OPTIMIZATION OF PROTEIN PRODUCTION DEVELOPMENT USING A COMBINATION OF CELL-FREE EXPRESSION AND HIGH-THROUGHPUT PROTEIN ANALYSIS

zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr. –Ing.)

der Fakultät für Chemieingenieurwesen und Verfahrenstechnik des

Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl. –Ing. Carolin Richter

aus Bruchsal

Referent: Prof. Dr. Jürgen Hubbuch

Korreferent: Prof. Dr. Christoph Syldatk

Tag der mündlichen Prüfung: 12.09.2014

Acknowledgements

This thesis is the result of my work at the DIARECT AG in Freiburg (im Breisgau) and at the Karlsruhe Institute of Technology. It would not have been possible to finish this project without the help and support of many people. Therefore, I would like to thank:

My doctoral advisor Prof. Dr. –Ing. Jürgen Hubbuch who gave me the opportunity to work on this PhD project. I would also like to thank him for his interest in the "very" biological topics I worked with and his time for many discussions about the progress of my work.

The CEO Prof. Dr. Liedvogel and the president of the DIARECT AG Dr. Haubruck for the opportunity to work in a multi-disciplinary team in the company's office and laboratories.

A very special thanks goes to Dr. Iris Asen, who supported me in every way someone can think of. Her endless time for discussions about my work, her high engagement, her critical eye on presentations and written work and her patience, motivated and encouraged me during ups and downs and gave my work the special features.

Furthermore I would like to thank my colleagues at the DIARECT AG. They always gave me thoughtful answers of every question I could think of. They really made my time special. I would like to thank Dr. Thomas Simon for his welcome in the molecular laboratory and his patience when I had other ideas how to organize the lab. His endless time for discussions and explanations especially about the biological side of my work. Dr. Richard Kneusel for his help during my work. Dr. Armin Neininger for his help with the codon usage of U1-68/70 and the organizations of the field days. Dr. Heike Schwarz for her help with my cell-cultures and a great time in the same office. Dr. Kosta Konstantinidis for his patience, teaching me how the printer works and the never ending proof-reading. Martina Reime and Haixa Ding for their office-sharing time. Maren Schütz and Jana Zink for their friendship and open ears when I was down and simply needed them to talk to, as well as all the cooking and jogging evenings. Vladimir Kempf for the jogging hours.

Dr. Anna Siudak for her help with all modeling based and statistical questions, her time, honesty and her encouraging words.

Natalie Rakel for her endless discussions and support about research and human topics while drinking coffee or on the telephone. Her encouragement when I was down and for her long lasting friendship.

For the friendly atmosphere and for giving me always a helping hand, I would also like to thank my colleagues at the KIT. Dr. Florian Dismer, Jörg Kittelmann, Katharina Lang, Kristina Schleining and Marc Hoffmann.

I would like to thank Frank Hämmerling and Christopher Ladd for the funny office time we had and the flower shopping tours.

My Master student Fabian Bickel for his interest in his project, his energy and patience when I had other ideas.

Outstanding thanks goes to my grandparents, parents and friends, who always believed in me, supported me during my research work and the writing part in all different ways one can think of. I'm sure that this was not always easy and I'm glad to have these loving people around me.

My dearests thank goes to Kai, who let me go to Freiburg and accepted a long-distance relationship. He always was my backing and supported me with his love and care. I love you!

What constitutes success?

He has achieved success who has lived well, laughed often, and loved much;
Who has enjoyed the trust of pure women, the respect of intelligent men and the love of little children;
Who has filled his niche and accomplished his task;
Who has never lacked appreciation of Earth's beauty or failed to express it;
Who has left the world better than he found it,
Whether an improved poppy, a perfect poem, or a rescued soul;
Who has always looked for the best in others and given them the best he had;
Whose life was an inspiration;
Whose memory a benediction.

Bessie Anderson Stanley, 1904

Abstract

This thesis deals with the optimization of high-throughput cell-free protein expression and subsequent protein analysis as well as the combination of these methods. The research outcome contributes to help simplifying and accelerating the biochemical protein production and analysis. The conventional cell-based expression methods are limited in expressing gene sequences, which could not yet been assigned by defined functions. Using cell-free expression, proteins with special properties, which are of high value in the personalized medicine and pharmaceutical research, can be produced from these gene sequences. In this research, not only the optimization of cell-free expression systems and their use in various protein expressions, but also the subsequent analysis of the proteins were examined. Each of the two steps, protein production and protein analysis, were further developed from laboratory scale to high-throughput methods to enlarge the number of experiments in a shorter timeframe with less material costs. To demonstrate the applicability of the new method, the results were compared to cell-based expression experiments.

In recent years, in biotechnology and particularly the field of proteomics, the demand for faster and easier methods for protein production has grown. This methodology expansion is focused on making genetic information more efficiently useable. Here, cell-free expression plays a key role as it provides a powerful technological platform to prepare proteins from DNA templates [1–3]. Cell-free expression consists of two steps, the production of the cell extract and the cell-free protein-producing reaction. At the beginning of this study cell lysates from three different genetically modified bacterial cells of the strain *Escherichia coli* and two eukaryotic insect cell types (Sf9 and High Five™) were prepared. The cell disruption method and further processing of the cell lysate contribute to increased protein yield; therefore three different methods (pressure digestion by French press or cell homogenization or freezing / thawing) were performed and optimized. Furthermore, the lysates differ in protein expression characteristics, depending on the cell type and strain. For example, the RNaseE mutant of the *E. coli* strain BL21 Star™ (DE3) proved to be very efficient in cell-free protein expression. The gene sequence of the RNaseE enzyme, which is usually responsible for the destruction of foreign mRNA, was removed and therefore the production of this enzyme is suppressed. Cell-free protein expression using this lysate is advantageous as the newly formed mRNA is not destroyed by the cell lysate

enzymes. By expression screenings in various cell-free expression systems, the cell-free expression, depending on the desired application of the protein, and the protein type has been selected to produce proteins either in a high yield or with the desired post-translational modifications.

The basis of all protein expressions is the gencoding DNA sequence that is translated during the reaction into the respective protein. In this study, only plasmids were used as DNA carriers. All plasmids contained the T7 promoter system, which produces high RNA levels using the T7 RNA polymerase and thus laying the basis for a high amount of protein. The cDNA sequences were cloned into the pET24d vector for the *E. coli* expression and into the plasmid pDT1 [4] for insect cell-free reaction. This differentiation between the cell types and the usage of these two vectors has proven to be appropriate to enhance the protein yield, which was shown in studies carried out at the beginning of this thesis. The DNA sequences of the respective proteins usually contain an N- or C-terminal His-tag for the protein specific detection or purification. The DNA for a special protein attachment, the HaloTag® [5] was additionally cloned behind or in front of the coding sequence for the protein to specifically immobilize them on treated surfaces after expression.

The cell-free protein producing reaction (herein referred to as cell-free expression) consists universally of the same two parts, the transcription and translation. The difference of prokaryotes and eukaryotes is the location of the expression processes. Whereas in prokaryotes the transcription and translation run simultaneously within the cell, in eukaryotes transcription takes place in the cell nucleus and the translation is then performed in the intracellular fluid. During transcription, an mRNA template is created from a DNA template, which in turn is translated into a peptide chain. Depending on the protein and cell type, the peptide chain is modified either through post-translational modifications or a folding of the proteins including helper proteins. These final protein modifications may constitute the decisive factor for a successful deployment of the proteins, depending on the desired application of the produced proteins. For example, the addition of various folding proteins has been successfully tested and used in the cell-free expression with *E. coli*. In this study, all *E. coli* made cell-free expressions were performed as ‘coupled-reactions’, meaning that the transcription and translation are executed simultaneously in one reaction vessel. In contrast to that, the *in vitro* insect cell expressions were performed connected in ‘linked-reactions’. This means that the mRNA is transferred to a second reaction vessel after the transcription, to perform the translation separately.

Prior to the application of the cell-free expression systems for the production of various proteins was the particular characterization and optimization of the reaction compositions. A cell-free expression consists of the cell lysate, including the tRNAs, various enzymes and the ribosomes. For a successful protein expression T7 RNA polymerase, 20 proteinogenic amino acids, energy in form of GTP and ATP as well as an energy regenerating system have to be added. The concentrations of these substances as well as other additives like a buffer system can be varied. It is possible to optimize the composition of the cell-free expression for every protein. However, for all optimizations of the *E. coli* cell-free system at laboratory scale the firefly luciferase (*Photinus pyralis*) was used as a reporter protein. This protein was chosen, because not only the protein yield, but also the activity of a respective protein is really important. The active luciferase can be detected directly out of the cell-free reaction mixture. By adding a buffer to the reaction mixture, positive expression results of firefly luciferase can be determined qualitatively and quantitatively in a light reaction which is measurable with a luminometer.

In contrast to its multiple benefits the insect cell-free reaction is far less developed than the *in vitro* expression in *E. coli*. This might be true, because previous characterization and optimization studies of the insect cell-free reaction were based on manual ‘one–factor–at–a-time’ methods, which are expensive and time consuming. In this study the insect cell-free expression systems of the two cell types Sf9 and High Five™ have been reproducible implemented on a robotic platform. Again the firefly luciferase was used as a reporter protein. The experimental design was performed using a statistical method (DoE) and the results were evaluated by multivariate data analysis (MVDA). For every insect cell type, 566 different reaction mixtures compositions were pipetted with the robotic workstation. This resulted in experimental data which could be successfully adjusted for both cell types to a second degree polynomial, a ‘response surface’. Furthermore, both empirical models could be validated by additional experiments. The results of the *in vitro* translation characterization, including the influence parameters and their interactions as well as their influence on the expression yield, could be quantitatively calculated and presented in sensitivity charts. The applicability of this new method was confirmed by a comparison to factors from previous studies. Therefore, the model-based characterization may be applied to other cell-free systems for a detailed description and an increase of the protein yield.

