-Tag (0,2 mg/ml in PBS pH7.4); Clone: 13/45/31. Dianova (Hamburg)
Rat Anti-Mouse mAB IgG1 conjugated with Alkaline Phosphatase; Clone: X56. BD Bioscience "Pharmingen"

Nitro blue tetrazolium (NBT), stock solution of 50 mg/ml:
0,1 g NBT dissolved in 2 ml 70% DMF solution and stored at 4°C

5-Bromo-4-chloro-3-indolyl phosphate (BCIP), stock solution of 50 mg/ml:
0,1 g BCIP dissolved in 2 ml 100% DMF and stored at 4°C

Phenylmethylsulfonylfluorid (PMSF)
10 mg/ml (57 mM) stock solution solved in isopropanol

Kanamycin (stock: 100 mg/ml)

DNaseI (stock: 5 mg/ml)

Activator-HBTU (0,2 L):
0,5 M HOBt (Hydroxybenzotriazole) dissolved in 200 ml DMF afterwards addition of
0,5 M HbTU (O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate)

Deprotector-Piperidine 20% (1L):
200 ml Piperidine (Sigma) + 800 ml NMP

Activator base-DIEA (N,N-Diisopropylethylamine), Roth

DCM (Dichloromethane), Roth

DMF (Dimethylformamide), Roth

NMP (N-Methyl-2-pyrrolidone), Roth

TFA (Trifluoro acetic acid)

III.2. Methods

III.2.1. Sequence analysis of G and F protein

In order to compare and find the most common subtypes for both HRSV genotypes, HRSV-A and HRSV-B, complete amino acid (aa) sequences of G and F proteins from available HRSV strains, were retrieved from NCBI and Expasy (Uniprot) databases. A multiple protein sequence alignment with the collected sequences in a "Fasta" format was performed using Clustal W tool (www.ebi.ac.uk/clustalw) and a Genedoc software (www.nrbsc.org/gfx/genedoc). Once multiple protein sequence alignment was executed, the most different sequences were used to perform a further multiple sequence alignment. On the basis of this alignment information, it was decided which subtypes are suitable for the production of recombinant proteins and peptide synthesis. The accession numbers of the top sequences used for the alignments were derived from UniProtKB/Swiss-Prot-Expasy: GLYC_HRSVA (accession number: P03423) and FUS_HRSV (accession number: P3420). In addition, the aa sequences of the G and F proteins were analyzed using two prediction programs: ProtParam (www.expasy.ch/tools/protparam) which provided general protein characteristic information and PsiPred (bioinf.cs.ucl.ac.uk/psipred) which delivered the data of predicted secondary structure of the proteins.

III.2.2. Plasmid transformation of *E. coli*

The pET-27b (+) plasmid (ATG: biosynthetics, Merzhausen, Germany) was used for the generation of recombinant HRSV G protein and the two HRSV F protein subunits, F1 and F2 (Figure 8). Amino acid sequences of the proteins and their accession numbers were retrieved from UniProtKB/Swiss-Prot-Expasy: GLYC_HRSVA (accession number: P03423) and FUS_HRSV (accession number: P3420). The cDNAs were codon-optimized for the expression in *Escherichia coli* and the constructs were inserted into 5' *NdeI* and 3' *XhoI* unique restriction sites of the pET-27 (+) vector. Each of the three inserts contained a hexa-histidine tag followed by a stop codon at the 3' end (Figure 8). In order to confirm the correct identity of the

constructs, *NdeI/XhoI* single and double restriction analysis followed by DNA sequencing was performed (Eurofins MWG Operon, Ebersberg, Germany).

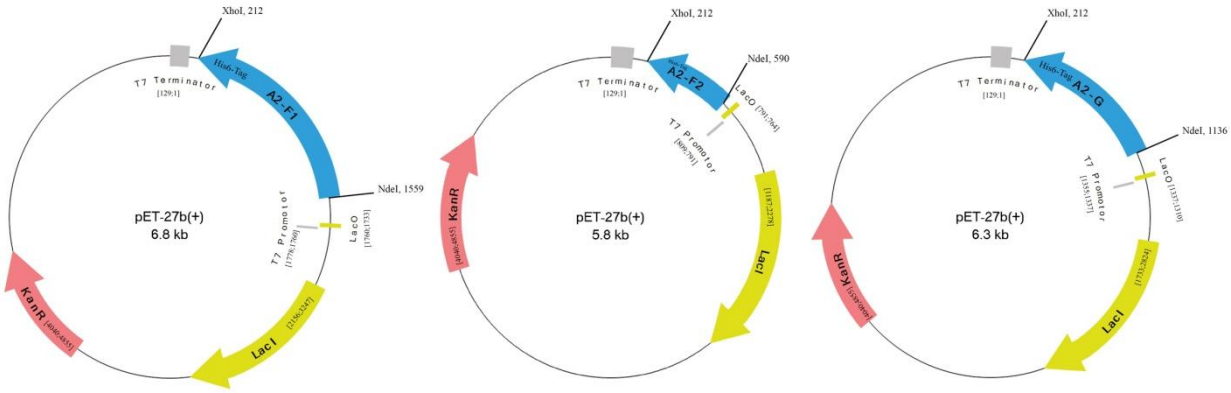


Figure 8: pET-27b (+) vectors containing cDNAs coding for 6x-His-tagged HRSV F protein subunits F1 and F2, and the HRSV G protein (from the left). Following features of the plasmid are shown: vector size (kb) in the middle of each vector, Kanamycin resistant gene *in red*, T7 promotor/terminator *in grey*, LacO and LacI *in yellow* and the inserted cDNA sequences coding for F1, F2 or G proteins between the two unique restriction sites *XhoI* and *NdeI in blue*.

All plasmids containing the synthetic genes were transformed into BL21-Gold (DE3) *Escherichia coli* competent cells (Agilent Technologies, USA). For each transformation (G, F1 and F2), 20 µl of the competent cells were used. Cells were thawed on ice for 15 minutes. Then, approximately 300 ng/2 µl of plasmid-DNA were added to the cells, gently mixed and incubated 30 min on ice. Afterwards, each of the transformation reaction was heat shocked at 42°C for 30 seconds to enable the uptake of the plasmid DNA and placed back on ice for 2 minutes. Subsequently, 250 µl of SOC medium was added to each transformation tube and placed for 1 hour at 37°C in a shaking incubator (600 rpm). This step allowed the bacteria to regenerate for the subsequent growing on the Kanamycin (50 µg/ml) containing LB-agar plates. After overnight incubation at 37°C, single colonies were selected for further experiments.

III.2.3. Expression of recombinant proteins in *E. coli*

For the expression experiments of the F protein (F1 and F2 subunits), 50 ml pre-cultures of LB-medium supplemented with 50 µg/ml Kanamycin were grown from single colonies at 37°C in a shaking incubator (200 rpm, ON). On the next day, 1 L of LB-medium containing Kanamycin (50 µg/mL) was inoculated with the pre-culture to a starting OD_{600nm} value of 0.1.[b] Cells were grown at 37°C with agitation (180 rpm) until OD_{600nm} between 0.5 and 0.6 was reached. At that point, the expression of the target protein was induced by adding IPTG to a final concentration of 1 mM/L. To determine the optimal induction period for the expression of F1 and F2 proteins, 1 ml samples from the *E. coli* cultures were taken before and 1h, 2h, 3h , 4h, 18h following IPTG induction for monitoring of protein expression.

Expression of the recombinant G protein was performed using a single and freshly isolated colony from a LB-agar plate. The 1 L culture was grown at 37°C and 180 rpm until the cells reached midlog growth (OD_{600nm} of 0.5) and was then transferred into two sterile Erlenmeyer flasks. The expression of the target protein was induced by the addition of IPTG (1 mM/L) and the incubation was performed either at 30°C or 25°C, respectively. In order to determine the optimal condition for the G protein expression, 1 ml samples were removed from the cultures (30°C and 25°C) before and 1h, 2h, 3h ,4h, 18h after IPTG induction and analyzed by SDS-PAGE.

III.2.4. Determination of protein solubility

Parallel to the time-course analysis of protein expression, 50 ml samples for the analysis of protein solubility were taken after 2h, 4h and 18h following 1 mM IPTG induction. These samples were centrifuged (10 min, 4°C, 4000 rpm) and the cell-pellets were frozen and thawed three times on ice using liquid nitrogen (N_2). This and the two following steps lead to the breakage of the cell in order to make the recombinant proteins accessible. Next, each

[b] Calculation of needed cell culture amount to reach a starting OD_{600nm} value of 0,1: The 1:5 Dilution of culture had an OD_{600nm} value of 0.524 (0,524*5=2,62). → 2,62*X ml=0,1*1000 ml → 38,17 ml of pre-culture were used to inoculate 1 L of LB-medium.

thawed cell-pellet was re-suspended in 5 ml native lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and further homogenized with an Ultra-Turrax type device (Janke & Kunkel-IKA Labortechnik, Staufen, Germany) by three times 30 seconds bursts. From each cell lysates, 1 ml samples were taken for the SDS-PAGE analysis (CL: clear lysate). The remaining 4 ml of each of the homogenized cell solution were centrifuged (20 min, 4°C, 18000 rpm) to obtain supernatants containing soluble proteins (Sol.: soluble) and the pellets containing cell membrane debris, unbroken cells, nucleic acids and insoluble proteins. Finally, each pellet was again re-suspended in 5 ml of native lysis buffer (Insol.: insoluble). Ten µl of each of the cell lysate fraction, soluble and insoluble protein fraction were mixed with 10 µl of 4 x SDS-sample buffer (4x, 10mM Tris/HCl, 10% ß-Mercaptoethanol, 4% SDS, 0.2% Bromphenol blue 20% Glycerine), denatured (95° for 5 min) and analyzed by SDS-PAGE and Western blotting.

III.2.5. Purification of recombinant F1 and F2 protein-subunits by Ni-NTA affinity chromatography under denaturing conditions

Recombinant F1 and F2 protein subunits were purified from insoluble fractions of a 1 L *E. coli* expression culture using an inclusion body preparation protocol. For this purpose, cells were harvested by centrifugation (20 min, 4°C and 4000 rpm) 2 hours (F1) and 4 hours (F2) after induction with 1 mM IPTG and frozen at -20°C until use. The cell pellet was thawed on ice and re-suspended in 20 ml of 25 mM imidazole, 0.1% Triton X-100, pH 7.4 and mechanically lysed with an Ultra-Turrax (3 x 30 sec bursts and cooled down on ice between bursts). In order to degrade DNA and make the lysate more liquid, 200 µg of *DNaseI* were added to the cell lysate and incubated at RT for 30 minutes.[c] Afterwards, 7 ml of 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0 were added in order to increase the permeabilisation of the membranes and the mixture was incubated for half an hour on ice. Next, inclusion bodies were separated from soluble cell components (SN1) by centrifugation (20 min, 4°C and 18000 rpm). Before freezing the obtained supernatant (SN1) at -20°C an 40 µl aliquot was taken for SDS-PAGE analysis. The obtained pellet containing insoluble aggregates and was re-suspended in 3,5 ml 6% Triton X-

[c] In the case of F1 protein purification also PMSF (2 mM end concentration) was added to inhibit protease. In all following resuspending steps of obtained pellets PMSF was added again at final concentration of 2 mM (SN1-SN3).

100, 60 mM EDTA, 1.5 M NaCl, pH 7.0. After centrifugation (20 min, 4°C and 18000 rpm) gained supernatant (SN2) was handled as SN1 and the pellet was dissolved in 3,5 ml of 25 mM imidazole, 0.1% Triton X-100, pH 7.4. Subsequently, a final centrifugation step was executed and the third supernatant (SN3) was obtained. The pellet containing the insoluble inclusion bodies was solubilised by adding 5 ml of lysis buffer (6 M GuHCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0) and incubated overnight at RT. On the next day mixture was centrifuged (20 min, 4°C and 16000 rpm) to separate insoluble aggregates (pellet) from solubilised recombinant protein (supernatant). Finally the solubilised inclusion body fraction was mixed with 2 ml of equilibrated Ni-NTA agarose (Quiagen GmbH, Hilden, Germany), binding capacity 5-10 mg/ml, and incubated for 8 hours (F1) or for 4,5 hours (F2) at RT. This step enabled the binding of Ni-NTA to the 6xHis-tag to the recombinant protein. Next, the protein-Ni-NTA resin mixture was loaded onto column (QIAGEN Polypropylene Columns 5 ml, Hilden, Germany) and the flow-through (FT) fraction was collected drop by drop.

In the case of F1 protein purification, the column was washed first with 15 ml wash buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 6.4) and washed further with the same buffer whereby 2x2 ml and 1x1 ml fractions were collected. The pH of 6.4 in the wash buffer compared to the pH of 8.0 in the lysis buffer allowed the elution of non-tagged contaminating *E. coli* proteins. Subsequently, the recombinant F1 protein was eluted stepwise by decreasing the pH values of elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 4.5- pH 4.0). Twelve fractions of 1 ml volume were obtained at ph 4.5 and further 5 fractions of the same volume were collected at pH 4.0 elution. All fractions obtained during washing and elution steps were collected and analyzed by SDS-PAGE and Western-blot. Pure protein fractions were pooled together and dialysed in a Slide-A-Lyzer dialysis cassette, cut-off of 3.5 kDa (Thermo Scientific, Pierce Protein Biology Products, Rockford, USA) slowly by stepwise decreasing urea content against dialysis buffer (100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 4.5) in order to get rid of the urea and to allow the protein to refold. Finally the F1 protein solution was dialysed against storage buffer (10 mM NaH_2PO_4, pH 4.5). After dialysis, the protein solution was concentrated by centrifugation (~6 min, 4000 rpm, 4°C) using Amicon Ultra-15 centrifugal filter tubes (Merck Millipore, Tullagreen, Carrigtwohill Co., Ireland) with cut-off of 30 kDa.

During F2 protein purification the column was washed twice with 10 ml of the wash buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 6.1). Then the recombinant F2 protein was

eluted stepwise by decreasing the pH values of the elution buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl), 5 fractions with 1,5 ml volume were collected after elution with pH 4.5 and 30 fractions with 1 ml volume were collected after elution with pH 3.5.

Pure protein fractions were pooled together and dialysed in a Slide-A-Lyzer dialysis cassette, cut-off of 3.5 kDa (Thermo Scientific, Pierce Protein Biology Products, Rockford, USA) against ddH_2O in order to get rid of the urea and to allow the protein to refold. After dialysis, the protein solution was concentrated by centrifugation (~8 min, 4000 rpm, 4°C) using Amicon Ultra-15 centrifugal filter tubes (Merck Millipore, Tullagreen, Carrigtwohill Co., Ireland) with cut-off of 10 kDa.

Protein concentration (F1 and F2) was measured using BCA Protein Assay (Novagen, Darmstadt, Germany) in a micro-plate format. Bovine serum albumin (BSA) (2 mg/ml) was used for the preparation of standard dilutions and the assay was measured at 562 nm.

III.2.6. Purification of recombinant F1 protein-subunit by Ni-NTA affinity chromatography under native conditions

Cells were harvested by centrifugation (20 min, 4°C and 4000 rpm) 2 hours after the addition of 1 mM IPTG and frozen at -20°C until use. Due to the fact that the recombinant protein was found in the insoluble fraction after the lysis of *E. coli* cells (Figure 9B), protein purification under native conditions was done with the inclusion body preparation protocol as described above. The cell pellet containing inclusion bodies was dissolved in 10 ml of lysis buffer containing 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0. In order to avoid protein degradation, the inclusion body solution was incubated overnight at 4°C with agitation and the addition of PMSF (2 mM). Separation of solubilized recombinant protein (supernatant) from insoluble aggregates (pellet) was achieved by centrifugation (20 min, 4°C and 16000 rpm). From this time point, all following steps were performed at 4°C. The supernatant containing the recombinant protein was transferred into a pre-treated dialysis tube (Spectrum Medical industries; Inc. Spectra/Por- Molecular porous membrane tubing MWCO 6-8 kDa. Housten, Texas, USA.) and dialyzed against a native lysis buffer (50 mM NaH_2PO_4, 300 mM

NaCl, 10 mM imidazole, pH 8.0), allowing the protein to refold. After dialysis, clear lysate was mixed with 2 ml of equilibrated Ni-NTA agarose (Quiagen GmbH, Hilden, Germany)[d] and incubated for the next 8 hours with agitation. This step enabled the binding of Ni-NTA to the 6xHis-tag on the recombinant F1 protein. Next, the protein-Ni-NTA resin mixture was loaded onto a polypropylene 5 ml column (QIAGEN, Hilden, Germany) and a flow-through (FT) fraction was collected drop by drop. This step was repeated twice. Afterwards, two washing steps were performed, first with native lysis buffer (3 x 2 ml, 1 x 1 ml) and the second with native wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imodazole, pH 8.0), where 2 x 2 ml and 1 x 1 ml fractions were collected. For the elution of the protein, the concentration of imidazole increased in the elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 60-100 mM imodazole, pH 8.0). First 7 fractions of 1 ml volume were collected after elution with 60 mM imidazole (50 mM NaH_2PO_4 , 300 mM NaCl, pH 8.0), following 7 eluates of 1 ml volume with 80 mM imidazole concentration in buffer and finally, 7 more fractions with 100 mM imidazole concentration in buffer. Pure F1-protein fractions were pooled together and dialysed in a Slide-A-Lyzer dialysis cassette (cut-off of 3.5 kDa) or Spectra/Por molecular porous dialysis tubes with a cut-off of 6-8 kDa (Spectrum Medical Industries, Inc., Houston, Texas, USA) against 10 mM NaH_2PO_4, pH 7.0. Amicon Ultra-15 centrifugal filter tubes (Merck Millipore, Tullagreen, Carrigtwohill Co., Ireland) with cut-off of 30 kDa were used to concentrate the protein solution by centrifugation (~10 min, 4000 rpm, 4°C). The protein concentration was measured in a micro-plate procedure Bradford Protein Assay kit (Bio-Rad Laboratories GmbH, München, Germany) and in addition, in micro-plate format by BCA Protein Assay kit (Novagen, Darmstadt, Germany). Bovine serum albumin (BSA) (2 mg/ml) was used for preparation of standard dilutions as recommended by kits and measured at 562 nm and 595 nm, respectively.

[d] Ni-NTA binding capacity 5-10 mg/ml.

III.2.7. SDS-PAGE and Western blot

Purity of recombinant proteins was determined by Coomassie Brilliant Blue staining of a 12.5% SDS polyacrylamide gel and the protein identity was confirmed by Western-blotting.

SDS-PAGE gel preparation

First the glass plate sandwich was assembled and checked for tightness with water. Then the resolving gel solution was prepared as described above in the material section (SDS-PAGE-separating gel). APS and TEMED was added last and mixed carefully to avoid formation of bubbles. Because polymerization begins as soon as APS is added to the mixture, all subsequent actions were performed promptly. The gel solution was poured between the glass plates but approximately 1/4 of the space was left free for the stacking gel. In order to separate the stacking gel from resolving gel in a straight linear boarder the top of the resolving gel was carefully covered with 80% isopropanol. After the separating gel polymerized the isopropanol solution was discarded and the glass plate sandwich with separating gel was washed gently with ddH_2O. Then, the prepared stacking gel solution (prepared as described above, APS and TEMED was added last) was poured carefully onto the polymerized resolving gel layer and the combs were inserted. After the upper gel polymerized the combs were removed carefully and the gel was transferred to the electrophoresis tank (Whatman/ Biometra, Göttlingen, Germany), filled bottom and top reservoirs with electrophorese buffer (1x working solution, 25 mM Tris-Cl, 250 mM glycine, 0.1% SDS, pH 8.3).

SDS-PAGE sample preparation

For this purpose culture sample or protein purification samples were mixed in a 1:1 ratio with Laemmli sample buffer (4x, 10mM Tris/HCl, 10% ß-Mercaptoethanol, 4% SDS, 0.2% Bromphenol blue 20% Glycerine of 85% Glycerol), denatured using a thermo cycler at 95° for 5 min and centrifuged (5 min, 14000 rpm for protein expression samples, short spin for all other samples).

The denatured proteins in the mixture were subsequently applied (18 µl per sample) onto the 12.5 % SDS-PAGE gel submerged in electrophorese buffer (1x working solution, 25 mM Tris-Cl,

250 mM glycine, 0.1% SDS, pH 8.3). In order to determine the molecular weight of unknown proteins, 4 µl of PageRuler Plus Prestained Protein Ladder (Thermo Scientific, Pierce Protein Biology products, Rockford, USA) were also loaded onto one gel lane. An electric current was applied across the gel, causing the negatively-charged proteins to migrate across the gel towards the cathode. After approximately 1 hour at 25 mAmp/gel the proteins migrated differentially based on their size; smaller proteins as F2 travelled farther down the gel, while larger ones (F1, G) remained closer to the point of origin. Following electrophoresis, the gel was stained by Coomassie Brilliant Blue staining solution (1% Coomassie Brilliant Blue, 0,5 L Methanol, 0,1 L Acetic acid, 0,4 L ddH_2O), allowing visualisation of the separated proteins or the gel was processed further for Western blotting.

Western-Blot

After separating the protein mixture by SDS-PAGE, the gel was transferred to a wet nitrocellulose membrane (Whantman-GE Healthcare Life Science, Vienna, Austria) soaked in blotting buffer (25 mM Tris, 192 mM glycine, and 20% methanol). Membrane was placed between the gel surface and the positive electrode in a sandwich. The sandwich included a fibre pad (sponge) at each end, and filter papers (Chromatography paper; Whatman-GE Healthcare Life Science, Vienna, Austria) to protect the gel and blotting membrane. In order to ensure a clear image, a close contact of gel and membrane and the placement of the membrane between the gel and the positive electrode were carried out. The membrane was placed in a kind so that the negatively charged proteins could migrate from the gel to the membrane. This type of transfer is called electrophoretic transfer and was executed in wet conditions using blotting buffer (25 mM Tris, 192 mM glycine, and 20% methanol) in a blotting unit (Mighty Small Transpho, Hoefer Inc., San Francisko, USA). The transfer was done using an electric field (240 mAmp, for exact 1 hour) oriented perpendicular to the surface of the gel, causing proteins to move out of the gel and to the membrane.

In order to prevent non-specific binding of the antibodies, the nitrocellulose membrane with proteins was blocked three times for 15 minutes, on a shaker at RT with Gold buffer (0,05 M Na_2HPO_4, 0,006 M NaH_2PO_4, 0.5% BSA, 0.05% NaN_3, 0.5% Tween 20). Subsequently, a 1:1000 dilution of anti-His-tag antibodies (Dianova/ Mouse mAb IgG_1 Anti 6-His-Tag; Clone: 13/45/31. Hamburg, Germany) diluted in Gold buffer was incubated with the membrane over night at

4°C with agitation. Then, again washing steps with Gold buffer were carried out (3 x 15 min) to remove unbound antibodies. Next, a 1:1000 dilution of alkaline phosphatise- conjugated antibodies (BD Bioscience/Pharmingen/ Rat Anti-Mouse mAB IgG1 conjugated with Alkaline Phosphatase; Clone: X56; Heidelberg, Germany) diluted in Gold buffer were incubated with the membrane for 2 hours on a shaker at RT. After this, final washing steps were performed at the same conditions and because of the same reason as mentioned before. Finally, protein detection could be done by the addition of 10 ml substrate alkaline phosphatase staining solution (100 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) comingled with 66 µl of NBT (50 mg/ml) and 33 µl of BCIP (50 mg/ml) solutions. This mixture yielded an intense, insoluble black-purple band if reaction with alkaline phosphatase, conjugated on secondary antibody-anti-His-tag-6xHis-tagged protein complex, on the washed nitrocellulose membrane occurs. When the protein signal could be observed, the reaction was stopped by discarding the substrate solution and rinsing the membrane with water.

III.2.8. Peptide synthesis and purification by HPLC

Synthetic peptides spanning the complete G protein were synthesized by solid phase synthesis with the 9-fluorenyl-methoxy-carbonyl-method, using PEG-PS preloaded resins (Tentagel resins, e.g. TentaGel™ S PHB-Arg(Pbf)Fmoc; Sigma-Aldrich, Austria) and a microwave Peptide Synthesizer CEM-Liberty (CEM Corp., Matthews, NC, USA). First, the required amount of all needed N-terminal Fmoc[e] protected amino acids (Nova-Biochem/Merck, Schuchardt OHG, Hohenbrunn, Germany) was calculated, weighted and dissolved in NMP.[f] Then, the appropriate resin with the correct C-terminal amino acid was weighted and dissolved in 3 ml

[e] The amino acid N-termini are protected with Fmoc (9-fluorenylmethoxycarbonyl) that blocks non-specific reaction during synthesis. The C-terminus of the C-terminal amino acid of the peptide is also protected to facilitate peptide extension in the correct orientation. To allow peptide bond formation these protecting groups are relatively easy to remove from the newly added amino acid (deprotection) just after coupling to allow the next incoming amino acid to bind to the growing peptide chain in the proper orientation. Fmoc is a base-labile protecting group that is removed with a mild base such as piperidine.

[f] *N*-Methyl-2-pyrrolidone (NMP) efficiently solvates the resin and can improve the yield of coupling. Most common peptide reagents are very soluble in NMP.

of NMP. Further reagents such as Activator-HBTU[g], Deprotector-Piperidine 20%[h], Activator base-DIEA, DCM and DMF were prepared.

After successive cycles of amino acid deprotection and coupling which were performed automatically by the machine, all remaining protecting groups (side chain protecting group as tBut and resin) had to be removed from the synthesized peptide. This cleavage step was carried out by hand. The resin- NMP solution was passed through a glass filter so that resin stayed in the filter-tube. Then the resin was washed 3 times with 25 ml DCM to eliminate NMP and to dry the resin. Afterwards, the dried resin was transferred into a falcon tube and the mentioned groups were cleaved by acidolysis, using TFA[i] (19 ml) and scavenger (1 ml). Water and Triisopropylsilane in a 1:1 ratio were used as scavengers, which reacted with free protecting groups. This resin-TFA-Triisopropylsilane mixture was incubated for 2 hours at RT with agitation. Next, peptide precipitation by ether was carried out. Briefly, the mixture passed through a new glass filter whereby the resin was separated from the TFA-peptide solution. Ice cold tert-Butylmethylether (25 ml) was added to the TFA-peptide solution in order to induce precipitation of the peptide. This mixture was incubated on ice for 2h.[j] Subsequently, the precipitate was harvested by centrifugation (10 min, 4°C and 3800 rpm). The gained ether supernatant was removed and the peptide precipitate was dissolved again in 25 ml pre-chilled tert-Butylmethylether to remove remaining cleavage reagents (TFA). These centrifugation and dissolving steps in ether were repeated two more times. After final centrifugation, gained peptide precipitate was dried at RT and air supply.

Synthesized peptides were further purified by reversed-phase HPLC using Dionex HPLC-UltiMate 3000 pump (Thermo Fisher Scientific Inc., Vienna, Austria) in a 10-70% acetonitrile gradient. For this purpose, approximately 50 mg of the dried peptide were dissolved in 2 ml

[g] Synthetic peptide coupling requires the activation of the C-terminal carboxylic acid on the incoming amino acid using 1-hydroxybenzotriazole (HOBt) together with (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU). HOBt activates the α-caboxyl group and reduce the risk of racemization. In the presence of HOBt and HBTU (helps to avoid side product formation) the rate of reaction is very rapid and the reaction is efficient with minimal side product formation.