Cell-free expression systems are often used to produce difficult to express proteins. These proteins can be either toxic to living cells or even inhibit the protein expression machinery of the cell. Here, the cell-free expression shows a tremendous advantage over conventional recombinant protein expression methods. If performing a high number of screening experiments, using cell-free expression can save a lot of time and material costs in contrast to cell-based expression. In this study, *E. coli* cell-free expression has been used to provide an *in vitro* expression optimization for the *in vivo* protein production of the full length U1-68/70 K protein. The autoantigen U1-68/70 K is the dominant diagnostic marker for the autoimmune 'Mixed Connective Tissue Disease', which could not be expressed in its full length form [6]. However, it was possible to produce a truncated version for the use as a diagnostic marker. With the use of the cell-free expression screening, the results of the *in vitro* expression could be successfully transmitted to the *in vivo* environment and thus the snRNP protein U1-68/70 K could be successfully produce in a soluble and full-length form in *E. coli* cells. The length and specificity of the protein was verified by Western blot analysis and an MS / MS approach. Furthermore, the reactivity of the protein has been tested and demonstrated for autoimmune diagnostics. The establishment of a cell-free expression system for the prediction of cell-based protein production parameters such as the applicability of the cDNA construct, the expression temperature or the folding properties can now be determined in a time and material saving manner.

As already stated previously, the cell-free expression is the method of choice when it comes to the production of a large number of proteins in a small scale. To identify and quantify high numbers of proteins and to analyze their function in biological processes, microarray assays are used. Microarray systems have great potential and are for example implemented in multiplex diagnostics. Lyme disease, caused by several species of *Borrelia*, is the most common tick-borne disease in North America and Europe. In Germany each year 1 million new infections are reported. In most countries, a two-test approach for the diagnosis of this infection is used. This includes a specific ELISA followed by an immunoblot. Since this technique is very expensive, an alternative testing is preferred. A method for linking the before optimized cell-free expression to the subsequent microarray printing of various *Borrelia* antigens on multi-well plates has been developed. In an *E. coli* cell-free system eleven immunodominant antigens of different *Borrelia* species have been successfully expressed and then purified. The reproducible immobilization on the microarray plates and the detection of antigen activity could be detected with the help of different blood sera from patients

suffering from Lyme disease and specific monoclonal antibodies. A comparison of the cell-free approach developed here to cell-based expressed, purified and printed *Borrelia* antigens confirmed the diagnostic value of the new assays. In summary, this approach serves as a ‘proof of principle’ for the identification of potential biomarkers and offers the possibility of a multiplex protein detection for diseases.

The research in this thesis shows different types of applications of cell-free expression. Despite the already optimized production processes for cell-free expression systems and the recently significantly improved reaction yields, the potential for optimization of cell-free expression systems is not yet exhausted. The linkage between protein production and protein analysis needs to be further examined in future studies to integrate folding proteins or other additives in cell-free expression to reach integration of desired protein properties for the analysis of diseases already on the level of protein production.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich mit der Hochdurchsatzoptimierung von zellfreier Proteinexpression, der nachfolgenden Proteinanalytik sowie einer Verknüpfung dieser Verfahren. Das Forschungsergebnis trägt dazu bei die biochemische Proteinherstellung und -analyse zu vereinfachen und zu beschleunigen. Die üblichen zellbasierten Expressionsmethoden können Gen-Sequenzen, denen bisher keine definierte Funktion zugeordnet werden konnte, teilweise nur unzureichend oder überhaupt nicht darstellen. Mit Hilfe der zellfreien Expression können aus diesen Gen-Sequenzen Proteine mit besonderen Eigenschaften, die vor allem in der personalisierten Medizin und Pharmaforschung von hohem Wert sind, hergestellt werden. Nach der in dieser Forschungsarbeit optimierten zellfreien Expression in verschiedenen Zellsystemen und dem Einsatz dieser zur Expression von diagnostischen Proteinsystemen, wurde weiterführend auch die Proteinanalytik untersucht. Bei den beiden Verfahrensschritten, Proteinherstellung und Proteinanalytik, erfolgte eine Weiterentwicklung vom Labormaßstab zur Hochdurchsatzmethode, um eine höhere Anzahl von Experimenten in kürzerer Zeit und mit weniger Materialkosten durchführen zu können. Um die Anwendbarkeit dieser neuartigen Verfahren zu zeigen, wurden die Ergebnisse der zellfreien Expression mit bisher gebräuchlichen zellbasierten Expressionen verglichen.

In den letzten Jahren ist vor allem in der Biotechnologie und dort im Bereich der Proteomik der Bedarf an einfacheren und schnelleren Verfahren zur Proteinherstellung gewachsen, um genetische Informationen effizienter nutzbar zu machen. Hierbei spielt die zellfreie Expression, mit deren Hilfe Proteine aus DNA Vorlagen hergestellt werden können, als sehr leistungsfähige Technologieplattform eine wichtige Rolle [1–3]. Sie besteht aus zwei Schritten, der Herstellung des Zellextrakts und der zellfreien proteinproduzierenden Reaktion. Aus drei unterschiedlich genveränderten Bakterienzellen des Stammes *Escherichia coli* (*E. coli*) sowie zwei eukaryotischen Insektenzellarten (Sf9 und High Five™) wurden zu Beginn dieser Arbeit Zelllysate hergestellt. Da die Proteinausbeute unter anderem von der Methode des Zellaufschlusses und der weiteren Verarbeitung des Zelllysates abhängig ist, wurden drei verschiedene Verfahren zur Zelllyse (Druckaufschluss durch French Press oder Zellhomogenisator bzw. Einfrieren und Auftauen) angewandt und im Hinblick auf die Proteinausbeute optimiert. Weiterhin unterscheiden sich die Lysate selbst je nach Zelltyp und Stamm im Hinblick auf die Proteinexpression. Zum Beispiel erwies sich die RNaseE Mutante des *E. coli* Stammes BL21 Star™ (DE3) als sehr effizient für die zellfreie Proteinexpression. Die Gensequenz für das Enzym RNaseE

wurde hierbei entfernt und somit auch die Proteinherstellung dieses Enzyms unterdrückt, welches normalerweise für die Zerstörung jeglicher Fremd-mRNA zuständig ist. Somit erfolgt bei einem Zelllysat mit dieser Eigenschaft erheblich weniger Abbau der neu gebildete mRNA durch das Enzym RNAse Je nach gewünschtem Einsatz und Art der Proteine, wurde durch Expressionsscreenings in den verschiedenen zellfreien Expressionssystemen für jedes Protein die zellfreie Expression gewählt, die die Proteine entweder in hoher Ausbeute oder mit den gewünschten posttranslationalen Modifikationen produziert.

Die Grundvoraussetzung aller Proteinexpressionen ist die eingesetzte gencodierende DNA-Sequenz, die während der Reaktion in das jeweilige Protein übersetzt wird. In dieser Studie wurden als DNA-Träger ausschließlich Plasmide genutzt. Diese enthielten alle das T7-Promotorsystem, welches mittels T7-RNA-Polymerase hohe RNA Mengen produziert und somit die Voraussetzung für eine hohe Proteinmenge legt. Die cDNA-Sequenzen wurden für die *E. coli*-Expression in den pET24d-Vektor und für die insektenzellfreie Reaktion in das Plasmid pDT1 [4] kloniert. Diese Unterscheidung zwischen den Zelltypen und die Festlegung auf diese beiden Vektoren hat sich in entsprechenden, zu Beginn durchgeführten Studien, als ausbeutesteigernd erwiesen. Die cDNA-Sequenzen der jeweiligen Proteine enthielten meist zusätzlich einen N- oder C-terminalen His-Tag zur spezifischen Proteindetektion oder -aufreinigung. Teilweise wurde auch zusätzlich ein spezieller Proteinanhang, der HaloTag® hinter oder vor die codierende Proteinsequenz kloniert [5], um die Proteine nach der Expression spezifisch auf entsprechend behandelten Oberflächen immobilisieren zu können.

Die zellfreie proteinproduzierende Reaktion, im Weiteren als zellfreie Expression bezeichnet, besteht zellunabhängig aus zwei Teilen, der Transkription und der Translation. Der Unterschied zwischen Prokaryoten und Eukaryoten ist der Expressionsort. Während in den Prokaryoten die Transkription und Translation gleichzeitig innerhalb der Zelle ablaufen, finden die Transkription bei Eukaryoten im Zellkern und die nachfolgende Translation im Zellplasma statt. Während der Transkription wird aus einer DNA-Vorlage eine mRNA-Vorlage erstellt, welche wiederum in der sogenannten Translation in eine Peptidkette übersetzt wird. Danach werden je nach Proteintyp und Zellart teilweise posttranslationale Modifikationen an die Peptidkette angehängt bzw. es findet eine Faltung der Proteine mittels Helferproteinen statt. Je nach gewünschter Anwendung der hergestellten Proteine können diese finalen Proteinmodifikationen den Ausschlag für einen erfolgreichen Einsatz der Proteine darstellen. Zum Beispiel wurde der Zusatz verschiedener bekannter Faltungshelferproteine bei der zellfreien Expression mit *E. coli* erfolgreich getestet und eingesetzt. In dieser Studie erfolgten alle *E. coli*-zellfreien Expressionen

in Form einer ‚coupled-reaction', d.h. dass die Transkription und Translation gleichzeitig in einem Reaktionsgefäß abliefen. Im Gegensatz dazu wurden die *in-vitro*-Insektenzellreaktionen verbunden (‚linked-reaction') betrieben. Das bedeutet, dass die mRNA nach der Transkription im Reaktionsgefäß in ein zweites überführt wird, um danach die Translation durchzuführen. Vor dem Einsatz der beiden zellfreien Expressionssysteme zur Produktion verschiedener Proteine, erfolgte die jeweilige Charakterisierung und Optimierung der Reaktionszusammensetzungen. Eine zellfreie Expression besteht aus dem im Vorigen beschriebenen präparierten Zelllysat, welches unter anderem die tRNAs, verschiedene Enzyme und die zelleigenen Ribosomen enthält. Zur erfolgreichen Proteinexpression müssen die T7-RNA-Polymerase, die 20 proteinogenen Aminosäuren, Energie in Form von GTP und ATP sowie ein energieregenerierendes System zugesetzt werden. Die Konzentrationen dieser Substanzen sowie weiterer Zusätze, wie z.B. eines Puffersystems können variiert werden. Es ist möglich die Zusammensetzung der zellfreien Expression für jedes zu exprimierende Protein zu optimieren. Exemplarisch wurde allerdings bei den Optimierungen des *E. coli*-zellfreien Systems, welche im Labormaßstab erfolgte, die Firefly Luciferase (*Photinus pyralis*) als Beispielprotein genutzt. Die Wahl fiel auf dieses Protein, da nicht nur der Proteinausbeute, sondern auch der Aktivität des jeweiligen Proteins eine große Bedeutung zugeschrieben wird. Die aktive Luciferase kann direkt aus dem zellfreien Reaktionsgemisch heraus *in vitro* detektiert werden. Durch Zugabe eines Puffers zum Reaktionsgemisch entsteht bei positiver Luciferase-Expression eine Lichtreaktion, die mittels eines Luminometers qualitativ und quantitativ ermittelt werden kann.