[h] Deprotector solution (20% Piperidine in DMF) is a mild base and is used to remove the Fmoc group from the N-terminus of the growing peptide chain and permits the next amino acid to be coupled.

[i] Trifluoroacetic acid (TFA) is used at high concentration as a strong acid to remove the remaining protecting groups after synthesis of full-length peptide chain.

[j] The resin was collected again and the cleavage with TFA and Silane was executed a second time with the same succession and reagents.

10%-30% ACN and 0.1% TFA[k] and injected onto HPLC-pump. The separation of the peptide from by-products involved continuous partitioning of the molecules between the mobile phase and the hydrophobic stationary phase. The collected fractions were separated based on their milli absorption units (mAU) and each fraction was then analyzed by MALDI-TOF mass spectrometer. Fractions containing high intensity peaks corresponding to the correct molecular weight of the peptides were pooled together and after freezing with liquid N_2, lyophilized using CHRiST-Alpha 2-4 LSC Lyophilizator (SciQuip; Newtown, Wem, Shropshire, UK).

III.2.9. MALDI-TOF analysis of synthetic G protein peptides and of the F2 protein subunit

Laser desorption mass spectra (synthetic peptides and F2 protein) were acquired with a Microflex MALDI-TOF mass spectrometer (Bruker Daltonics Inc, Billerica, MA). The protein sample was dialyzed against ddH_2O as described in the F2 protein purification paragraph. For sample preparation a 1:1 mixture of purified peptide/protein and a matrix solution (α-Cyano-4-hydroxycinnamic acid dissolved in 70% ACN, 0.1% TFA for peptides and sinapinic acid dissolved in 50% ACN, 0.1% TFA for proteins, respectively)[l] was deposited onto a target and air dried. Acquired spectra were analyzed by Bruker Daltonics FlexAnalysis software.

[k] Acetonitrile (ACN) was used as a solvent. Its low viscosity and low chemical reactivity make it a popular choice for high-performance liquid chromatography (HPLC). Trifluoroacetic acid (TFA) at a low concentration was used as an ion pairing agent in liquid chromatography (HPLC) of peptides.

[l] α-Cyano-4-hydroxycinnamic acid and sinapinic acid are commonly used as matrix in MALDI mass spectrometry of a wide variety of peptides and proteins, respectively. They serve well as a matrix for MALDI due to their ability to absorb laser radiation and to also donate protons to the analyte of interest.

III.2.10. Circular dichroism (CD) analysis

Circular dichroism spectra of the recombinant F2 protein were measured on a Jasco J-180 spectropolarimeter (Japan Spectroscopic Co., Tokyo, Japan) using a high transparency 1 mm light path quartz cuvette adjusted to 25°C and washed with ddH_2O. For each measurement, 200 µl of protein sample dissolved in ddH_2O at a concentration of 0.1 mg/ml were used. Spectra were recorded at 25°C between 190 and 260 nm, with a resolution of 0.5 nm and at a scan speed of 50 nm/min. Three scans of each measurement (3 x ddH_2O, 3 x protein sample) were performed and the average of the obtained data was used for the analysis. The final spectra were baseline corrected by subtracting the corresponding water spectra obtained under identical conditions. Results were expressed as the mean residue ellipticity [θ] at a given wavelength. Temperature scans were performed according to a step-scan procedure, where the sample was heated stepwise at 2°C/min from 25°C to 95°C and finally cooled down to a starting point at the same rate. Three spectra were recorded every 5°C, and averaged. The secondary structure conformation of the F2 protein was calculated using the secondary structure estimation algorithm CDSSTR with Reference Set number 4, optimized for the wavelength range 190-260 nm as well as algorithms SELCON3 and CONTIN using DichroWeb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml). In addition, results from the CD measurements were compared to the sequence-based predictions of the protein secondary structure performed by PsiPred (bioinf.cs.ucl.ac.uk/psipred).

IV Results

IV.1. Expression, solubility and purification of recombinant F1 protein

IV.1.1. Expression and solubility screening of F1 protein

Figure 9A shows the time-course analysis of the expression of the F1 protein. When protein expression before and at different time points after induction with IPTG was compared, expression levels after 2 to 3 hours were found to be the highest as indicated by broader and more intense protein bands migrating at the correct molecular weight of 48,8 kDa (Figure 9A). No protein of interest could be detected before the addition of IPTG confirming that the expression was IPTG-dependent. Based on this result the induction period of 2h was considered to be the most optimal and was used for the subsequent protein purification experiments. In order to confirm the correct identity of the F1 protein, a Western-Blot using monoclonal anti-His-tag antibodies which reacted specifically with the over-expressed protein, was also performed (data not shown). Next, the solubility of the F1 protein under native conditions was analysed and the results are shown in Figure 9B. Unfortunately, the protein of interest could only be detected in the insoluble fraction of the cell lysates. In order to ensure that the cells were completely lysed, the pellet was additionally treated with 6% Triton X-100 to solubilise the protein in its native state and EDTA was added to facilitate the permeabilization of the outer membrane of *E. coli* cells (data not shown). Nevertheless, the protein remained insoluble under this condition (Figure 9B).

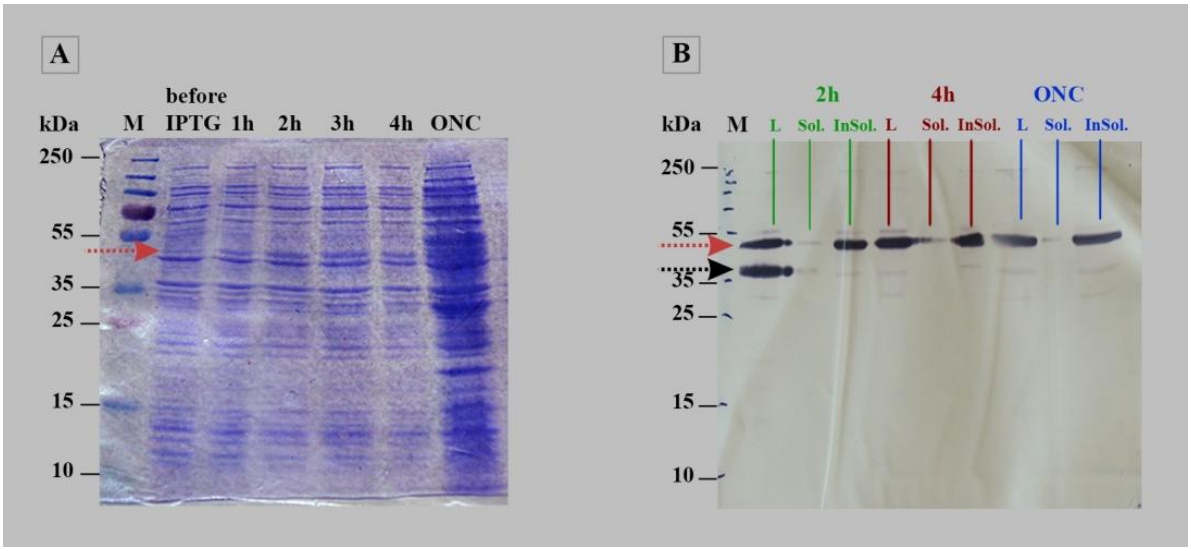


Figure 9: Time-course analysis after F1 protein expression and solubility. **A)** Coomassie blue-stained gel containing aliquots taken **before** and **1**, **2**, **3**, **4, 18** (overnight culture= **ONC**) hours following 1 mM IPTG induction. Molecular weights (kDa) are indicated on the *left margin* and the *red arrow* shows the position of F1 protein at approximately 48.8 kDa. **B)** Western blot, stained with anti-His-tag antibodies, containing fractions for the analysis of the F1 protein solubility. Coloured lines show the position of the loaded samples (**L**: lysate, **Sol.**: soluble fraction, **InSol.**: insoluble fraction at different time points (**2h**: green, **4h**: red, **ONC:** blue) after 1 mM IPTG induction. Molecular weights (kDa) are indicated on the *left margin*. The first slot contains additionally a VP1 protein at 37 kDa used as a positive control, indicated by a *black arrow*.

IV.1.2. Purification of recombinant F1 protein subunit under denaturing conditions

Due to the fact that the F1 protein could almost exclusively be detected in the insoluble fraction of the cell lysates (Figure 9B), purification under denaturing conditions was performed primarily. Figure 10 shows results obtained after purification of the recombinant protein from 1 L *E. coli* culture and 2 hours incubation time using an inclusion body preparation protocol followed by Ni-affinity chromatography. Because of the low expression level of the F1 protein, Western blots were always performed in addition to Coomassie blue-stained SDS-PAGE in order to verify the presence or absence of the F1 protein. With the exception of supernatant 3 (SN3) which contained slight amounts of the protein, the majority of the F1 protein was found in the fraction containing solubilized inclusion bodies (CL) (Figure 10B, left picture). Although washing fraction I (WI) contained a protein band corresponding to the correct molecular weight in Coomassie blue-stained polyacrylamide gel (Figure 10A, left picture), this band could not be confirmed by a Western blot (Figure 10B, left picture). Elution of the protein was achieved by decreasing the pH value of the elution buffer. His-tagged F1 protein was found in the first 4 elution fractions and its identity was verified using a monoclonal anti-his-tag antibody (Figure 10A and 10B, right pictures). Furthermore, protein bands migrating at the molecular weight level of approximately 13 kDa were additionally observed when a Western blot analysis was performed (Figure 10B, right picture). These bands could be explained by either fragmentation of the huge F1 protein subunit during purification steps or the co-purification of one of the native *E. coli* proteins (Table I).

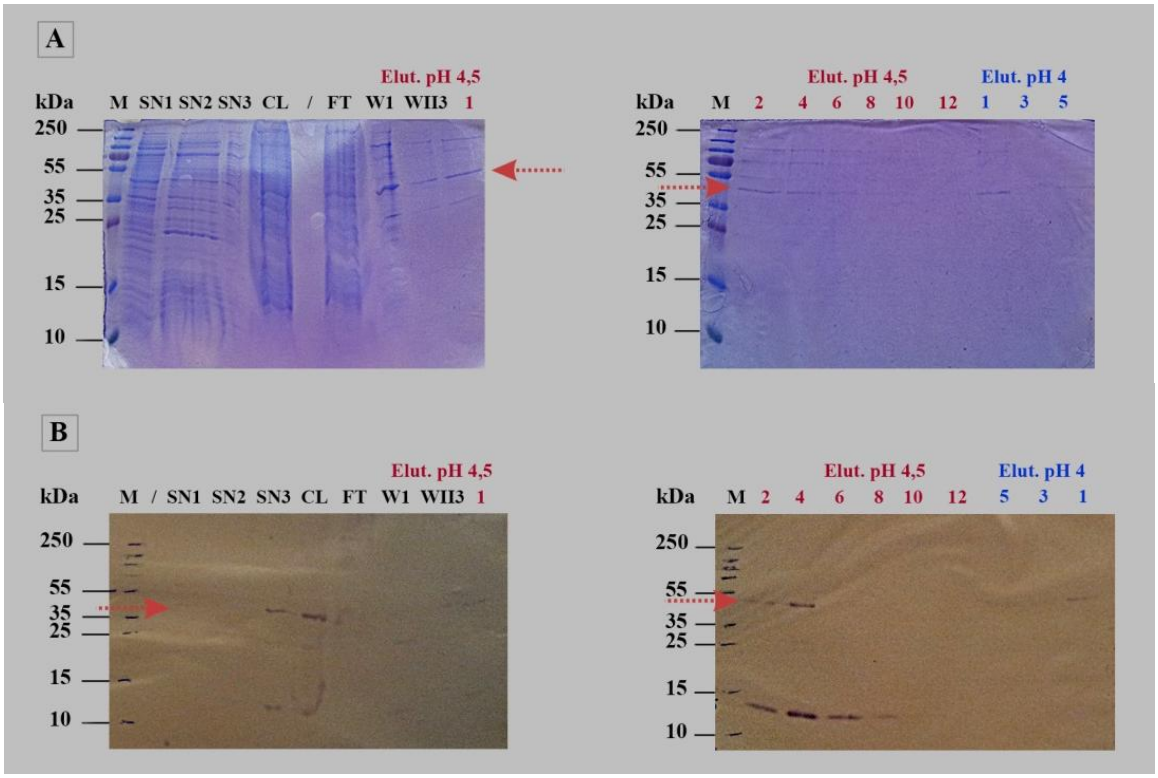


Figure 10: Purification of the His-tagged F1 protein using Ni-NTA chromatography under denaturing conditions. **A)** Coomassie blue-stained gels and **(B)** Western blot, stained with anti-His-tag antibodies, containing fractions obtained during inclusion body preparation: **SN1:** supernatant after cell lysis in 25 mM imidazole, 0.1% Triton-X-100, pH 7.4, treatment with DnaseI, PMSF and 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN2:** soluble fraction after washing with 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN3**: soluble fraction after washing with 25 mM imidazole, 0.1% Triton-X-100, pH 7.4; **CL:** clear lysate containing solubilized inclusion bodies in 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0; **FT:** flow through of the CL-Ni-NTA-mixture; **WI, WII3:** wash fractions with 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 6.4; **E 1-12 *in red*:** eluates at pH 4.5. **E 1, 3, 5 *in blue***: eluates at pH 4.0. The first lane of each gel was loaded with the molecular weight marker (M) as indicated on the *left side* and *red arrows* show the position of the F1 protein at approximately 48,8 kDa.