Im Widerspruch zu ihren vielfältigen Vorteilen ist die insektenzellfreie Reaktion bisher weniger gut etabliert als die *in-vitro*-Expression in *E. coli*. Dies könnte auch daran liegen, dass bisherige Charakterisierungen und Optimierungen der insektenzellfreien Reaktion auf manuellen ‚one-factor-at-a-time'-Methoden basieren, die teuer und zeitaufwändig sind. In dieser Thesis wurden die insektenzellfreien Expressionssysteme der beiden Zellenarten Sf9 und High Five™ auf einer Roborterplattform reproduzierbar implementiert. Als Reporterprotein kam auch hier wieder die Firefly Luciferase aus den vorher genannten Gründen zum Einsatz. Die Versuchsplanung erfolgte mittels statistischer Methodik (DoE) und das Ergebnis wurde durch eine multivariate Datenanalyse (MVDA) ausgewertet. Pro Zelltyp wurden 566 verschieden zusammengesetzte Reaktionsansätze mit Hilfe des Roboters pipettiert. Daraus resultierten für beide Zelltypen experimentelle Daten, die erfolgreich an ein Polynom zweiten Grades, eine ‚Response Surface', angepasst werden konnten. Weiterhin konnten beide empirischen Modelle durch zusätzliche Reaktionsansätze validiert werden. Das Ergebnis der *in vitro*-Translations-Charakterisierung durch den Einbezug der Einflussparameter und deren Interaktionen

sowie ihres Einflusses auf die Expressionsausbeute konnte quantitativ berechnet und in Sensitivitätsdiagrammen dargestellt werden. Die Anwendbarkeit dieser neuen Methode wurde durch einen Vergleich der Einflussgrößen zu früheren Studien bestätigt. Daraus folgt, dass die modellbasierte Charakterisierung jederzeit auf weitere zellfreie Systeme zur näheren Beschreibung und Steigerung der Ausbeute angewandt werden kann.

Zellfreie Expressionssysteme werden häufig dazu genutzt, schwer zu exprimierende Proteine herzustellen. Diese Proteine sind oft entweder für lebende Zellen toxisch oder hemmen den Proteinexpressionsapparat der Zellen. Hierbei besitzt die zellfreie Expression einen enormen Vorteil gegenüber herkömmlicher rekombinanter Proteinexpression. Weiterhin kann mit Hilfe der zellfreien Expression eine hohe Anzahl von Screeningexperimente durchgeführt werden, welche im Vergleich zu zellbasierter Expression viel Zeit und Materialkosten sparen. Im weiteren Verlauf der Arbeit wurde die *E. coli* zellfreie Expression dazu genutzt, durch eine *in-vitro*-Expressionensoptimierung die *in-vivo*-Proteinproduktion des Proteins U1-68/70 K in seiner vollen Länge zu ermöglichen. Das Autoantigen U1-68/70 K ist ein dominierender diagnostischer Marker der Autoimmunerkrankung ‚Mixed connective tissue disease', der bis vor kurzem nicht in seiner vollen Länge exprimiert werden konnte [6]. Es war allerdings möglich eine verkürzte Version als diagnostischen Marker zu produzieren. Durch den Einsatz eines zellfreien Expressionsscreenings ist es gelungen, die Ergebnisse der *in-vitro*-Expression auf das *in-vivo*-Umfeld zu übertragen und damit das snRNP Protein U1-68/70 K in seiner löslichen Form in voller Länge erfolgreich in *E. coli*-Zellen zu produzieren. Die Proteinlänge und -spezifität konnte durch Western-Blots und eine MS/MS-Analyse verifiziert werden. Zusätzlich wurde die Reaktivität in der Autoimmun-Diagnostik getestet und nachgewiesen. Durch die Etablierung eines zellfreien Expressionssystems zur Voraussage von Proteinproduktionsparametern, wie z.B. der Anwendbarkeit des cDNA-Konstrukts, der Expressionstemperatur und den Faltungseigenschaften in zellbasierenden Systemen, können zukünftig die vorher genannten Parameter Zeit und Material sparenden bestimmt werden.

Wie schon im Vorigen ausgeführt, ist die zellfreie Expression die Methode der Wahl, wenn es um die Produktion einer großen Anzahl an Proteinen in kleinem Maßstab geht. Um diese große Anzahl von Proteinen zu identifizieren und zu quantifizieren, sowie ihre Funktion in biologischen Prozessen zu analysieren, werden Assays im Mikroarrayformat verwendet. Mikroarraysysteme besitzen ein sehr großes Potenzial und werden zum Beispiel in der Multiplex-Diagnostik angewendet. Lyme-Borreliose ist die häufigste durch Zecken übertragene Krankheit in Nordamerika und Europa. In Deutschland werden jährlich 1 000 000 Neuinfektionen gemeldet. Für die Diagnostik

dieser Infektion wird in den meisten Ländern ein Zwei-Test Ansatz genutzt. Dieser beinhaltet einen ELISA gefolgt von spezifischen Immunoblots. Da diese Technik sehr aufwendig ist, wurde in dieser Arbeit nach einer Alternative gesucht. Durch die Kombination von zellfreier Expression und der Mikroarraytechnik wurde ein Verfahren für die Expression und das anschließende Drucken verschiedener *Borrelia*-Antigene auf Multi-Well-Platten erarbeitet. In einem eigens hergestellten *E. coli*-zellfreien System wurden elf immundominante Antigene der Lyme-Borreliose von verschiedenen *Borrelia*-Arten erfolgreich exprimiert und anschließend gereinigt. Mit Hilfe von Blutseren von an Borreliose erkrankten Patienten und spezifischen monoklonalen Antikörpern, konnten die erfolgreiche Immobilisierung und Aktivität der Antigene reproduzierbar auf den Mikroarrayplatten nachgewiesen werden. Ein Vergleich zu zellbasiert exprimierten, gereinigt und gedruckten *Borrelia*-Antigenen bestätigt die diagnostische Aussagekraft des neuen Assays. Zusammenfassend ist zu sagen, dass dieser Ansatz als ‚proof of principle' für die Identifizierung von potentiellen Biomarkern dient und die Möglichkeit einer Multiplex-Protein-Erkennung für Krankheiten bietet.

Wie diese Arbeit zeigt, sind die Arten der Anwendung der zellfreien Expression sehr vielfältig. Trotz bereits optimierter Herstellverfahren für zellfreie Expressionssysteme und in den vergangenen Jahren erheblich verbesserter Reaktionsausbeuten, ist das Potenzial der Optimierung von zellfreien Expressionssysteme noch lange nicht erschöpft. Auch die Vernetzung von Proteinherstellung und Proteinanalyse muss noch weiter untersucht werden, um zukünftig durch Zusatz von Faltungsproteinen oder weiteren Substanzen zur zellfreien Expression die gewünschten Proteineigenschaften für die Analyse von Krankheiten bereits auf der Proteinproduktionsebene zu integrieren.

Contents

1 Introduction

The objective of this work was the optimization and subsequent application of cell-free expression to improve protein production, where cell-based systems have natural boundaries. Not only the increase of protein yield, but the combination of protein expression and protein analysis in a high-throughput mode was a goal. Overall, an improvement of biochemical protein production based on this combination is desired.

1.1 Cell-free expression

The terms ‘cell-free expression’ or ‘*in vitro* protein synthesis’ are nowadays on everyone’s lips. Cell-free expression is a kind of protein production without living cells. The expression process is taking place in an artificial environment; specified as an “open” system [7]. This term is used, because the protein is expressed outside the cell without having the compartment of a cell membrane. This implies advantages and disadvantages.

1.1.1 Advantages and disadvantages of cell-free expression

The production of small quantities of proteins can be performed quickly and economically with cell-free expression. This leads to their adaptability to high-throughput experiments, in which high numbers of experiments are desired. In cell-free expression systems, no cell viability concerns are necessary. Therefore, toxic proteins, which would destroy the cell metabolism or simply the expression apparatus, can be produced. Before cell-free expression techniques were developed, biochemical and structural characterization of membrane proteins for example have been in its infancy. Nowadays, sufficient amounts of functional membrane proteins even for crystallography and biochemical analysis are producible [7,8]. In cell-free expression systems additives like detergents, metal ions, cofactors or binding partners can simply be added to the expression reaction. This can only be accomplished by owning the “open” system characteristic. Additionally, incorporating isotopic labels and non-natural amino acids into the peptide chain of the produced proteins is easy. Even the simultaneous expression of more than one protein in one cell-free reaction is possible. For example, heterotrimers of human laminin-322 LCC domains were successfully produced in a cell-free system. Furthermore, the three peptide chains were formed and assembled in the *in vitro* reaction [9].

A disadvantage of cell-free expressions is the relatively low protein yield, depending on the expression system. Additionally, cell-free expression can be expensive, depending on the used system and if it is commercially purchased. One has to keep in mind, that the *in vitro* reaction has no sustained metabolism to convert cheap energy sources like sugars. Recently, glucose [10,11] or polymeric carbohydrates [12] were successfully used in cell-free expressions as an energy supply. Another disadvantage, which is solving itself by the years, is the less characterization of cell-free expression systems and the less usage experience in laboratories compared to organisms like *E coli*.

1.1.2 The fundament of cell-free expression – vector cloning

1.1.2.1 DNA and genetic engineering

In the DNA the information about the protein sequence and its localization in the organism is stored. Already in 1953, Watson and Crick developed that the information carrier for heredity is the DNA [13]. This universal code, being identical for pro- and eukaryotes, only varies in the adjustment of the four different nucleotides (adenosine, thymine, cytosine and guanine) in codon triplets. Each of these triplets is corresponding to one of only 20 amino acid building blocks which form the amino acid sequence, being the primary structure of all proteins. Nevertheless, the use of the genetic code (also called 'codon usage bias') is variable in different organisms. This variability refers to differences in the frequency of the occurring codons in the DNA sequence. Therefore, it is important to keep in mind that the codon usage of the protein expressing organism should be applied for any DNA template to reduce expression difficulties or unwanted stops.

The methods of genetic engineering are consisting of different basic applications such as isolation, replication, enzymatic modification and characterization, sequencing and chemical synthesis of the molecule DNA. A bacterial cell has two different types of DNA, the genome and a number of 50 to 100 plasmids. These plasmids are circular and with 3 000 to 100 000 nucleotide base pairs (bp) relatively small compared to the genome. Since plasmids are consisting of double-stranded DNA, they can be replicated on their own, but can also be integrated into the bacterial genome. They have a replication starting point and mostly one or two genes, which are important for the survival of the bacterial cell. For example, antibiotic resistance genes can usually be found on the plasmid DNA. By looking at the before mentioned characteristics of plasmids, they own the abilities for the use as DNA vectors, integrating foreign DNA into bacterial cells. Plasmids can be extracted from the bacterial cell and new gene sequences can be cloned into them *in vitro*. They are smaller than 10 000 bp for an

easy handling and by integrating antibiotic resistant genes, negative selection pressure can be established. Vector plasmids own an ‘origin of replication’ for a bacterial cell. If a transformation of the vector into other cell types is preferred, another ‘origin of replication’ can be added optionally to the DNA sequence of the vector plasmid. This would enable a following transformation of the vector into another host cell organism. For the cloning procedure, the plasmids have to be purified and opened with restriction enzymes on before determined positions. The foreign DNA is integrated into the ‘multiple cloning site’ where every ‘restriction enzymes cutting site’ exists only once to guarantee the correct insertion of a DNA piece cut with the identical restriction enzyme. The invention of the polymerase chain reaction (PCR) made it possible to produce extremely high numbers of DNA sequences *in vitro*. The three basic phases of a PCR are denaturation, annealing and synthesis of a DNA sequence after a DNA template. It became really important to generate specific DNA segments for cloning or sequencing, as well as detecting the presence of specific genetic defects. One disadvantage is that the enzyme Taq-polymerase [14], which synthesizes the DNA sequence with the nucleotide building blocks, doesn’t have proofreading abilities and can introduce errors. If this happens early in the process, the false DNA is also further amplified. Another problem using PCR is the possibility of contamination with false DNA fragments. If this happens and the primer sequences can also anneal to the foreign DNA it is amplified as well. But nevertheless, PCR is the method of choice for producing DNA fragments *in vitro* [15]. In addition to the insertion of the coding DNA sequence also the DNA sequence for protein tags to simplify purification or protein analysis is preferred. After the ligation of the foreign DNA sequence into the plasmid sequence, the plasmid needs to be replicated. Therefore, a transformation into a bacterial cell is necessary. There are different transformation methods. A chemical method uses $CaCl_2$ for the perforation of the cell wall to integrate the plasmid DNA [16]. Electroporation is another method to physically open up the cell wall. The analysis of a successful plasmid DNA transformation and foreign DNA sequence integration is performed with an analytical PCR run and a following agarose gel analysis. Here, the DNA sequence is analysed after its bp size. For protein expression in different host cells, many plasmid types were studied and optimised to increase protein yields. Most plasmid vectors were developed for *E. coli*, but there are also systems for *Bacillus, Pseudomonas, Streptomyces*, *Lactobacillus* and some others.

1.1.2.2 Template generation: Plasmids for cell-free expression

The choice of plasmid and the codon usage of the gene of interest correspond strongly to the host cell organism in which the protein is expressed originally. This is also true for cell-free expression, but is here more influenceable than in cell-based

expression. This is something one has to keep in mind when cloning a plasmid for *in vitro* protein synthesis. An alternative to plasmids are PCR-templates [1,17–19], because the time-consuming steps of plasmid generation, multiplication and purification are eliminated. For a protein expression, the plasmid or PCR-product template needs to contain a promoter sequence and a translation initiation signal at the 5'-region of the gene of interest [2,20–22]. Especially in eukaryotic cells this 5' untranslated region (UTR) is important for initiation, translation and mRNA stability [20]. The optimization of 5'-UTRs for eukaryotic cell-free expression systems is described by Ezure *et al.* [23,24]. For stopping the expression of a protein, a transcription and translation stop at a selected location, a so called termination sequence at the 3'UTR-region is necessary [2].

The most time consuming step while applying *in vitro* protein synthesis is the vector preparation. In order to reach high-throughput applications this step needs to be diminished. This can be reached through the application of PCR products instead of using plasmid as vectors for cell-free expression. The shortcoming of using PCR products is the problem of having less DNA and accordingly linear DNA instead of circular plasmids, which is degraded faster. This problem was solved in the past by either using multiply-primed rolling circle amplification [25] or applying optimized reaction conditions [26,27]. Thus comparable protein yields using PCR products vs. plasmids can be reached.

1.1.3 Cell-free extract preparation

1.1.3.1 Which kind of cell-free expression systems are available?

The question of the cell type to use for cell-free expression strongly depends on the end product application. Though, the cell-free expression system should be similar to the protein origin and biochemical background to achieve good results like a high protein yield, the necessary protein complexity, desired downstream conditions and low costs [3]. Overall, every cultivatable cell type can be used for extract preparation and therefore protocols for preparation and optimization of cell lysates from different cells have been studied to establish different cell-free expression systems and to increase the protein expression level [2]. To date *E. coli* [28,29], wheat germ [30,31], rabbit reticulocytes [32,33], yeast [34,35] and insect cells [36] are mostly applied for cell-free expression. A number of unusual hosts like *Leishmania tarentolae* [37], human HeLa-cells cells [38,39], *Drosophila* embryo [40], Xenopus oocyte or egg [41] and hyperthermophilic archaeon [42] have also been successfully prepared.

1.1.3.2 Coupled vs linked cell-free expression

The difference between coupled and linked cell-free expression is the expression template. While in coupled reactions a DNA template is used, linked systems have mRNA templates. This differentiation originates from the used cell systems. Prokaryotes are single-cell creatures which proliferate through splitting. They lack a nucleus and their DNA is balled into a nucleoid. This fact influences protein expression, because only one compartment for transcription and translation exists. While the transcription of the DNA takes place, the ribosomes can already start translation on the other end of the built mRNA. The complete process is running in the cytosol (Figure 1). This is where 'coupled cell-free reaction' comes from. In this expression type, the DNA is directly added into the cell-free expression system and after transcription and translation are over, the protein can be analyzed. This kind of *in vitro* expression is mostly applied with prokaryotic systems, but can also be adopted in eukaryotic systems. In contrast to prokaryotic cells, eukaryotes own a nucleus where the DNA is stored. Therefore, the protein production process is locally parted. The transcription process, in which the DNA sequence is read and the mRNA is built takes place in the nucleus. For the now following translation, the mRNA has to move out of the nucleus into the cytosol where the ribosomes can translate the nucleic acid sequence into a protein (Figure 1). This process is the template for the 'linked cell-free expression', where first the transcription takes place *in vitro* followed by a traditional purification of the mRNA and afterwards the translation can take place in another tube.

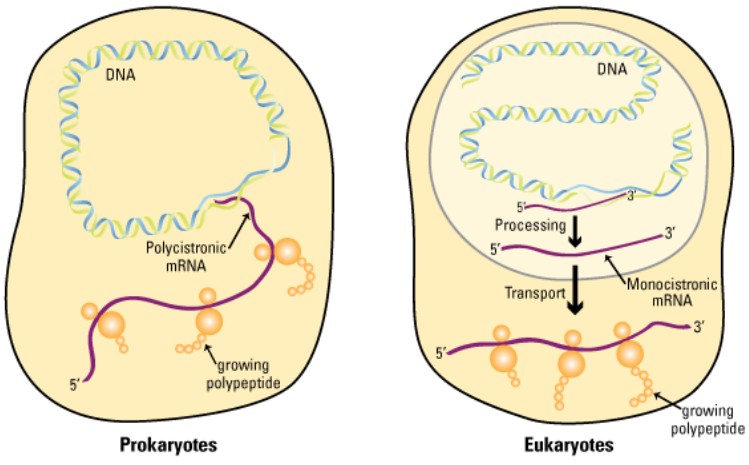


Figure 1: Prokaryotic (a) vs eukaryotic (b) cell: www.piercenet.com (thermo scientific)

1.1.3.3 Advantages and disadvantages of the different cell systems

The 'protein synthesis machinery' of different cell types may constitute of various cellular components (e.g. ribosomes, initiation and elongation factors, metabolic

enzymes) or co-factors (e.g. chaperones, foldases) which influence protein expression, folding and modifying proteins [2].

The fastest, most effective and cheapest system to date and therefore mostly used is the cell-free expression in *E. coli*. It is high-throughput compatible, tolerant to auxiliaries and the proteins expressed by this system are suitable for structural analysis [43,44]. High amounts of extract can be prepared easily since *E. coli* cells can be fermented at large scale. For the use of only a few expressions, *E. coli* cell-free extracts are commercially available [1]. Furthermore, well-established tools for modifications are available since *E. coli* cells are studied well [3]. The shortcomings of this system are in heterologous protein expression. *E. coli* cell-free system may produce truncated proteins, protein fragments or insoluble proteins and eukaryotic co- and posttranslational modifications are not possible. Their codon usage bias is different to eukaryotic cells, what can cause the aforementioned problems (see paragraph 1.1.2.1). It is difficult to generate folded eukaryotic proteins in this system [1,17,45,46].

Eukaryotic cell-based systems are required for the expression of correctly folded heterologous proteins which are also suitable for functional studies, because of owning post-translational modifications. But one has to keep in mind that they are less productive and more expensive than *E. coli* cell-free expression systems [41,42]. Rabbit reticulocytes and insect cells own the highest post-translational versatility for heterologous protein expression and are both commercially available. Furthermore, insect cells are a fast growing platform and their extract preparation in contrast to other eukaryotic cells is easy and quick [23]. Their disadvantage is that cell cultivation is expensive and time-consuming. Furthermore, they are non-mammalian cells and therefore not owning mammalian post-translational modifications, but still the insect cells can built many types of posttranslational modifications which have been functionally similar to authentic proteins [47]. Shortcomings of rabbit reticulocytes are that they have a low efficiency [48], are not able to glycosylate proteins [49] and their production is highly complex, because a manipulation of animal tissue is required. Furthermore, they often co-express unwanted byproducts. The highest protein yield even for complex proteins in eukaryotic cells is possible with wheat germ. They are relatively cheap compared to other eukaryotic cells [48], commercially available and even high-throughput compatible [50]. The disadvantages of this system are the little genetic modification tools, the lack of mammalian specific protein modifications and the laborious extract preparation [3].