IV.1.3. Purification of recombinant F1 protein subunit under native conditions

In order to obtain higher yields of the recombinant F1 protein, purification under native conditions was also performed (Figure 11). *E. coli* cells were harvested by centrifugation after 2h incubation period and the inclusion body pellet was solubilized overnight using lysis buffer containing 8 M urea. After separation from other insoluble components, the F1 protein containing solution was dialysed against native lysis buffer in order to allow the protein to refold. Next, the protocol for the purification *via* Nickel-affinity chromatography under native conditions was used and all subsequent purification steps were performed at 4°C. Although a protein band corresponding to the molecular weight of F1 protein was again observed in washing fractions, this could not be detected by monoclonal anti-His-tag antibodies (data not shown). Elution of the recombinant protein was performed by increasing the concentration of imidazole in the elution buffer (Figure 11B, 11C). The clearest fractions were obtained when the elution buffer contained 80 mM imidazole. These data were also confirmed by a Western blot analysis using monoclonal anti-His-tag antibodies which reacted specifically with a protein band at 48.8 kDa (Figure 11D). No other impurities were observed in the eluted fractions. The unspecific binding of other proteins to the Ni-NTA resin in the elution fractions 1-5 (60 mM imidazole) might be explained by a too long incubation time of the Ni-NTA-protein solution-mixture (Figure 11B). However, shorter incubation periods (4h, 2h and 45 min) did not increase the purity of the eluted fractions (data not shown).

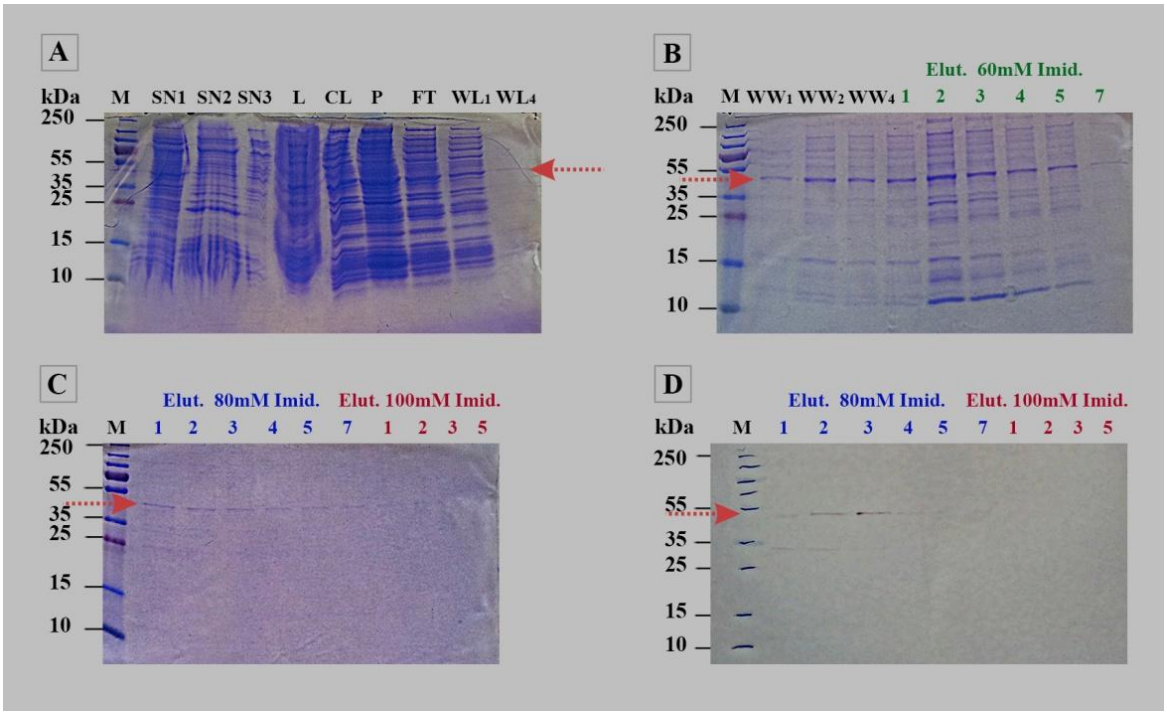


Figure 11: Purification of the His-tagged F1 protein using Ni-NTA chromatography under native conditions. Coomassie blue-stained gels (**A, B, C**) and Western blot (**D**) stained with anti-His-tag antibodies, showing fractions of different purification steps : **SN1:** supernatant after cell lysis in 25 mM imidazole, 0.1% Triton-X-100, pH 7.4, treatment with DNaseI, PMSF and 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN2:** soluble fraction after washing with 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN3**: soluble fraction after washing with 25 mM imidazole, 0.1% Triton-X-100, pH 7.4; **L:** lysate containing solubilized inclusion body pellet in 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0; **CL:** cleared lysate obtained after centrifugation step; **P:** pellet obtained after final centrifugation step; **FT:** flow through Ni-NTA column; **WL_1 and WL_4:** wash fractions with 50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0 **WW_1-WW_4:** wash fractions with 50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0; **E 1-7 *in green*:** eluates with 50 mM NaH_2PO_4, 300 mM NaCl, 60 mM imidazole, pH 8.0; **E 1-7 *in blue*:** eluates with 50 mM NaH_2PO_4, 300 mM NaCl, 80 mM imidazole, pH 8.0. **E 1-5 *in red*:** eluates with 50 mM NaH_2PO_4, 300 mM NaCl, 100 mM imidazole, pH 8.0. Molecular weights (kDa) are indicated on the *left margins* and *red arrows* show the position of the F1 protein at approximately 48.8 kDa.

IV.2. Expression, solubility, purification and characterization of recombinant F2 protein

IV.2.1. Expression and solubility screening of F2 protein

In order to optimize the expression yield of recombinant F2 protein, expression levels before and at different time points after IPTG induction were studied as shown in Figure 12A. A protein band migrating at approximately 14 kDa, which corresponded to the molecular weight of the F2 protein subunit, appeared already after 1h after IPTG induction. The intensity of this band increased remarkably until 4h and remained largely unaltered even after overnight incubation (Figure 12A). Highest levels of F2 expression were observed 3 to 4 hours after addition of IPTG and therefore, cells were harvested at these time points. No protein of interest could be detected before the addition of IPTG confirming that the expression was IPTG-dependent (Figure 12A). Again, the identity of F2 protein was verified using a monoclonal anti-His-tag antibody which reacted specifically with the protein band of 14 kDa (data not shown). Figure 12B shows a Western-Blot result of F2 protein solubility analysis under native conditions. After the lysis of the cells, most of the recombinant protein was found in the insoluble fraction of the cell lysates (Figure 12B). Most likely, the high expression level of F2 led to its accumulation in *E. coli* inclusion bodies. Therefore, a procedure under denaturing conditions was used to separate and solubilise these insoluble aggregates as described in the methods section.

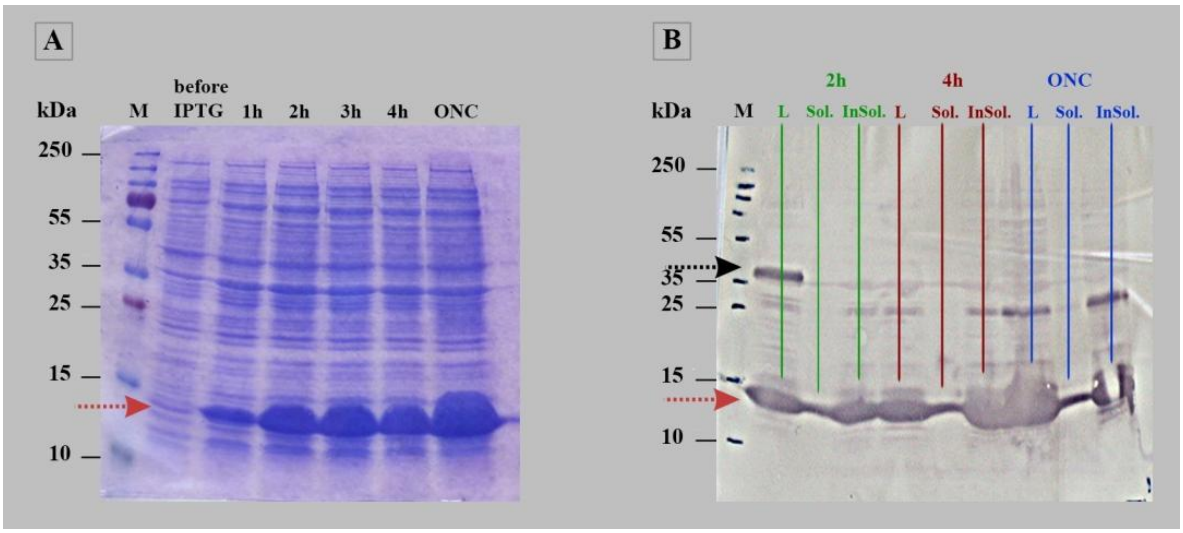


Figure 12: Time-course analysis after F2 protein expression and solubility. **A)** Coomassie blue-stained gel containing aliquots taken **before** and **1**, **2**, **3**, **4, 18** (overnight culture= **ONC**) hours following 1 mM IPTG induction. Molecular weights (kDa) are indicated on the *left margin*, the *red arrow* shows F2 protein position at 14 kDa. **B)** Western blot, stained with anti-His-tag antibodies, containing fractions from the analysis of the F2 protein solubility. Coloured lines show the position of the loaded samples (**L**: lysate, **Sol.**: soluble fraction, **InSol.**: insoluble fraction) at different time points (**2h**: green, **4h**: red, **ONC:** blue) after 1 mM IPTG induction. Molecular weights (kDa) are indicated on the *left margin*. The first slot contains additionally a VP1 protein at 37 kDa used as a positive control, indicated by a *black arrow*.

IV.2.2. Purification of recombinant F2 protein subunit under denaturing conditions

Due to the high protein expression level, F2 protein subunit was purified from 1 L *E. coli* culture using an inclusion body preparation protocol followed by Ni-NTA affinity chromatography under denaturing conditions. Remarkably, inclusion bodies were only solubilized when a strong denaturing solution, i.e. 6 M GuHCl, was used (CL) (Figure 13, upper left). Although the presence of the F2 protein was observed in the flow through (FT) and washing fractions (W1, W5), protein bands of very intensity were eluted using elution buffers of decreasing pH values (pH = 5.0; 4.5; 3.5). High amounts of protein migrating at 14 kDa in Coomassie blue-stained polyacrylamide gel were obtained in the first 5 fractions eluted with 8 M urea and pH of 4.5 followed by 10 fractions eluted with 8 M urea and pH of 3.5. However, these fractions also contained an additional band at approximately 25 kDa, which could either be a dimeric form of F2 protein or another co-purified *E. coli* protein (Figure 13, lower left). Interestingly, further elution steps led to the complete elimination of the 25 kDa protein band resulting in much higher purity level of the F2 protein subunit (Figure 13, lower right) which could also be proven by Western blotting (data not shown).

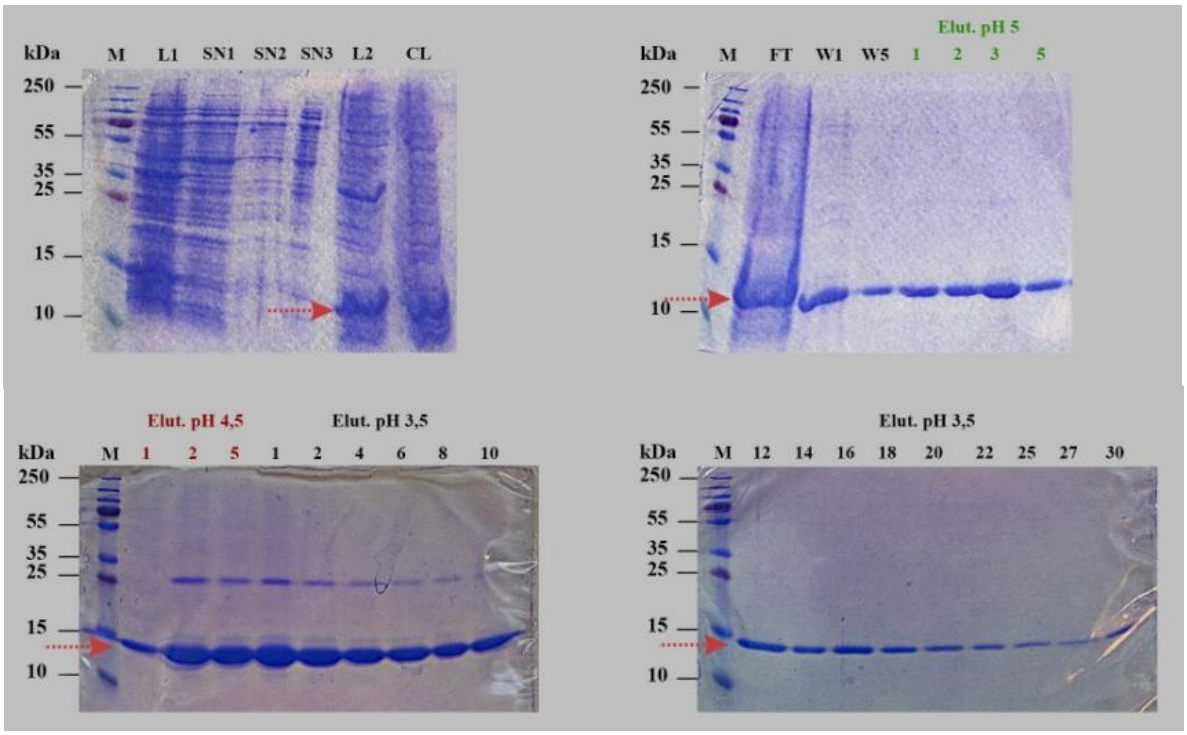


Figure 13: Purification of the His-tagged F2 protein using Ni-NTA technology under denaturing conditions. Coomassie blue-stained gels containing fractions obtained during inclusion body preparation protocol: **L1**: cell lysate obtained after cell lysis in 25 mM imidazole, 0.1% Triton X-100 pH 7.4, sonication by Ultra-Turrax and treatment with 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl pH 7.0 and DNaseI; **L2**: lysate of the solubilized inclusion body pellet in 6 M GuHCl, 100 mM NaH_2PO_4, 10 mM Tris, pH 8.0; **SN1:** supernatant after cell lysis in 25 mM imidazole, 0.1% Triton-X-100, pH 7.4, treatment with DnaseI, PMSF and 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN2:** soluble fraction after washing with 6% Triton X-100, 60 mM EDTA, 1.5 M NaCl, pH 7.0; **SN3**: soluble fraction after washing with 25 mM imidazole, 0.1% Triton-X-100, pH 7.4; **L2:** solubilized inclusion body pellet in 6 M GuHCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0. **CL:** clear lysate containing solubilized inclusion bodies obtained after centrifugation; **FT:** flow through the Ni-NTA column; **W1, W5:** wash fractions with 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 6.1; **E 1-5 *in green*:** eluates at pH 5.0; **E 1, 2, 5 *in red***: eluates at pH 4.5 and **E 1-30 *in black*:** eluates at pH 3.5. The first position of each gel was loaded with the molecular weight marker as indicated on *the left side* and the *red arrows* show the position of the F2 protein at approximately 14 kDa.

IV.2.3. Reducing and non-reducing SDS-PAGE of the F2 protein subunit

In order to verify the tendency of the F2 protein to form multimers, SDS-PAGE under reducing and non-reducing conditions were performed and the results are shown in Figure 14. For this purpose, purified F2 subunit (4,3 µg/lane) was mixed with the sample buffer with or without reducing agent, ß-Mercaptoethanol. Under reducing conditions, the F2 protein forms a single band at 14 kDa and a dimer at 28 kDa (Figure 14A and 14B, red colour). On the other hand, under non-reducing conditions, which leave disulphide bonds intact, the F2 protein forms several multimers mainly due to the presence of two cysteine residues within the 122 aa long subunit (Figure 14A and 14B, green colour).

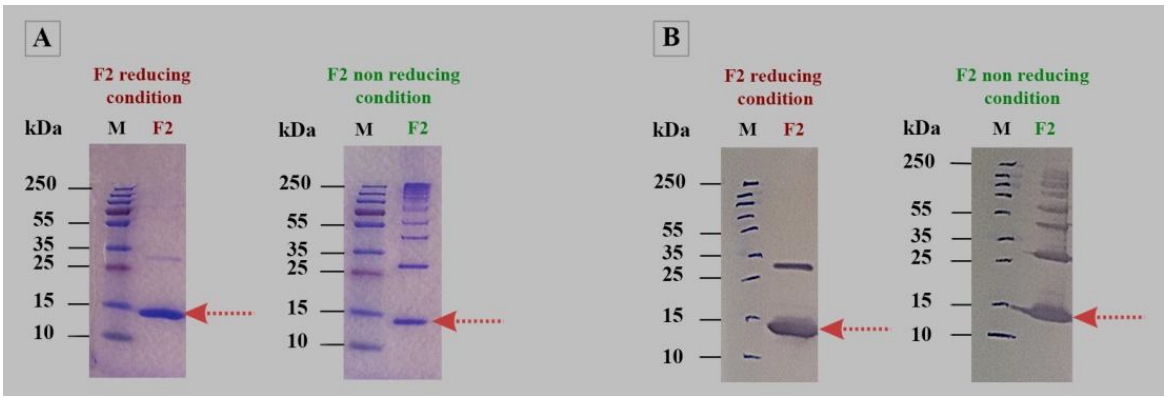
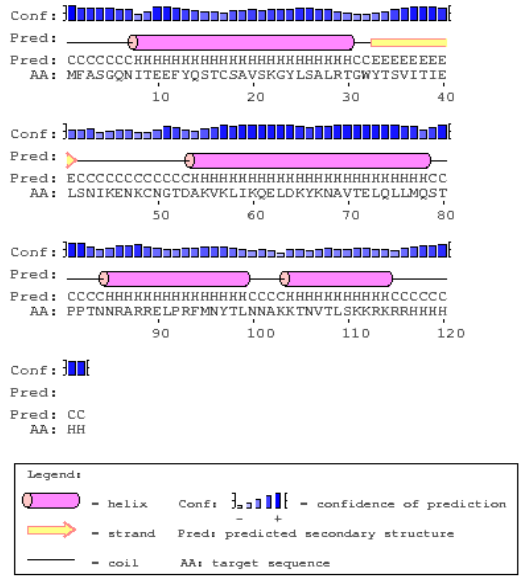


Figure 14: Purified and refolded F2 protein under reducing and non-reducing conditions. **A)** Coomassie blue-stained gel and **(B)** Western blot, stained with anti-His-tag antibodies, showing the recombinant F2 protein under reducing (*left side, in red*) and non-reducing (*right side, in green*) conditions. Molecular weights (kDa) are indicated on the *left margins* and *red arrows* show the position of the protein at 14 kDa.

IV.2.4. Circular dichroism analysis of the F2 protein subunit

To assess whether recombinant F2 protein subunit was folded after the removal of denaturing components, circular dichroism (CD) analyses was carried out. When analysed by far-UV CD, F2 protein showed one minimum at 208 – 210 nm and the second at 218 - 220 nm indicating a mixture between alpha-helix and beta-sheet structures (Figure 15B). In order to estimate the number of secondary structure segments in protein, the software programs CDSSTR, CONTIN and SELCON3 were used for the calculation (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml). Analysis with the reference data set 4 yielded very similar results between the different programs, in average 15% alpha-helix, 34% beta-sheet, 20% turns and 31% unordered structure (Figure 15A). These results were in accordance with Psi-Pred sequence-based predictions (http://bioinf.cs.ucl.ac.uk/psipred/) (Figure 15A). In addition, circular dichroism was further used to evaluate thermal stability of the F2 protein. Temperature scans were performed according to a step-scan procedure, in which the sample was heated stepwise from 25°C to 95°C with a rate of 2°C/min. Surprisingly, the protein of interest remained largely intact and did not show a thermal unfolding transition even at the temperature of 95°C, which usually have a denaturing effect on the proteins (data not shown).

A



B

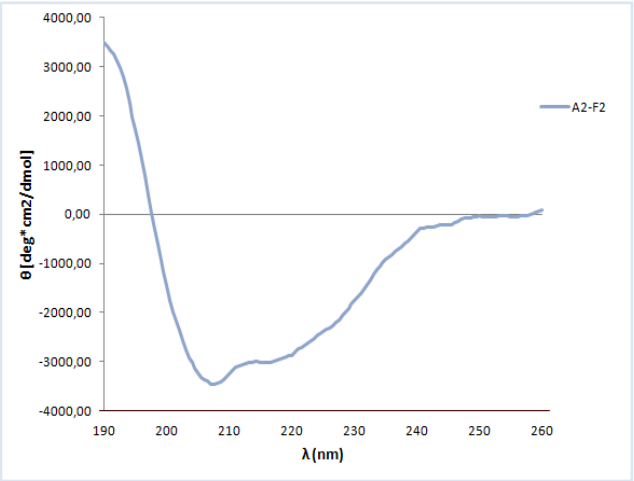


Figure 15: Analysis of the secondary structure of the recombinant F2 protein subunit. **A)** PsiPred sequence-based prediction of the F2 protein structure. Shown are potential structural elements (Pred) together with the confidence level of the prediction (Conf) indicated in blue bars and the amino acid sequence of the F2 protein (AA). Alpha-helixes are represented by a pink cylinder and the letter **H**; beta-sheets as yellow arrows and the letter **E**; coiled coil or unordered structures are marked as black dashes and the letter **C**. **B)** Far-UV CD analysis of purified F2 protein subunit. The spectrum was recorded at 25°C and is expressed as mean residue ellipticities of three measurements (Θ, y-axis) at given wavelengths (nm, x-axis).

IV.2.5. Mass spectrometry of the purified F2 protein subunit

MALDI-TOF spectrometry was used to determine the molecular weight of the purified F2-protein sample. The MALDI-TOF spectrogram shows a main peak of 14147,958 Da which is in agreement with the predicted mass of 14150 Da by ProtParam software (Figure 16). Beside the prominent protein peak, a small and flat rise in the range of m/z 15000 can be observed. This is probably due to metal ion adduct formation as its appearance is common in biological samples because of high endogenous concentration of various salts but also because of other salts which might have been added during the sample preparation process. Especially the widely used matrix (sinapinic acid, α-Cyano-4-hydroxycinnamic acid or 2,5-dihydroxybenzoic acid) which facilitates the ionization of proteins and peptides in MALDI-TOF-MS leads to matrix-analyte-adduct formation induced by sodium and potassium ions. [68] These elements are present in solvents and buffers, and are extracted from many plastics that are used for sample preparation and storage. Matrix adducts are ubiquitous in MALDI-TOF mass spectra, in the range of m/z 800–1100 and are particularly evident at low sample concentrations.[69] Single

adduct peaks in the range of m/z 800-1100 were not observed due to the fact that measurement started at this point (calibration with protein standard in the range of m/z 1000 to 35000). There are also other peaks with very low intensity signal and lower mass to charge values (7068,330 Da, 8305,413 Da) which may represent other impurities e.g. components after insufficient isolation from biological material or fragments of protein of interest which molecular weights could be compared with the predicted masses by fragmentation software ACD/MS-Fragmenter. Another peak of very low signal intensity, with a mass of 28412,514 Da, may represent F2 in its dimeric form, since Metthews *et al.* showed that RSV F2 subunit region HR3 (Heptad repeat 3, 53-100 aa) is able to form dimeric helices (Figure 16).[70]

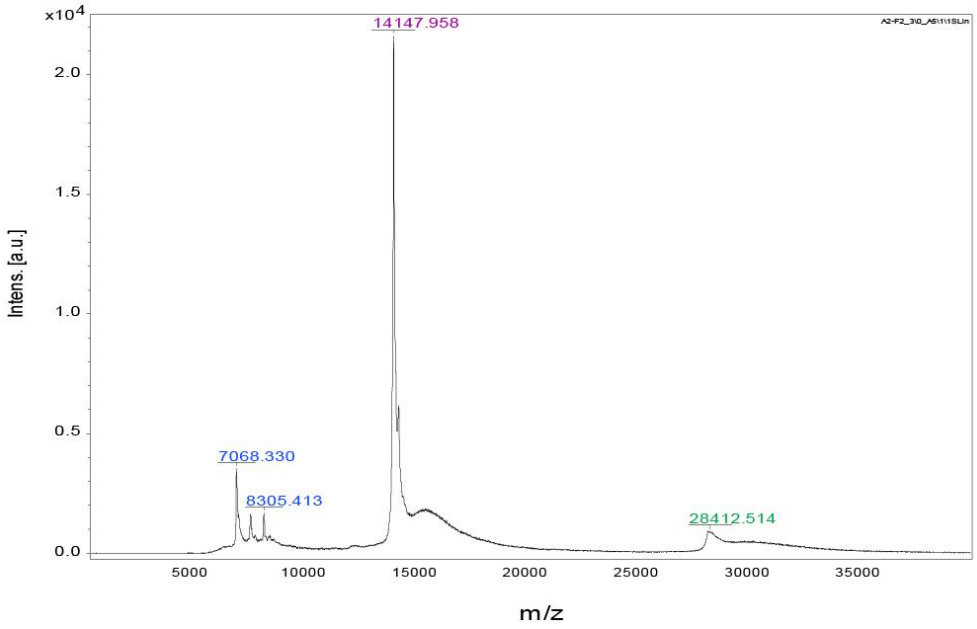


Figure 16: MALDI-TOF mass spectrometry analysis of the purified F2 protein subunit. The x-axis represents the mass/charge (m/z) ratio and the y-axis displays the signal's intensity in arbitrary units (a.u.). 14147.958 Da *in purple* indicates the measured molecular weight of the sample with the highest intensity that corresponds to the predicted molecular weight by ProtParam (14150 Da). The color *blue* indicates the molecular weight of by-products and *green* a possible F2 dimer.

IV.3. Production of recombinant G protein and peptide synthesis

IV.3.1. Expression and solubility of recombinant G protein

Figure 17A and 17B show a time-course analysis of the expression levels of recombinant G protein. For this particular protein, which contains a lot of hydrophobic regions, the incubation temperature was reduced from 37°C to either 30°C or 25°C after IPTG induction. Furthermore, no overnight starter culture was prepared and the number of generations before induction was kept to minimum of OD_{600nm} = 0.4 in order to reduce the possible toxic effect of this protein.[71] For the analysis of protein solubility, a Western blot was performed. A monoclonal anti-His-tag antibody detected a band of approximately 40 kDa almost exclusively in the insoluble fraction of cell lysates (Figure 17C). Although the expression of recombinant G protein could be induced in *E. coli* cells, it was not enough to proceed with the purification process.