In many studies cell-extract preparation methods from different cells were optimized, but it is still difficult to eliminate protein or nucleic acid degradation by proteases or nucleases. To date, inhibitor substances can suppress the activity of the natural present proteases and nucleases in cell extracts, but not eliminate them. In 2001, the *E. coli* PURE (Protein synthesis Using Recombinant Elements) system has been made. It constitutes of an *in vitro* reconstituted mixture composed of purified components. All 32 recombinantly produced components are linked to a histidine (His)-tag. Therefore, the possibility of an easier purification of the target protein using affinity chromatography is warranted. It still is a very expensive *in vitro* protein production system, but a valuable tool for studying the translation process under predefined conditions [51].

1.1.3.4 A short overview: How to make cell-free extracts

The basic steps of a cell-free extract preparation are shown in Figure 2. First, the cells need to be grown in a cell culture, followed by the cell harvest while they are still having a good viability and productivity, which is warranted while the cells are in a rapid growing phase [52,53]. The cell lysis depends on the cell type and can be performed in different ways. For example, *E. coli* cells are traditionally opened by a French press or high pressure homogenizations. Whereas eukaryotes only need liquid nitrogen in nitrogen bombs or a simple freeze and thawing method to be ruptured [23]. In the following centrifugation steps unwanted substances are eliminated, which also dependents on the extract type prepared. For *E. coli* cells a following incubation step can be performed which “activates” the extract. Recently, a new method for *E. coli* extract preparation was published, reducing time and effort [54]. Overall, the extract should be handled with care and be cooled during the complete process in order to prevent proteases and nucleases to reduce the expression abilities of the different extract types. At the end, a quick freezing step in liquid nitrogen and the storage at -80°C is essential.

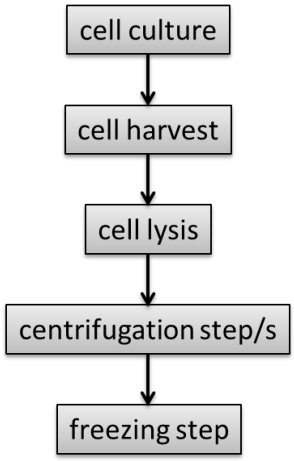


Figure 2: Cell extract preparation process

1.1.4 Cell-free reaction composition

Cell-free gene expression is a modern method of protein production. It is possible to assemble only the desired protein without having to deal with the protein production system especially with the metabolism of the cell. Still, the prepared cell lysate is the main substance used in cell-free expression. Up to the extract preparation method

important proteins for the following expression can be preserved and don't have to be added additionally. The *in vitro* expression systems are using the transcription and translation apparatus of the lysed cells. The most literally described and laboratory employed cell-system is still the *E. coli* system. The first and most cited paper about *E. coli* extract preparation and reaction composition is the publication of Pratt [55] from 1984. It explicates the coupled transcription and translation system derived from Zubay *et al.* [56] and firstly introduces the *E. coli* S30 extract which is named after the centrifugal step with 30 000 x g after cell rupture of the *E. coli* strain MRE600. Since then many improvements of the S30 extract system have been realised. For example in 1996 Kim *et al.* [57] introduced a new preparation method with condensed *E. coli* extract for higher protein yields, using the *E. coli* strain A19. Nowadays, every research group applies their own cell-free expression method by changing the preparation method slightly for individual needs. In 2004 Swarz *et al.* summarized the whole process of extract preparation and cell-free expression with *E. coli* cells in *Methods of Molecular Biology* still using the *E. coli* strain A19 [58]. The production of *E. coli* extracts from a different strain was published by Kigawa *et al.* in 2004 [59]. The specialisation went on to prepare extracts suitable for membrane protein expression [60]. The selective mutation of *E. coli* strain A19 and BL21 to repress RNA-degrading enzymes and therefore improve cell-free expression yields was also accomplished [61,62]. Decreased temperatures while incubating *E. coli* cells are known to improve productivity as well [63] because the RNAse activity are decreased. An iodoacetamide treated *E. coli* extract can efficiently introduce disulfide bonds into the expressed proteins [64]. Further on the improvements of the *E. coli* extract moved to a simplified and cost effective S12 system, which is supposed to include higher yields than the S30 extract [54]. In the new method, only a centrifugal step with 12 000 x g was applied in order to achieve more productivity and consistency of the extract. With this new method, different *E. coli* strains can be used. For example the BL21 Star™ (DE3) strain contains a mutation in the gene encoding RNaseE (*rne*131), which is one of the major sources of foreign mRNA degradation. Using this strain stabilizes the *in vitro* expression newly transcript mRNA which can lead to higher protein yields. The BL21 Rosetta strain is also optimized for protein production. It contains extra tRNAs for rarely used codons in *E. coli* in order to enhance the express results of mammalian proteins. In this study, both of the before mentioned strains were used for *E. coli* cell-free expressions.

Different reaction schemes are accomplished with coupled cell-free reactions. The continuous exchange cell-free system also known as the 'continuous flow system' was first introduced by Spirin in 1988 [29]. Furthermore the continuous exchange [65] and hollow fiber system [66] as well as a bilayer system were established [67]. But still the batch system is the most applied and for use in parallel, multiplexed and rapid protein

expression experiments appropriate system. The standard reaction mixture has only slightly changed since Pratt [55]. But many trials have been made to improve the secondary energy system to regenerate ATP throughout the cell-free expression. Hence, this has turned out to be the most critical reason behind the short duration and the low protein yields [68,69]. The improvements in energy supply make a wide stretch from fed-batch wise addition of energy sources and magnesium ions [69] over pyruvate as an energy source that does not accumulate inorganic phosphate [68]. With the invention of the pyruvate/CoA/NAD/oxalate (PANOX) system, which uses the addition of the cofactors NAD and CoA and the *E. coli* intrinsic enzymes pyruvate dehydrogenase (PDH) and phosphotransacetylase (PTA) to elongate the cell-free reaction as well as sodium oxalate to retard the non-productive degradation of phosphoenol pyruvate (PEP) [70], the first step of higher yield was set. Further investigations were made to use glycolytic intermediates as energy sources. It has been shown that glucose-6-phosphate, fructose-1,6-bisphosphate and 3-phosphoglycerate serve as efficient and less expensive energy sources [70,71]. Glucose itself could only be utilized, if the pH of the reaction is not decreasing during the *in vitro* protein synthesis [10]. Kim *et al.* used glucose as an energy source by employing an appropriate buffer system and the optimized S12 extract [72]. Protein synthesis in cell-free reactions could also be prolonged using a dual energy regenerating system. Creatine phosphate and creatine kinase as well as glucose were used to expand *in vitro* translation to 3 h with a 2-3 times higher protein yield in contrast to a single energy system [73]. All methods for energizing cell-free protein synthesis were reviewed by Kim and Kim 2009 [74]. A further improvement of the batch reaction was the application of large scale cell-free reactions [75,76].

1.2 Overview of protein analysis after cell-free expression

The analysis of protein structure and function keep science occupied since a long time. But the development of efficient purification strategies, with which a single protein could be purified out of protein mixtures and the revolutionary techniques of protein analysis, enabled our understanding of protein structure. Still protein analysis after cell-free expression can be challenging by finding a small amount of protein in a higher concentrated protein mixture like the cell-free expression premix. The traditional method of [^{35}S]methionine addition to the cell-free expression reaction and incorporation into the expressed protein allows the detection of the desired protein by autoradiography. Due to high costs, regulations, radioactive exposure, waste and

disposal issues many researchers are not executing this method anymore. Hence, a need for different new detection methods exists.

1.2.1 Gel-based methods

1.2.1.1 SDS-PAGE

In general, for separation of protein mixtures according to their molecular weight, SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is used. A linear relationship between the logarithm of the molecular mass and the migration routes of the SDS-polypeptide-micelles in certain areas is obtained. The use of protein standards can determine the molecular weights of the Coomassie stained proteins. However, Coomassie stained SDS-Pages detect all proteins and are therefore unsuitable for cell-free expression protein detection, because too many bands appear. Since cell-free expressed proteins are existent in low amounts in the expression mixture their detection is challenging. To date, the incorporation of a lysine-charged tRNA that is labeled at the ε position of the lysine with a fluorophore BODIPY®-FL into the emerging protein can be added to the cell-free reaction. This FluoroTect™ $Green_{Lys}$ *in vitro* Translation Labeling System (Promega GmbH, Mannheim) can be purchased and used for fluorescent labeling of proteins. After a SDS-PAGE electrophoresis, the nascent protein can be detected with a fluorescent scanner.

1.2.1.2 Western Blot

Western Blotting is another method of choice. In doing so, the separated proteins from the SDS-PAGE gel are electroblotted onto a nitrocellulose membrane. In the following an immune detection with two antibodies is attached. The primary antibody is a specific antibody to the protein which should be detected. The secondary antibody is a labeled antibody, which binds specifically to the first antibody. This method has a higher specificity than SDS-PAGE colored with Coomassie and is therefore a better choice for the detection of cell-free expressed proteins. But still, due to the specificity of the primary mAb the detection sensitivity can be too low. Therefore, a specific Western Blotting method for cell-free reactions was introduced by Promega. It is the Transcend™ Chemiluminescent Non-Radioactive Translation Detection System (Promega GmbH, Mannheim), which relies on the incorporation of a charged ε-labeled biotinylated-lysine-tRNA complex into nascent proteins during translation. Proteins can be visualized in a western blotting format by binding Streptavidin-AP or Streptavidin-HRP as a primary antibody and followed by colorimetric or chemiluminescent detection. With this method a similar sensitivity to the

autoradiographic detection with [^{35}S]methionine incorporation can be achieved. Therefore, radioactivity in the protein detection after cell-free reactions is eliminated.

1.2.2 Mass spectrometry (MS)

With the before mentioned methods, the protein weight and therefore a characteristic can be approximately determined or compared. Still, for a more detailed protein detection, the molecular weight needs to be determined accurately. Mass spectrometry represents an analytical technique for determining the mass to charge ratio of ions under high vacuum. The discovery and application of Matrix assisted laser desorption/ionization (MALDI) as a gentle ionization method, made the transfer of larger intact biological macromolecules like proteins in the gas phase and thus their mass spectrometric analysis possible. When using MALDI the exact mass of biological molecules can be determined and their chemical composition can be verified. If the amino acid sequence is known, it is possible to construe from the difference between the calculated and the measured mass of a protein directly to the post translational modifications. In addition, unknown protein samples can be quickly and easily identified directly after proteolytic cleavage, based on the exact masses of the released peptides using a sequence database comparison [77].

1.2.3 High-throughput screening protein analysis

1.2.3.1 Protein specific spectroscopic methods

1.2.3.1.1 Fluorescence

The family of fluorescent proteins is originating from marine organisms. These organisms from coral reefs are characteristic for displaying bright fluorescence in almost the entire visible spectrum (450 nm to 655 nm). The Green Fluorescent Protein (GFP) from *Aequorea victoria* is the most famous and most studied fluorescent protein [78]. The fluorescent nature of this protein can be used as a reporter protein for cell-free expression, because it is always detectable against the protein background.

1.2.3.1.2 Luminescence

Synthesized firefly luciferase from *Photinus pyralis* is detectable in a protein mixture using a luciferase assay substrate and a following luminescence measurement. In the Firefly Luciferase assay, the addition of luciferin, ATP and O_2 converts luciferase and its cofactor Mg^{2+} to oxyluciferin, AMP, PP_i, CO_2 and measurable luminescence emission [79]. Hence, this protein detection method can also be used in solution, right after the expression reaction.

1.2.3.2 Protein Microarray

Array systems are analytical systems which allow a large number of simultaneous measurements in one experiment. Samples are placed in an array format in a defined manner. A microarray is the miniaturized format allowing a highly parallel implementation of experiments. A protein microarray can identify and quantify a high number of proteins in a single experiment. Therefore, with a protein microarray expression studies, as well as global interaction studies and functional studies can be analyzed. The outstanding strength of the microarray technology is the high sensitivity of the measurements and on the other side the possibility to determine dozens to hundreds of relevant experimental parameters from extremely small samples [77]. The array systems are characterized by two points, on the one hand, the high degree of parallelism and on the other hand the extreme reduction of the analyte detection area. In many cases, protein microarrays are used for the analysis of antigen-antibody interactions. In an array a plurality of antigens is immobilized on a microarray surface and then detected by a labeled antibody.

2 Motivation and Research Proposal

For industrial processes in the biotechnological field, which traditionally are divided into upstream and downstream applications, each part of the process is still characterized and optimized separately. The upstream part includes genetic engineering and bacterial or mammalian strain development. The optimization of fermentation processes, involving feed strategies as well as media composition is also embedded. The downstream processing starts with the cell harvest and include product purification. Nowadays, research and development has recognized the need for combining the characterization and optimization of the two parts in order to integrate development of sections within the overall process. However, with 'established' approaches in process development often based purely on experience and sequential experimentation, integration has shown not to be feasible. The development of high-throughput methods enables faster and material-saving processes. These methods include biochemical, genetical and pharmacological tests, automated on liquid handling stations for a faster and reproducible performance. Originally, high-throughput methods were used for target screening in the pharmacological industry, but are now applicable for all kinds of research projects. One of the major challenges under this methodology of process optimization is the minimization of processes to high-throughput modifications, including a robotic implementation as well as an application improvement and optimization of the single process steps. The optimization of the complete process can only be conducted when the combination of the upstream and downstream part during the optimization is aspired. The following scale up of optimized process parameters therefore always includes upstream and downstream aspects (Figure 3). In the area of upstream processing, cell-free expression techniques recently came into the picture. They are a fast expression tool for small quantities of proteins, and therefore ideal for screening experiments. The cell-free expression systems are also complementary for diverse cell types. The most used and investigated cell type is *E. coli*, but also insect cells, wheat germ or even mammalian cells are proving promising results. For example insect cells own the ability to express heterologous proteins with post-translational modifications. Therefore, the cell-free expression performance of insect cells comprises lots of new features over *E. coli*. However, they are not as well established as *E. coli* systems. In high-throughput process optimization, data generation using a statistically relevant design of experiments is important. Well-designed approaches enable robustness of the system and description of experimental sensitivities. Furthermore, statistical data analyses will provide interaction quantification and reliable predictions for optimal substance concentration ranges as well as highlighting

influencing factors and predicting substance interactions. In the following, the characterization of two insect cell strains on behalf of their capability of being a cell-free expression system are tested and confirmed by a model based analysis (for further reading: “High-throughput characterization of an Insect cell-free expression”).

After enhancing cell-free reaction systems with high-throughput screening and implementation on robotic platforms, the resulting systems are usable for screening and improving cell-based expression. Many proteins are hard to express because they are toxic to the cell metabolism. Therefore, the combination of high-throughput parameter optimization and transfer to cell-based expression is a valid approach for folded and soluble difficult-to-express proteins. With cell-free expression, variables such as the comparison of eukaryotic or prokaryotic systems, requirements for auxiliary factors (e.g. the addition of chaperones, detergents or cofactors) or protein- and process-specific (e.g. temperature, time) can be estimated [80]. However, the full-length expression of the autoantigen U1-68/70 K has not been reported in literature so far. In the publication: “Soluble full-length expression and characterization of snRNP protein U1-68/70 K”, the cell-free expression optimization of U1-68/70 K provides an example for the adaptability of the cell-free systems in screening experiments and shows that scale up is both manageable and process parameters are transferable.

The last and most important challenge is the linkage of upstream to downstream processes. By linking cell-free expression to microarray diagnostics, two high-throughput methods were combined, eliminating the laborious and time-consuming steps of cell-based protein expression and subsequent purification (“Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera”). Cell-free expression and microarray technology are already linked for applications, such as biomarker detection for diagnosis of autoimmune disease, immunological studies, vaccine development, protein-protein interactions and toxin detection. However, the linking of these two high-throughput methods are currently accomplished by DNA printing and a subsequent protein expression on the microarray surface. Therefore, the comparability to protein microarrays is challenging. Printing DNA molecules instead of proteins is evidently different. The reproducibility of on-chip expression may limit the use of this technology in commercial diagnostics. The possibility of rapid and cost-reduced biomarker screening to improve the protein conditions for printing may be possible. An advanced question is the comparability between arrays developed on the basis of cell free expression and those manufactured by the large-scale compatible cell-based variant.

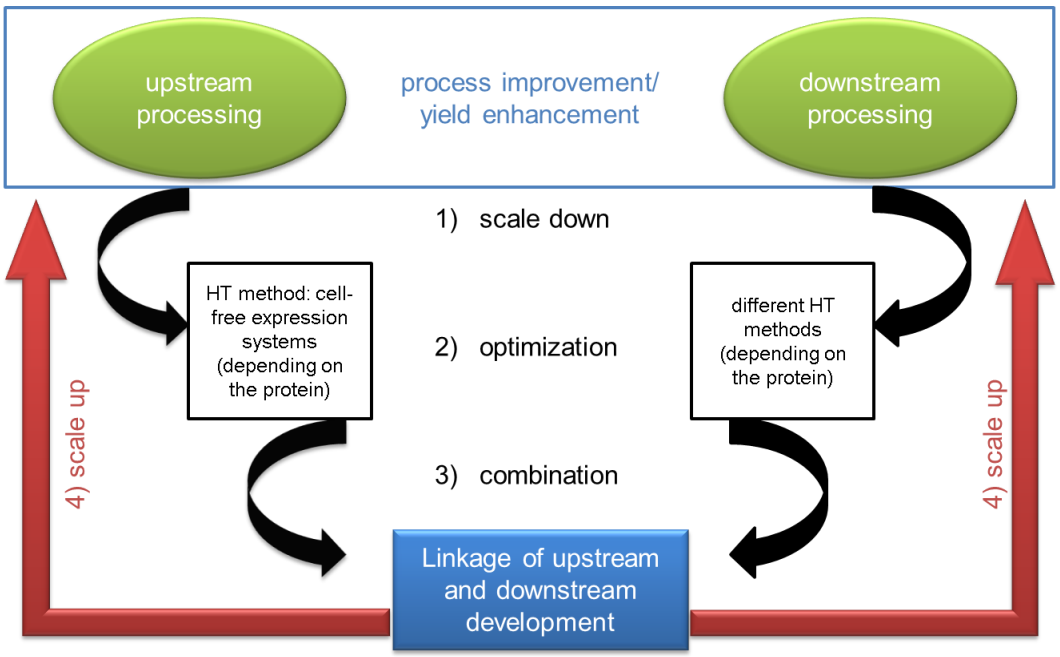


Figure 3: Concept of research proposal for this thesis.

3 Publications and Manuscripts

High-throughput characterization of an Insect cell-free expression

Carolin Richter, Fabian Bickel, Anna Osberghaus, Jürgen Hubbuch
submitted to *Engineering in Life Sciences*

This paper describes the reproducible implementation of two insect cell-free expression systems on a robotic platform. The characterization of the *in vitro* translation process included experimental planning by statistical design of experiments (DoE) and visualization of the parameter influences on the expression yield as well as the fit of the experimental data to quadratic response surface models by multivariate data analysis (MVDA). The results were compared to previous studies, which confirmed the applicability of the new method.

Soluble full-length expression and characterization of snRNP protein U1-68/70 K

Carolin Richter, Thomas Simon, Iris Asen, Gerald Brenner-Weiss, Jürgen Hubbuch
submitted to *Protein Expression and Purification*

In this study, expression parameters such as the application of the cDNA construct, the expression temperature and folding properties for the cell-based expression in *E. coli* of the autoantigen U1-68/70 K could be determined in a cell-free expression screening, followed by a successful protein production. Until now, this protein could not be expressed in its full-length form.

Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray based multiplex detection using different patient sera

Carolin Richter, Kosta Konstantinidis, Iris Asen, Richard Kneusel, Jürgen Hubbuch
accepted by *Engineering in Life Sciences* (DOI: 10.1002/elsc.201300109)

In this paper, we established a procedure for the cell-free expression and subsequent printing of different *Borrelia* antigens onto several multi-well microarray plate surfaces. The eleven immunodominant antigens of Lyme borreliosis from different *Borrelia* species proteins could be reproducibly detected on the microarray plates. This approach serves as a proof of principle for the identification of potential biomarkers using cell-free expressed proteins.