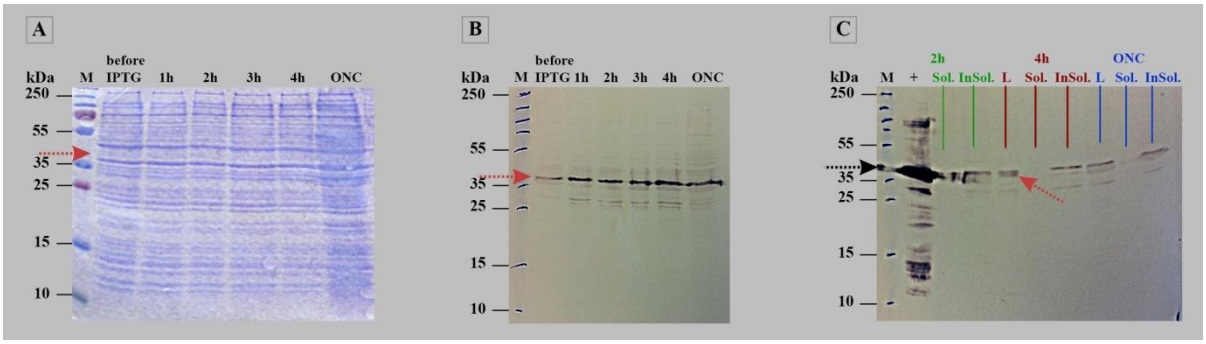


Figure 17: Time-course analysis of the G protein expression at 30°C. **A)** Coomassie blue-stained polyacrylamide gel and **(B)** Western blot, stained with anti-His-tag antibodies, containing 1 ml culture samples taken **before** and **1, 2, 3, 4, 18** (overnight culture= **ONC**) hours following 1 mM IPTG addition. **C)** Western blot, stained with anti-His-tag antibodies, containing fractions from the analysis of the G protein solubility. Coloured lines show the position of the loaded samples (**L**: lysate, **Sol.**: soluble fraction, **InSol.**: insoluble fraction) at different time points (**2h**: green, **4h**: red, **ONC:** blue) after 1mM IPTG induction. The first slot contains additionally a VP1 protein at 37 kDa used as a positive control and indicated by a *black arrow*. Molecular weights (kDa) are indicated on the *left margins,* the *red arrows* show the position of the G protein at approximately 40 kDa.

IV.3.2. Synthesis of peptides spanning the G protein

In order to map the surface glycoprotein G, synthetic 5 aa overlapping peptides with a length of 25 to 27 aa spanning the ectodomain of the subtype A2 were synthesized by solid phase synthesis with the 9-fluorenyl-methoxy-carbonyl-method. Peptides were designed to avoid disruption of special sequence motifs within the aa sequence. In order to avoid a cut in the conserved domain or the CX3C motif within the A2-G-10 peptide, this product has a length of 27 aa. Furthermore, to avoid the formation of disulfide bonds, two Cys residues on one peptide were separated which allowed better synthesis outcome and a product with higher purity. Similarly, proline residues on C- terminal end were also avoided as this aa can cause significant deletions during peptide synthesis process. [72]

Altogether, 16 peptides were synthesized and purified by HPLC and their biochemical properties are demonstrated in Table II. These peptides will subsequently be used to establish a serological diagnostic assay for measuring antibody responses to HRSV following experimental human infection.

No.	Peptide	Sequence	species	lenght (AA)	MW	pI	total charge	Properties
1	A2-G-1	MSKNKDQRTAKTLERTWDTLNHLLF	RSV-A-A2	25	3047,4	9,69	-2	Intravirion part
2	A2-G-2	KDQRTAKTLERTWDTLNHLLFISSC	RSV-A-A2	25	2977,3	8,21	-1	Intravirion part
3	A2-G-3	FISSCLYKLNLKSVAQITLSILAMI	RSV-A-A2	25	2770,4	9,20	-2	intravirion part/transmembran part
4	A2-G-4	ILAMIISTSLIIAAIIFIASANHKV	RSV-A-A2	25	2624,2	8,76	-1	transmembran part
5	A2-G-5	ANHKVTPTTAIIQDATSQIKNTTPT	RSV-A-A2	25	2651,9	8,64	-1	virion surface
6	A2-G-6	NTTPTYLTQNPQLGISPSNPSEITS	RSV-A-A2	25	2660,8	4,00	1	virion surface (2 potential O-glyc.)
7	A2-G-7	SEITSQITTILASTTPGVKSTLQST	RSV-A-A2	25	2564,8	5,72	0	virion surface (4 potential O-glyc.)
8	A2-G-8	STLQSTTVKTKNTTTTQTQPSKPTT	RSV-A-A2	25	2680,9	10,30	-3	virion surface (5-6 potential O-glyc., 1 potential N-glyc.)
9	A2-G-9	SKPTTKQRQNKPPSKPNNDFHFEVF	RSV-A-A2	25	2972,3	10,00	-3	virion surface (2 potential O-glyc.)/ conserved domain
10	A2-G-10	DFHFEVFNFVPCSICSNNPTCWAICKR	RSV-A-A2	27	3178,6	6,72	0	conserved domain/CXC3 Motif
11	A2-G-11	TCWAICKRIPNKKPGKKTTKPTKK	RSV-A-A2	25	2857,5	10,49	-9	Heparin Binding Domain/ virion surface (2 potential O-glyc.)
12	A2-G-12	KPTKKPTLKTTKKDPKPQTTKSKEV	RSV-A-A2	25	2838,3	10,30	-5	virion surface (2 potential O-glyc.)
13	A2-G-13	KSKEVPTTKPTEEPTINTTKTNIIT	RSV-A-A2	25	2772,1	8,43	-1	virion surface (2 potential O-glyc., 1 potential N-glyc.)
14	A2-G-14	TNIITTLLTSNTTGNPELTSQMETF	RSV-A-A2	25	2728,0	3,79	2	virion surface (3 potential O-glyc., 1 potential N-glyc.)
15	A2-G-15	QMETFHSTSSEGNPSPSQVSTTSEY	RSV-A-A2	25	2718,8	4,24	3	virion surface (2 potential O-glyc.)
16	A2-G-16	PSPSQVSTTSEYPSQPSSPPNTPRQ	RSV-A-A2	25	2656,8	6,43	0	virion surface (5 potential O-glyc.)

Table II: Amino acid sequences of peptides spanning the complete G protein of the A2 strain (subgroup A). Length (AA), molecular weight (MW), pI value and total charge of each peptide are shown. Total charge is calculated from ProtParam (Expasy) prediction as follows: number of negatively charged residues minus number of positive charged amino acids. Peptide property information was derived from UniProtKB (Expasy). Special motifs of peptides (conserved domain, CXC3 motif and heparin binding domain) are colored in *blue*, *green* and *red*.

IV.3.3. Mass spectrometry of purified peptides spanning the G protein

MALDI-TOF spectrometry was used to verify the molecular weights of the purified peptides. Figure 18 shows spectrograms of three representative purified G protein peptides. Main peaks of 2970,943 Da for A2-G-9, 3180,686 Da for A2-G-10 and 2856,556 Da for A2-G-11 corresponded to the molecular masses calculated by ProtParam software. The peptide A2-G-10 showed a lower intensity signal than the other two peptides most likely due to a longer peptide length. In fact the concentration of full-length peptide synthesized in a reaction is inversely correlated with the length of the proposed peptide; therefore, as the length of the peptide increases, the yield is reduced due to increased difficulties in purifying the low-abundant product from the crude mixture.[73] The purity of the synthesized peptides was 91.25%, 92.86% and 80.56% respectively.

The remaining 13 peptides were purified to a similar or even higher purity and the obtained intensity values were enhanced. The only exceptions were peptides A2-G-3, A2-G -4, A2-G-6 and A2-G-14 which either did not lead to the formation of a precipitate or did not match the masses corresponding to the predicted molecular weights. In the case of A2-G-3 or A2-G-4 peptides, it was even not possible to obtain enough peptide precipitate after synthesis. This could be explained by a high amount of hydrophobic residues of isoleucine and leucine which can cause the formation of β-sheets. As a consequence, it can prevent the growing peptide from dissolving completely during synthesis, causing deletion in the synthesized peptide sequence.[74] Both peptides contained transmembrane parts of the G protein and did not lead to a satisfying product outcome even after usage of ready to use double-coupled amino acids (I-S, S-I, I-T, K-S and T-S). Two other synthetic peptides A2-G-6 and A2-G-14 yielded enough peptide precipitate but, unfortunately, resulted in incorrect molecular weights. In the case of A2-G-6 very high amounts of serine and proline residues could not be avoided which lead probably to deletions during synthesis. It is known that especially proline residues can undergo cis/trans isomerization and thus reduce the purity of a peptide. Furthermore, this peptide has an asparagine at its N-terminal end which also should have been avoided, because the asparagine N-terminal protecting group can be difficult to remove during the cleavage. [75] The A2-G-14 peptide consisted of 32% threonine and 20% leucine/isoleucine

residues which are known to cause β-sheets, which in turn lead to incomplete solvation during peptide synthesis, resulting again in deletions.

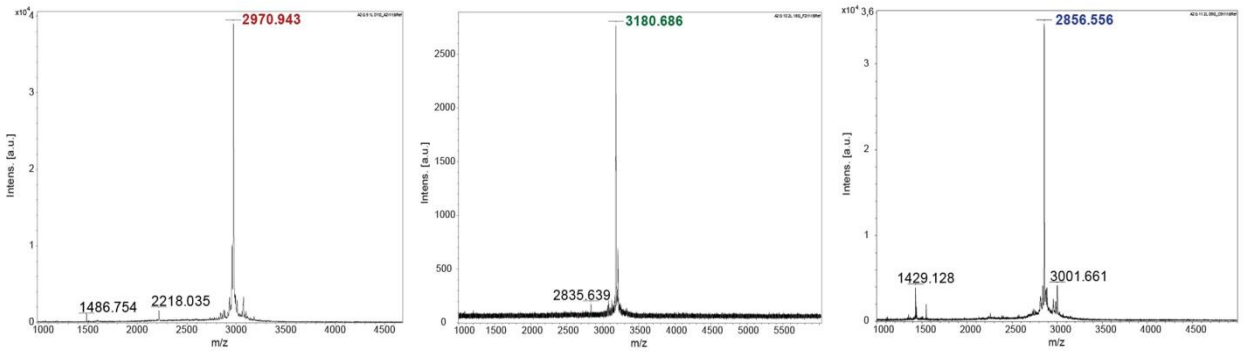


Figure 18: MALDI-TOF mass spectrometry analysis of representative purified G peptides A2-G-9, A2-G-10 and A2-G-11 (from the left). The x-axis represents the mass/charge (m/z) ratio and the y-axis displays the signal's intensity in arbitrary units (a.u.). 2970,943 Da *in red,* 3180,686 *in green* and 2856,556 *in blue* indicate the measured molecular weights of the peptide samples with the highest intensities that correspond to the predicted molecular weights by ProtParam (2972,3 Da, 3178,6 Da and 2857,5 Da). The color *blue* indicates the molecular weight of possible by-products in much lower signal intensity.

V Discussion

The major goal of the current thesis was to construct, express and purify two important HRSV-derived proteins, F and G, using gene- and peptide-synthesis technologies. In a first step, sequences of the viral proteins from all known HRSV strains, including the two antigenic groups, HRSV-A and HRSV-B, were retrieved from sequence data bases and aligned to determine the most distantly related strains (Figure 2, 3 and 4). Next, synthetic genes with codons optimized for the expression in *Escherichia coli* were made for F1, F2 and G proteins from two representative strains, A2 and BA, respectively.

In the course of this work, particular attention was given to optimize the expression and purification strategies using Nickel affinity chromatography to obtain soluble and folded recombinant proteins. To maximize the expression yield, protein expression before and at different time points after induction with IPTG was studied for all three proteins. When compared with F2 recombinant protein expression, F1 protein subunit was expressed at very low concentration most likely due to the tendency of this protein to fuse to *E. coli* cells.[76] This toxic effect has already been observed during bacterial growth as IPTG induction resulted in severe inhibition of the cell growth.[77, 78] Therefore, it is very likely that the expression efficiency decreased due to the cellular death, which at the end resulted in a low amount of the purified protein, making it difficult to perform experiments necessary for its biophysical characterization. Unfortunately, changes of the expression conditions did not improve the expression efficiency. For example, in order to reduce possible protein toxicity, the incubation temperature was reduced from 37°C to 30°C.[79] Also lowering the IPTG concentration from 1 mM to 0,10 µM or 0,4 µM as it was done by Arcuri *et al.* did not lead to radical improvement.

Nevertheless, efforts were made to purify the recombinant F1 protein subunit. First, a purification protocol using the Ni-NTA technology under denaturing conditions was performed, once using the clarified lysate of cell homogenate, in which cell pellets were directly re-suspended in 6 M GuHCl, and also using the cleared lysate obtained after inclusion body preparation. Unfortunately, both experimental setups did not yield pure and sufficient amounts of the F1 protein, whereby a better result was obtained when an inclusion body preparation prior to purification was performed. Therefore, protein purification was carried out under native conditions after inclusion body preparation procedure. This led to an

improvement of the binding of the recombinant protein to the Ni-NTA resin, as ticker protein bands were observed in SDS-PAGE after purification process (Figure 11B and 11C). However, an increase of un-specific products in purified protein fractions was also noticed. Besides of the F1 protein migrating at molecular weight of 48.8 kDa, a 13 kDa protein band was also found (Figure 11B). This could be explained by the fragmentation of the huge F1 protein subunit or by co-purification of a native *E. coli* protein. The co-contaminant expression of native bacterial proteins that exhibit a relatively high affinity for divalent cations during the expression in *E. coli,* frequently results in their co-purification during Ni-NTA affinity chromatography. This is a very common problem in Ni-NTA technology, especially when the expression level of the recombinant protein is low.[80] Our results are in accordance with this statement as very low expression levels of F1 protein were detected. Furthermore, it is also very likely that the detected contaminant is the *Hfq* protein (*Host factor-I protein*) (Table I).

Data of Vytvytska *et al.* indicate that *Hfq* participates in the regulation of *ompA* mRNA stability in response to changes in growth rate due to environmental stress conditions. *OmpA* mRNA in *E. coli* is coding for the major outer membrane protein A and the stability of the transcript correlates inversely with the bacterial growth rate as it was observed that in slow-growing cells its half-life is reduced. *Hfq* acts most likely as an RNA chaperone altering the *ompA* transcript stability and translation efficiency leading in further consequence to transcript degradation. In addition, it was reported that the amount of *Hfq* was observed to be larger in slow growing cultures than in cells dividing more rapidly.[81]

To overcome the problem of co-purification of native *E. coli* proteins along with the recombinant protein, the imidazole concentration in native lysis buffer was decreased from 10 mM to 5 mM. Moreover, the amount of Ni-NTA resin as well as the incubation time was reduced in order to lower the un-specific binding of contaminating proteins. Further attempts to reduce the amount of contaminants included rinsing the loaded protein with native lysis buffer, followed by wash buffer and finally elution with 150 mM imidazole concentration in buffer. As a result, the most pure fractions of F1 could be obtained during washing instead of elution steps, in which a very high amount of contaminating proteins could also be detected. In another experiment, the concentration of imidazole in the elution buffer was reduced to ensure a gentle and stepwise purification procedure. Although, this way of handling led to an acquisition of more pure fractions of F1, the obtained yield was still very low (Figure 11). The reduction of imidazole concentration in the wash buffer did also not influence the binding

efficiency. By repeating the experiment several times, it was possible to obtain enough pure fractions of F1. However, a new problem arose during the concentration procedure as the protein tended to precipitate out of solution.

Precipitation of proteins is sometimes coupled with the inhomogeneity of the sample, either in the form of contaminating protein or alternate folding states. Precipitation can also occur by aggregation due to the presence of hydrophobic or hydrophilic patches on the surface of the target protein, regions that are known to exist on F1 (Figure 3 and 4). Moreover, this problem worsens as the protein concentration increases.[82] Therefore, efforts focused on finding a suitable buffer. In order to determine conditions that enhance protein solubility, 7 different buffers were screened simultaneously for their ability to prevent protein aggregation (referred to as protein storage buffer A-G). The best result could be achieved in buffer A (10 mM NaH_2PO_4, pH 5.5) and buffer G (1x PBS: 10 mM NaH_2PO_4, 10 mM Tris, 140 mM NaCl; 2,7 mM KCl, 1,8 mM KH_2PO_4 pH 7.3). Nevertheless, F1 protein subunit precipitated at the concentration of 0,025 mg/ml making further screenings for a suitable buffer necessary.

Also the G protein expression did not yield a sufficient protein amount. This is most likely due to its hydrophobic regions which are known to have a toxic effect on host cells as they tend to associate with or incorporate into vital membrane systems.[83] Moreover, a level of basal transcription which occurred in the absence of IPTG induction ("leakiness") should be avoided because of the possible protein toxicity to cells. Optimization of expression conditions did also not lead to better expression efficiency. Interestingly, the recombinant G protein migrated in SDS-PAGE with a molecular weight of about 40 kDa, which is higher than that deduced from the amino acid sequence by ProtParam-ExPASy-tool (33,408 kDa) (Figure 17). However, un-glycosylated G protein synthesized *in vitro* using cell-free translation of mRNA as well as the recombinant G protein expressed in *Salmonella typhimurium* also showed a retarded electrophoretic mobility with respect to its calculated molecular weight.[84, 85]

In summary, both F1 as well as G proteins were expressed in *E. coli* in very low amounts which made the protein purification and concentration difficult. There are still some more possibilities which can be tried to obtain better results. For example, in order to increase the yield of the recombinant protein, it could be helpful to cut the sequence within the region of transmembrane anchor in the F1 protein subunit and the signal peptide/membrane anchor part in the G protein as this part seemed to be responsible for the toxic effect in expressing cells. This procedure was used for the production of vesicular stomatitis virus G protein where

the deletion of NH_2-terminal domain (signal peptide) allows evading the lethal effect of this region to *E. coli* cells.[86] The removal of predicted membrane-spanning regions and the prevention of hydrophobic residues at the termini are suggestions which are widely accepted and could lead to a better protein expression outcome.[87]

Furthermore, the protein expression level could be enhanced by using a different type of media such as Terrific Broth (TB) which may support higher levels of membrane protein expression.[88] In order to inhibit basal expression level, as it was observed in the case of G protein (Figure 17), the addition of 1% glucose to the TB media should be considered.[89] Alternatively, *E. coli* strains in which the T7 RNA polymerase is under the control of the more tightly regulated arabinose promoter can also be tried.[90] The use of another prokaryotic expression system e.g. *Salmonella typhimurium* cells can be considered as this type of bacteria was used for the successful recombinant HRSV G glycoprotein production as described by Martin-Gallardo *et al*.[91]

In contrary to F and G proteins, there were no problems with the expression and purification of protein subunit F2. Noteworthy, a protein band of 25 kDa was observed in the purified fractions (Figure 13). Possible reasons for this result are either the co-purification of a native *E. coli* protein or the formation of aggregates in the purified fractions.

By comparing the molecular weights of all possible contaminants (Table I), the YadF protein seems to be an appropriate candidate. The *yadF* gene of *E. coli* encodes a β-class carbonic anhydrase (CA), an enzyme which interconverts CO_2 and bicarbonate. Its expression is influenced by cell growth rate and is maximal in slow-growing cultures with high density. Furthermore, the expression of the YadF protein is enhanced during starvation or heat stress conditions.[92] These assumptions are partially in accordance with the experimental condition during F2 expression. The expressing cell culture showed high cellular density and starvation was also likely due to induction period of 4 hours with an end OD_{600nm} of 3,67. The detection of both protein bands, 14 kDa and 25 kDa, with monoclonal anti-his-tag antibodies can also be explained by the fact that the YadF protein contains 5.5% of histidine residues. Another possibility is that the F2 protein tended to form dimers as shown by Metthews *et al.*[93]

In summary, purified recombinant F2 protein subunit, and later on also F1 and G proteins, might be used for the characterization of immune responses in RSV infected patients as not much is known about the humoral immune responses directed against recombinant RSV

proteins. Eventually, recombinant proteins may be used to develop new therapeutic strategies for RSV infections.

The production of synthetic peptides spanning the whole G protein of the A2 subtype of HRSV group- A was very successful even though certain peptides could not be produced (A2-G-3, A2-G-4, A2-G-6 and A2-G-14). But by shortening peptide sequences of all those peptides and by avoiding the N-terminal argenine (A2-G-6) it is very likely that also these peptides can be produced.

VI Conclusion and future perspectives

This thesis describes the production conditions of recombinant HRSV proteins in *E. coli*. In addition, it shows the successful production of synthetic peptides spanning the ectodomain of HRSV attachment G protein.

The findings obtained during the course of this work may be useful for the development of diagnostic tests for HRSV infection. Furthermore, new insights might be gained for the fusion protein subunit F2. So far, little is known about the smaller subunit F2 due to the fact that most of the studies deal with the F1 subunit. Because the evidence of a secondary conformation of the F2 subunit could be provided by CD-spectroscopy, there is a chance that this recombinant protein displays some of its original activities such as possible interactions with host molecules. The fact that this protein product is folded, allows further tertiary structure analysis such as NMR and crystallization, which would provide help to gain further information about its regulation and biological tasks as well as the identification of its interacting partner in host cells. These new findings could help to clear up the mystery of binding partners of HRSV entry receptors as Schlender *et al.* show that host specificity function depends on F2 subunit alone. All in all the obtained results could present a step towards the development of a new treatment approaches and diagnostic tools for HRSV infection.

References

[1] Robert A. Dudas and Ruth A. Karron. Clinical Microbiology Reviews; 1998/ 11 (3). *Respiratory Syncytial Virus Vaccines*. Page 430-439.

[2] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1619.

[3] W. Paul Glezen et al. Formerly Archives of Pediatrics & Adolescent Medicine/ JAMA Pediatrics; 1986/ 140 (6). *Risk of Primary Infection and Reinfection with Respiratory Syncytial Virus*. Page 543-546.

[4] Rafael Lozano et al. Lancet 2012/ 380. *Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010*. Page 2095-2128.

[5] N. Sigurs et al. American Journal of Respiratory and Critical Care Medicine; 2005/ 171 (2). *Severe respiratory syncytial virus bronchiolitis in infancy and asthma and allergy at age 13.* Page 137–141.

[6] J. C. Rooney and H. E. Williams. The Journal of Pediatrics; 1971/79 (5). *The relationship between proved viral bronchiolitis and subsequent wheezing*. Page 744–747.

[7] P. D. Sly and M. E. Hibbert. Pediatric Pulmonology; 1989/ 7 (3). *Childhood asthma following hospitalization with acute viral bronchiolitis in infancy*. Page 153–158.

[8] C.A. Heilman. The Journal of Infectious Disease; 1990/ 161 (3). *From the National Institute of Allergy and Infectious Diseases and the World Health Organization. Respiratory syncytial and parainfluenza viruses.* Page 402-406.

[9] http://www.cdc.gov/rsv/about/transmission.html (25.5.2014)

[10] S. Yusuf et al. Epidemiology and Infection; 2007/ 135 (7). *The relationship of meteorological conditions to the epidemic activity of respiratory syncytial virus.* Page 1077-1090.

[11] Alfonsina Trento et al. Journal of Virology; 2010/ 84 (15). *Ten Years of Global Evolution of the Human Respiratory Syncytial Virus BA Genotype with a 60-Nucleotide Duplication in the G Protein Gene.* Page 7500-7512.

[12] P. R. Johnson et al. Journal of Virology; 1987/ 61 (10). *Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity.* Page 3163-3166.

[13] E. E. Walsh et al. The Journal of Infectious Diseases; 1997/ 175 (4). *Severity of respiratory syncytial virus infection is related to virus strain.* Page 814-820.

[14] W. M. Sullender. Clinical Microbiology Reviews; 2000/ 13 (1). *Respiratory syncytial virus genetic and antigenic diversity.* Page 1-15.

[15] AliReza Eshaghi et al. Public Library of Science One; 2012/ 7 (3). *Genetic Variability of Human Respiratory Syncytial Virus A Strains Circulating in Ontario: A Novel Genotype with a 72 Nucleotide G Gene Duplication.* E32807.

[16] Guanglin Cui et al. Emerging Microbes & Infection; 2013/ 2 (4). *Emerging human respiratory syncytial virus genotype ON1 found in infants with pneumonia in Beijing, China.* e22.

[17] Zhiwu Sun et al. Viruses 2013/ 5 (1). *Respiratory Syncytial Virus Entry Inhibitors Targeting the F Protein*. Page 211-225.

[18] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1632.

[19] Peter L. Collins et al. Elsevier; Virology 2002/ 296 (2). *Respiratory Syncytial Virus: Reverse Genetics and Vaccine Strategies.* Page 204-211.

[20] J. Schlender et al. Journal of Virology; 2000/ 74 (18). *Bovine respiratory syncytial virus nonstructural proteins NS1 and NS2 cooperatively antagonize alpha/beta interferon-induced antiviral response.* Page 8234–8242.

[21] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1602-1609.

[22] Peter L. Collins et al. Proceedings of the National Academy of Science of the United States of America; 1996/ 93 (1). *Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus.* Page 81-85.

[23] A. Bermingham et al. Proceedings of the National Academy of Science of the United States of America; 1999/ 96 (20). *The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription*. Page 11259-11264.

[24] D.S. Steck et al. Elsevier; Virology 1991/ 183 (1). *Sequence analysis of the polymerase L gene of human respiratory syncytial virus and predicted phylogeny of nonsegmented negative-strand viruses.* Page 273-287.

[25] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1607.

[26] A. Bukreyev ·et al. Journal of Virology; 2008/ 82(24). *The secreted form of respiratory syncytial virus G glycoprotein helps the virus evade antibody-mediated restriction of replication by acting as an antigen decoy and through effects on Fc receptor-bearing leukocytes.* Page 12191-12204.

[27] D. A. Hendricks et al. Journal of Virology; 1988/ 62(7). *Further characterization of the soluble form of the G glycoprotein of respiratory syncytial virus.* Page 2228-2233.

[28] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1607.

[29] Wathen M. W. et al. Biochemistry; 1991/ 30(11). *Characterization of oligosaccharide structures on a chimeric respiratory syncytial virus protein expressed in insect cell line Sf9*. Page 2863-2868.

[30] *Fields Virology* Volume Two; Editors-in-chief: David M. Knipe, Peter M. Howley; Associate Editors: Diane E. et al. Lippincott Williams & Wilkins 2007, Fifth Edition. Page 1607-1608.

[31] R. A. Tripp et al. Nature Immunology; 2001/ 2 (8). *CX3C chemokine mimicry by respiratory syncytial virus G glycoprotein.* Page 732-738.

[32] Fernando P. Polack et al. Proceedings of the National Academy of Science of the United States of America; 2005 /102(25). *The cysteine-rich region of respiratory syncytial virus attachment protein inhibits innate immunity elicited by the virus and endotoxin.* Page 8996-9001.

[33] J. P. Langedijket al. Elsevier; Virology 1998/ 243 (2). *Structural homology of the central conserved region of the attachment protein G of respiratory syncytial virus with the fourth subdomain of 55-kDa tumor necrosis factor receptor.* Page 293–302.

[34] K. M. Neuzil et al. The American Journal of Medical Scinces; 1996/ 311(5). *Protective Role of TNF-alpha in respiratory syncytial virus infection in vitro and in vivo.* Page 201-204.