3.1 High-throughput characterization of an Insect cell-free expression

Carolin Richter[1], Fabian Bickel[2], Anna Osberghaus[1], Jürgen Hubbuch[1]

[1]Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany

Corresponding author: Juergen.Hubbuch@kit.edu
Phone +49 721 608 42557
Fax +49 721 608 46240

[2]Biberach University of Applied Science, Institute of Applied Science, Hubertus-Liebrecht-Straße 35, 88400 Biberach / Riss, Germany

Keywords

Insect cell-free expression, model based analysis, Firefly Luciferase, high-throughput screening, design of experiments

Abbreviations

adenosine triphosphate (ATP)
amino acids (AA)
coefficient of variation (CV)
creatine kinase (CK)
creatine phosphate (CP)
cytosin triphosphate (CTP)
design of experiments (DoE)
guanosine triphosphate (GTP)
high-throughput screening (HTS)
multivariate data analysis (MVDA)
magnesium acetate ($MgO(Ac)_2$)
potassium acetate (KOAc)
Prediction Sum of Squares (PRESS)
root mean square error (RMSE)
transfer ribonucleic acid (tRNA)
uracil triphosphate (UTP)

Abstract

Cell-free protein expression is a promising tool for improving protein specific expression techniques. Despite their advantages, Insect cell-free expression systems are not as well established as *Escherichia coli* cell-free systems. In most studies, characterization and optimization strategies are based on manual ‘one factor at a time’ investigations that are expensive and time consuming. In this paper, two insect cell-free expression systems (Sf9 and High Five™) were reproducibly (CV=2.9%) implemented on a robotic platform with integrated analytics. All experiments were planned by statistical design of experiments (DoE) using central composite designs and analyzed by multivariate data analysis (MVDA). Quadratic response surface models were fitted to the experimental data and model predictivity was validated successfully for both insect cell types. The characterization of the complete *in vitro* translation process included quantification and visualization of the parameter influences on the expression yield and the robustness of the systems. The results were compared to previous studies, which confirmed the applicability of the new method. In the future, yields from insect cell-free expression can be enhanced using a comprehensive system characterization based on optimally designed high-throughput screenings on robotic systems.

1 Introduction

In the growing field of proteomics there is an increasing demand for purified recombinant proteins for structural and functional studies. However, the adaptation of cell-based protein expression for high-throughput procedures is difficult and laborious [46,80]. Cell-free expression is a promising solution to this bottleneck. It offers the expression of large numbers of proteins in a short time frame and examination of protein-specific reaction conditions, for example co-translational or post-translational modifications, which support protein folding or solubilization [1].

Today there are different well-established sources of cell lysates, reaching from *E. coli* over wheat germ, yeast, insect to mammalian cells [44,81,82]. The choice of system depends on the biochemical nature of the desired protein. Intrinsic enzymes present in insect cell lysates are advantageous for many types of eukaryotic-specific post-translational modifications [47]. Additionally, extract preparation is easy and quick. Consequently, the insect cell *in vitro* expression system appears to be the “fastest growing cell-free expression platform” [3]. At the same time, however, insect cell-free expression systems are not as advanced as can be found for *E. coli* cell-free systems [49,83,84] consequently leading to a lower performance in terms of productivity than found in *in vitro* expression of *E. coli* lysates [44]. Characterization and optimization of insect cell-free expression systems is still conducted by changing ‘one-factor-at-a-time’, a laborious and time consuming procedure where no substance interactions within the cell-free reaction can be analyzed. Optimization procedures have already been conducted to the lysate composition, preparation of extracts from genetically engineered cells, the choice of the regenerating energy system and the DNA template [2]. Particularly, the insect cell-free lysate production and reaction substance composition was characterized and optimized by Ezure *et al.* [23] and Sato *et al.* [24] with one-factor-at-a-time experimentation. Sato *et al.* [85] developed the direct input of the mRNA into the insect cell-free translation reaction without any purification in order to pave the way for high-throughput screening (HTS) experiments.

HTS would allow for a thorough characterization of the reaction substance composition in an insect cell-free system using modern robotic platforms for automated experimentation combined with model-based analysis. With data based on a statistically relevant design of experimental points, the robustness of the system and the sensitivities can be described. Influencing factors can be highlighted and furthermore, statistical data analysis provides interaction quantification and reliable predictions for optimal substance concentration ranges [86].

In this study, we successfully and reproducibly implemented a well-designed HTS for insect cell-free expression on a robotic pipetting platform. For the linked insect cell-

free expression this approach included transcription and translation, and the protein analysis. The generated data was fitted to a response surface model and analyzed by MVDA, providing a remarkable system characterization of the insect cell-free translation reaction. Based on the validated model, the optimal concentration of each single substance in the insect cell-free expression and their interactions could now be predicted for the two insect cell-free lysates from Sf9 and High Five™ cells. This new approach significantly extends the results from the cited previous linear studies. In the future, this approach can be used for optimization of *in vitro* expression techniques for different cell systems.

2 Material and Methods

2.1 Materials. HEPES, DTT, adenosine triphosphate (ATP), amino acids (AA), cytosin triphosphate (CTP), EGTA, glycerine, glycylglycine, guanosine triphosphate (GTP), magnesium acetate ($MgO(Ac)_2$), potassium acetate (KOAc), uracil triphosphate (UTP) were purchased from Carl Roth GmbH + Co. KG (Karlsruhe, Germany). Creatine phosphate (CP) and spermidine were from AppliChem GmbH (Darmstadt, Germany). Creatine kinase (CK) in a specific activity of 508 U/mg (Lot. 12861621) and transfer ribonucleic acid (tRNA) from baker's yeast were obtained from Roche Diagnostics Deutschland GmbH (Mannheim, Germany). RNAse Inhibitor RNasin® (40U/µL) and Steady-Glo® Luciferase Assay System were from Promega GmbH (Mannheim, Germany). Firefly Luciferase/Photinus Lampyris (recombinant) for calibration of Steady-Glo® Luciferase Assay System was purchased from PJK GmbH (Kleinblittersdorf, Germany).

2.2 Molecular basics. For all Insect cell-free expression reactions the plasmid pF25A ICE T7 Flexi Vector (Promega GmbH, Germany - GenBank® accession no. EU754721) implementing the luc+ gene [87] was used. The vector included a T7 RNA polymerase promoter, a 5’ and 3’ UTR sequence and a poly adenosine tail.

2.3 Cell culture conditions. Spodoptera frugiperda insect cells (Sf9; IPLB-SF-21-AE) [88] and High Five™ (BTI-TN-5B1-4) (Trichoplusia ni) insect cells (LifeTechnologies Corporation, Germany) [89] were grown in HINK'S TNMFH (SAFC®Global) supplemented with 10 % FKS at 27°C in adherent cultures. For the following suspension cultures Gibco® Express Five® Serum Free medium supplemented with 4 mM L-glutamine in Ex Cell ® 420 (SAFC®Global) for Sf9 cultures were used for higher cell densities prior to extract preparation. All media components were purchased at Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany).

2.4 Preparations for Insect cell-free expression. The insect cell cultures were harvested after reaching the exponential growth phase (10^6 cells/ml), washed and resuspended to a density of $1.5x10^8$ cells/ml. Insect cell extract was prepared as previously described [23] and stored at -80°C until further use. The above-mentioned plasmid was transcribed with 0.5 U/µl T7 RNA polymerase (Promega GmbH, Germany) in transcription buffer [90] using 16 mM $MgO(Ac)_2$ as Mg^{2+} source and incubated at 37°C for 3 h. For the purpose of a high-throughput application, the mRNA was not purified. The absorbance at 260 nm and 280 nm was measured to determine the purity and existence of the mRNA in the transcription buffer before and after the reaction. The transcription mixture was directly used or stored at -80°C.

2.5 The response surface modeling approach. Response surface modeling is a useful tool for empiric modeling and predicting a response of different input variables with statistical techniques [91]. It is able to specify the relationships among the response and the input factors. Response surface modeling includes the design of an experimental measurement series of the response, the process of developing a mathematical model with the best fit, finding optimal input variable values producing a maximum response and representing direct and interactive effects of process parameters. For the DoE, the central composite design is one of the most popular designs. It is a full-factorial design combined with a central composite design star, additionally including three center points. Therefore, all combinations are systematically varied, resulting in 2^k factorial design points, several center points and 2^k axial star points. The distance between the center of the design space to a star point is α. This DoE allows for the detection of nonlinearities and interactions in the factor-response-relationship. The here described characterization experiments are based on a central composite faced design (α=1.0) with three levels for each factor, three center points and a duplicate determination of all measurements (see Figure 1). For the characterization of the reaction substance composition in an insect cell-free system all additives of the translation premix, besides mRNA, insect cell-free extract and the protease inhibitor were investigated by the experimental design. Insect cell-free extract and mRNA were not included, because both are containing variable substances, depending on cell batches. To handle this variability, the insect cell-free extracts originated from one batch. Furthermore, the same mRNA preparation was used for one experimental setup. The unpurified mRNA still includes Mg^{2+} from the transcription reaction. It is experimentally added to the translation mixture. The protease inhibitor was directly given in an appropriate concentration to the insect cell-free extract for stabilization.

A number of 566 experiments was generated by the DoE software Modde 8 (Umetrics, USA) varying the twelve translation premix substances within reasonable

preset ranges, expanding limitations from [92] (Table 1) and including repeat determination of each setup. The experimental order was randomized to exclude any regional effects. For a general first approach, a smaller number of experiments would have been possible, but in order to assure a good coverage of the design space and to obtain a robust model even in case of many outliers, the highest number of experiments possibly performable within one week was chosen. The same experimental design was applied to Sf9 and High Five™ insect cell-free extracts.

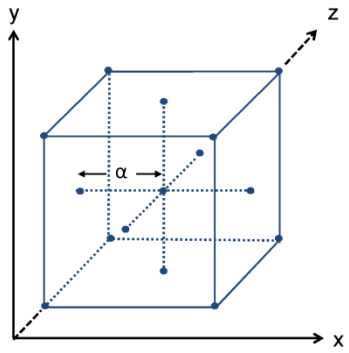


Figure 1: A typical face-centered composite design for three factors

Table 1: Translation premix components with the corresponding concentration range; lower and upper limits for DoE

translation premix component	concentration range limits	
HEPES-KOH (pH 7.9) [mM]	20	50
ATP [mM]	0.1	2
GTP [mM]	0.1	2
DTT [mM]	0.2	5
tRNA [mg/ml]	0.01	0.5
KOAc [mM]	50	150
$MgO(Ac)_2$ [mM]	1	3
EGTA [mM]	0.1	10
AA [mM]	0.01	0.1
CP [mM]	1	100
creatine kinase 508 U/mg (in 0.25 M pH 7.4 Glycylglycine)	10 mg/µl (0.005 U/µl)	404 mg/µl (0.20 U/µl)
RNase Inhibitor [U/µl]	0.3	0.65

2.6 Implementation of the Insect cell-free reaction on a robotic platform. For a study of 566 experiments as investigated here, pipetting with a robotic platform is preferred and therefore the platform ‘Freedom Evo’ (Tecan, Crailsheim, Germany) was used. The robotic workstation was equipped with a carrier cooling rack to control the temperature during pipetting, a robotic manipulator arm to move the microtiter plates, two incubators with integrated shaking function, a liquid handling arm (LiHa) and an Infinite microwell plate reader. Two dilutor volumes (1000 µL and 250 µL capacity), each connected to four tips of the before mentioned LiHA were used. More information about robotic systems can be accessed in former publications of our reseach group [93–95]. Since viscosity and ionic strength influence the pipetting process, liquid classes for every stock solution of the twelve translation premix substances were established and calibrated according to procedures described in [96]. Up to three different stock solutions per component were used to adjust the final concentrations. Using a dilutor with a capacity of 250 µl, the minimal pipetted volume is restricted to 1 µl when aiming at high accuracy. The tips of the liquid handling arm were washed and sterilized with 70% ethanol after finishing a pipetting step with a specific substance. Pipetting high sample numbers on a robotic platform and intensive intermediate cleaning of the pipetting tips can last 12-24 hours, therefore all substances were cooled down to approximately 9°C in the carrier cooling rack (Tecan, Crailsheim, Germany).