[35] Michael N. Teng et al. Journal of Virology; 2002/ 76 (12). *The Central Conserved Cysteine Noose of the Attachment G Protein of Human Respiratory Syncytial Virus Is Not Required for Efficient Viral Infection In Vitro or In Vivo.* Page 6164-6171.

[36] S. Levine et al. Journal of General Virology; 1987/ 68 (9). *Demonstration that the glycoprotein G is the attachment protein of respiratory syncytial virus.* Page 2521-2524.

[37] T. Krusat and H.J. Streckert. Archives of Virology; 1997/ 142 (6). *Heparin-dependent attachment of respiratory syncytial virus (RSV) to host cells.* Page 1247–1254.

[38] J.T. Gallagher et al. Biochemical Journal; 1986/ 236 (2). *Structure and function of heparan sulfate proteoglycans.* Page 313–325.

[39] Steven A. Feldman et al. Journal of Virology; 1999/ 73(8). *Identification of a Linear Heparin Binding Domain for Human Respiratory Syncytial Virus Attachment Glycoprotein G.* Page 6610–6617.

[40] Alfosina Trento et al. Journal of Virology; 2010/ 84 (15). *Ten Years of Global Evolution of the Human Respiratory Syncytial Virus BA Genotype with a 60-Nucleotide Duplication in the G Protein Gene.* Page 7500-7512.

[41] Abdul s. Yunus et al. Elsevier; Virology 2010/ 396 (2). *Elevated temperature triggers human respiratory syncytial virus F protein six-helix bundle formation.* Page 226-237.

[42] L.J. Calder et al. Virology; 2000/ 25 (271). *Electron microscopy of the human respiratory syncytial virus fusion protein complex that it forms with monoclonal antibodies.* Page 122-131.

[43] Lydia Tan et al. Journal of Virology; 2013/ 87 (14). *The Comperative Genomics of Human Respiratory Syncytial Virus Subgroups A and B: Genetic Variability and Molecular Evolutionary Dynamics*. Page 8213-8226.

[44] C.M. Carr et al. Proceedings of the National Academy of Science of the United States of America; 1997/ 94 (26). *Influenza hemagglutinin is spring loaded by a metastable native conformation*. Page 14306-14313.

[45] D.M. Eckert and P.S. Kim. Annual Review of Biochemistry; 2001/ 70. *Mechanism of viral membrane fusion and its inhibition*. Page 777-810.

[46] Xun Zhao et al. Proceedings of the National Academy of Science of the United States of America; 2000/ 97 (26). *Structural characterization of the human respiratory syncytial virus fusion protein core.* Page 14172-14177.

[47] Zhiwu Sun et al. Viruses; 2013/ 5 (1). *Respiratory Syncytial Virus Entry Inhibitors Targeting the F Protein*. Page 211-225.

[48] Jacqueline M. Methews et al. Journal of Virology; 2000/ 74 (13). *The Core of the Respiratory Syncytial Virus Fusion Protein Is a Trimeric Coiled Coil.* Page 5911-5920.

[49] Luis González-Reyes et al. Proceeding of the National Academy of Sciences of the United States of America; 2001/ 98 (17). *Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion.* Page 9859-9864.

[50] Amanda E. Gardner and Rebecca E. Dutch. Journal of Virology; 2007/ 81 (15). *A Conserved Region in the F2 Subunit of Paramyxovirus Fusion Proteins Is Involved In Fusion Regulation.* Page 8303-8314.

[51] Jörg Schlender et al. Journal of Virology; 2003/ 77 (8). *Respiratory Syncytial Virus (RSV) Fusion Protein Subunit F2, Not Attachment Protein G, Determines the Specificity of RSV Infection.* Page 4609-4616.

[52] S. Techaarpornkul et al. Journal of Virology; 2001/ 75 (15). *Functional analysis of recombinant respiratory syncytial virus deletion mutants lacking the small hydrophobic and/or attachment glycoprotein gene.* Page 6825-6834.

[53] Jörg Schlender et al. Journal of Virology; 2003/ 77 (8). *Respiratory Syncytial Virus (RSV) Fusion Protein Subunit F2, Not Attachment Protein G, Determines the Specificity of RSV Infection.* Page 4609-4616.

[54] The QIAexpressionist handbook, page 71. www.quiagen.com

[55] Susanne Gräslund et al. Nature Methods-Author manuscript; 2008/ 5 (4). *Protein production and purification.* Page 135-146.

[56] Victor Martin Bolanos-Garcia et al. Biochimica et Biophysica Acta; 2006/ 1760 (9). *Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography.* Page 1304–1313.

[57] Victor Martin Bolanos-Garcia et al. Biochimica et Biophysica Acta; 2006/ 1760 (9). *Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography.* Page 1304–1313.

[58] http://www.piercenet.com/method/peptide-synthesis (20.05.2014)

[59] http://wolfson.huji.ac.il/purification/PDF/ReversePhase/VYDAChandbookRPC.pdf (25.05.2014)

[60] http://www.mayomedicallaboratories.com/articles/communique/2013/01-maldi-tof-mass-spectrometry/index.html (20.05.2014)

[61] Norma J. Greenfield. Nature Protocols; 2006/ 1 (6). *Using circular dichroism spectra to estimate protein secondary structure.* Page 2876-2890.

[62] Ronald B. Corley. *A Guide to Methods in the Biomedical Sciences*; Springer Science + Business Media Inc.; 2005. Page 7.

[63] http://www.chemeurope.com/en/encyclopedia/SDS-PAGE.html (20.05.2014)

[64] Institute of Technology and Management. The Journal of Department of Applied Sciences & Humanities; 2007 (Vol. 6). *SDS-PAGE-Sodium dodecyl sulfate polyacrylamide gel electrophoresis.* Page 98-100. http://www.itmindia.edu

[65] Ronald B. Corley. *A Guide to Methods in the Biomedical Sciences;* Springer Science + Business Media Inc.; 2005. Page 8.

[66] U.K. Laemmli. Nature 1970/ 227 (5259). *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Page 680-685.

[67] Tahrin Mahmood and Ping-Chang Yang. Northern American Journal of Medical Sciences; 2012/ 4 (9). *Western Blot: Technique, Theory and Trouble Shooting.* Page 429-434.

[68] Xianwen Lou et al. Journal of The American Society for Mass Spectrometry; 2013/ 24 (9). *Unusual Analyte-Matrix Adduct Ions and Mechanism of Their Formation in MALDI TOF MS of Benzene-1,3,5-Tricarboxamide and Urea Compounds.* Page 1405-1412.

[69] Xiangping Zhu and Ioannis A. Papayannopoulos. Journal of Biomolecular Techniques; 2003/ 14(4). *Improvement in the Detection of Low Concentration Protein Digests on a MALDI TOF/TOF Workstation by reducing α-Cyano-4-hydroxycinnamic Acid Adduct Ions*. Page 298–307.

[70] Jacqueline M. Methews et al. Journal of Virology; 2000/ 74 (13). *The Core of the Respiratory Syncytial Virus Fusion Protein Is a Trimeric Coiled Coil.* Page 5911-5920.

[71] The QIAexpressionist; June 2003. *A handbook for high-level expression and purification of 6xHis-tagged proteins*. Page 52.

[72] http://www.lifetein.com/blog/peptide-antigen-design/ (26.5.2014)

[73] http://www.piercenet.com/method/peptide-design (26.5.2014)

[74] http://www.lifetein.com/blog/peptide-antigen-design/#sthash.EE5xid39.dpuf (26.5.2014)

[75] http://www.piercenet.com/method/peptide-design (26.5.2014)

[76] N. G. Davis and M. C. Hsu. Proceedings of the National Academy of Science of the United States of America; 1986 /83 (14). *The fusion-related hydrophobic domain of Sendai F protein can be moved through the cytoplasmic membrane of Escherichia coli.* Page 5091-5095.

[77] Antonia Martin-Gallardo et al. Virology; 1991/ 184 (1). *Expression of the F Glycoprotein Gene from Human Respiratory Syncytial Virus in Escherchia coli: Mapping of a Fusion Inhibiting Epitope.* Page 428-432.

[78] Helen A. Arcuri et al. Elsevier; Protein Expression and Purification 2008/ 62 (2). *Expression and purification of human respiratory syncytial virus recombinant fusion protein.* Page 146-152.

[79] The QIAexpressionist; June 2003. *A handbook for high-level expression and purification of 6xHis-tagged proteins.* Page 52.

[80] Victor Martin Bolanos-Garcia et al. Biochimica et Biophysica Acta; 2006/1760 (9). *Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography.* Page 1304–1313.

[81] Oresta Vytvytska et al. Proceedings of the National Academy of Science of the United States of America; 1998/ 95 (24). *Host factor I, Hfq, binds to Escherichia coli ompA mRNA in a growth rate-dependent fashion and regulates its stability.* Page 14118-14123.

[82] Susanne Gräslund et al. Nat Methods Author manuscript; Feb 2008/ 5 (2). *Protein production and purification*. Page 135-146.

[83] The QIAexpressionist; June 2003. *A handbook for high-level expression and purification of 6xHis-tagged proteins*. Page 52.

[84] M. Satake et al. Nucleic Acids Research; 1985/ 13 (21). *Respiratory syncytial virus envelope glycoprotein (G) has a novel structure.* Page 7795-7812.

[85] Antonia Martin-Gallardo et al. Journal of General Virology; 1993/ 74 (3). *Expression of the G glycoprotein gene of human respiratory syncytial virus in Salmonella typhimurium.* Page 453-458.

[86] J. K. Rose and A. Shafferman. Proceedings of the National Academy of Science of the United States of America; 1981/ 78 (11). *Conditional expression of the vesicular stomatitis virus glycoprotein gene in Escherichia coli.* Page 6670-6674.

[87] Susanne Gräslund et al. Nature Methods-Author manuscript; Feb 2008/ 5 (2). *Protein production and purification.* Page 135-146.

[88] Wiley-VCH Verlag GmbH & Co KGaA 2011. Chapter 1/ James Samuelson/ Bacterial Systems. Editor: Anne Skaja Robinson. *Production of Membrane Proteins: Strategies for Expression and Isolation*; Editor: Anne Skaja Robinson. Page 24.

[89] S.H. Pan and B.A. Malcolm. Biotechniques; 2000/ 29 (6). *Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3).* Page 1234 – 1237.

[90] D. R. Wycuff and K.S. Matthews. Elsevier; Analytic Biochemistry 2000/ 277 (1). *Generation of an AraC-araBAD promoter-regulated T7 expression system.* Page 67-73.

[91] Antonia Martin-Gallardo et al. Journal of General Virology; 1993/ 74 (3). *Expression of the G glycoprotein gene of human respiratory syncytial virus in Salmonella typhimurium.* Page 453-458.

[92] Christophe Merlin et al. Journal of Bacteriology; 2003/185 (21). *Why is carbonic anhydrase essential to Escherichia coli?* Page 6415–6424.

[93] Jacqueline M. Methews et al. Journal of Virology; 2000/ 74 (13). *The Core of the Respiratory Syncytial Virus Fusion Protein Is a Trimeric Coiled Coil.* Page 5911-5920.

Curriculum Vitae

CONTACT INFORMATION

Name:	Borochova Kristina
E-Mail:	kristina84@gmx.at

PERSONAL INFORMATION

Date of Birth:	19.11.1984
Place of Birth:	Dushanbe/Tajikistan
Citizenship:	Austria

EDUCATION

2013	-	2014	Master thesis: *"Expression and purification of recombinant surface proteins G and F of Human Respiratory Syncytial Virus (HRSV)"* Supervisor: Prof. Dr. Rudolf Valenta Institute of Pathophysiology and Allergy Research, 3Q Medical University of Vienna, Spitalgasse 23, 1090 Vienna
2012	-	2014	Master degree course molecular Microbiology and Immunology University of Vienna, Dr. Karl-Lueger-Ring 1; 1010 Vienna Max F. Perutz Laboratories, Dr. Bohrgasse 9, 1030 Vienna Medical University of Vienna, Spitalgasse 23, 1090 Vienna
2008	-	2012	Bachelor degree course biology (Microbiology and Genetics) Bachelor thesis: *"Spt6 and H2A.Z und ihre Auswirkung auf die Kinetik der Stressgentranskription"* Supervisor: Priv. Doz. Dr. Christoph Schüller University of Vienna, Dr. Karl-Lueger-Ring 1; 1010 Vienna Max F. Perutz Laboratories, Dr. Bohrgasse 9, 1030 Vienna
2004	-	2008	Bfi Wien, Margaretenstraße 65; 1050 Vienna (University entrance exam)
2001	-	2004	Nursing-School AKH Wien, Lazarettgasse 14; 1090 Vienna
1999	-	2001	Realgymnasium, Landstraße Hauptstraße 70; 1030 Vienna

1995	-	1999	secondary modern school, Alsegger Straße 45; 1180 Vienna
1992	-	1995	elementary school, Bischof Faber-Platz 1; 1180 Vienna
1990	-	1992	elementary school, Termez/Uzbekistan

EMPLOYMENT HISTORY

2013	-	2014	Department of Pathophysiology and Allergy Research, Währingergürtel 18-19/ 3Q, 1090 Vienna Practical work for my master thesis
2008	-	2013	Wiener Rotes Kreuz, Nottendorfergasse 21, 1030 Vienna Field of work home health care
2004	-	2008	AKH Wien, OP-Group 2 Cardiac-Thorax-Vascular Surgery. Währingergürtel 18-19/ Ebene 9 OP-Gr. 2, 1090 Vienna Field of work anaesthesia nurse with additional function as a hygiene nurse

ADDITIONAL QUALIFICATIONS

Languages:	German, English and Russian
Computer:	Microsoft Office, Adobe Illustrator, Photoshop, Gene-Doc, Phylogenetic analysis tools.