Prior to the implementation of insect cell-free reaction on the robotic system (Figure 2) the stability of the translation premix, the transcribed mRNA and the insect cell lysate and its influence on the following expression performance were determined separately. All components of the translation premix were pipetted together and incubated at 9°C. After different time periods up to 18 h, samples were taken and the mRNA and insect cell-free extract were added. The cell-free expressions were incubated at 27°C for 3h and their protein yields were compared. The stability of the mRNA and insect cell-extract at 9°C at different times up to 40 minutes were tested in the same way. Furthermore, the influences of different mixing conditions on the expression yield were checked. The mixing of the insect cell-extract after thawing and the mixing step immediately before the cell-free expression including the translation premix, the mRNA and the insect cell extract were investigated. It was distinguished between no mixing, gentle mixing or excess mixing using a pipette and mixing on a vortex mixer. After the mixing step, the insect cell-free expressions were incubated at 27°C for 3h and the protein yields of firefly luciferase were compared.

For all experiments on the robotic system 96-Well Half Area White Flat Bottom Polystyrene NBS™ Microplates (Corning, NY, USA) were used. The overall volume of the insect cell-free translation was 50 µl, consisting of 40% translation premix, 10%

mRNA and 50% insect cell-free extract. The translation premixes consisted of twelve premix substances (Table 1). The insect cell-free extract was thawed immediately prior to the translation reaction and a concentration of 0.1% [v/v] protease inhibitor was added. After pipetting the different translation premix compositions, mRNA and insect cell extract were added and the insect cell-free reaction was incubated at 27°C for 3h.

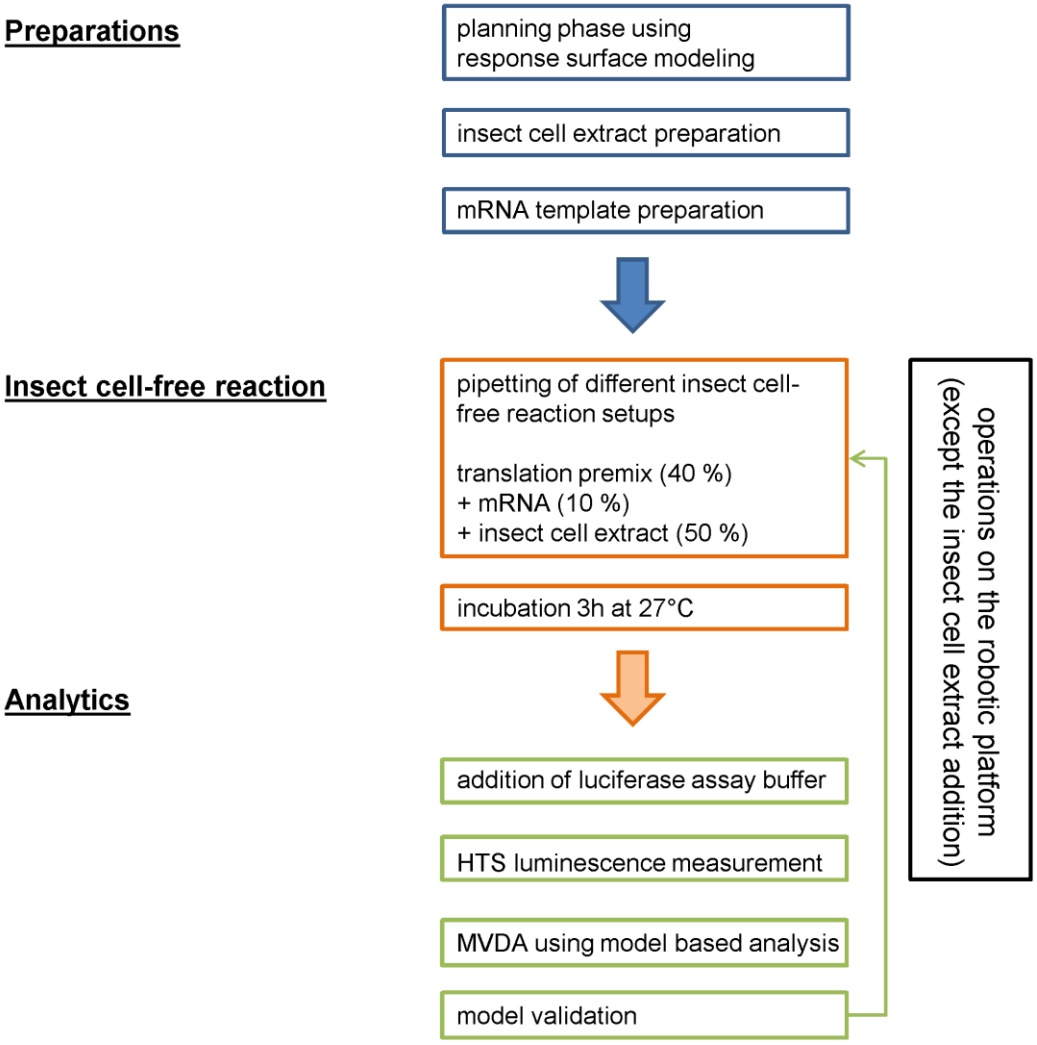


Figure 2: Flow diagram of the insect cell-free expression characterization combined with a model-based approach.

2.7 Luminescence assay. Synthesized firefly luciferase from *Photinus pyralis* (EC 1.13.12.7) was detected in HTS format using the Steady-Glo® Luciferase Assay System (Promega GmbH, Germany) according to manufacturer's instructions. In the Firefly Luciferase assay, the addition of luciferin, ATP and O_2 converts luciferase and its cofactor Mg^{2+} to oxyluciferin, AMP, PP_i, CO_2 and measurable luminescence emission [79]. Since ATP and Mg^{2+} as well as the chelating agent EGTA are existent in different concentrations in the translation premix, their influence on the assay was determined beforehand. For the firefly luciferase yield calculations, the calibration curves also included the three different EGTA and Mg^{2+} concentrations, resulting in 9 different calibration curves. Luminescence results were measured and data translation performed before use in model-based data analysis.

2.8 Multivariate data analysis and model validation. The luminescence assay results for Sf9 and High Five™ were both separately fitted to an empiric quadratic model, created with the 'Model-Based Calibration Toolbox' in Matlab (R2011a, Mathworks, Germany). Equation 1 shows the regression fit of an n-variate quadratic function of m responses z_1, ..., z_m with x_1, x_2, ..., x_n as the n selected factors for expression description and z_i the response value to a specific factor setting x_{1i},x_{2i}, ..., x_{ni} for all $1 \leq i \leq m$.

$$z_i=a_1+b_1*x_{1i}+b_2*x_{2i}+\ldots+b_n*x_{ni}+c_1*x_{21i}+c_2*x_{22i}+\ldots+c_n*x_{2ni}+d_{1,2}x_{1i}x_{2i}+d_{1,3}x_{1i}x_{3i}+\ldots+d_{n-1,n}x_{n-1,i}x_{ni}$$

Equation 1: Regression fit of n-variate quadratic function

The parameter a_1 is an added constant like an intercept term in linear regression. The values of the parameters b_k for $1 \leq k \leq n$ display the magnitude of linear influence of the factors x_k. The values of the parameters c_k with $1 \leq k \leq n$ quantify quadratic influences of the factors x_k and the mixed effects/interaction terms of two-components are quantified by the parameters d_{12} to $d_{n-1,n}$. For a better fit of the model, outliers were eliminated based on the linearity of the N-probability plot. For Sf9 only two outliers were found, whereas for HighFive™ six data points were eliminated. A Box Cox transformation of the data was applied. For model validation 15 random experiments were predicted within the design space, including one replicate each, and experimental data was compared with the RSM results.

3 Results

The focus of this study was the implementation of the insect cell-free expression on a robotic platform and the characterization of the reaction substance composition in an insect cell-free system using a model-based analysis.

3.1 Implementation of insect cell-free reaction on the robotic platform

The insect cell-free reaction consists of mRNA, insect cell-free extract and the translation premix. Prior to the robotic implementation, stability experiments regarding the insect cell-free extract, mRNA, translation premix and the mixing conditions were performed. The expression performance of the translation premix did not change after incubating the pipetted substances at 9°C for 18 hours. This is a pre-requisite for pipetting 566 wells, each including different premix substance concentrations. Therefore the pipetting order started with the translation premix (Figure 2). The stability tests for the mRNA in the transcription buffer showed 100% translational activity after the incubation at 9°C for 40 min (Figure 3 A). This time range is enough for pipetting mRNA into each of the 566 wells before the incubation. By comparing the absorption values of the transcription mixture before and after the transcription process, an accumulation of mRNA is measurable with absorption spectroscopy (Figure 3 B). Additionally, the quotient of A260/A280 between 1.8 and 2.0 is a measure for mRNA purity; the value of 1.87 for the finished transcription reaction confirmed insignificant protein contamination. The crude insect cell extract lost 40% of its expression activity after 40 minutes incubation at 9°C. Furthermore, by excessive mixing or using a vortex mixer to mix the extract after thawing, its expression activity declined to less than 40% (Figure 3 C). Therefore gentle handling, gentle mixing and rapid processing after thawing is critical to maintain 100% expression activity. Hence, the insect cell extract was added to the cell-free reaction by multichannel dispension right before the incubation start. Since this is a time dependent step and the robotic configuration did not allow a simultaneous pipetting of the desired volume range with eight tips, multichannel dispersion was performed by hand. The last mixing step before the incubation is also very important, especially when keeping the fragile insect cell-free extract in mind. Surprisingly, it turned out, that the insect cell-free translation activity is not influenced by the mixing method, and the expression reproducibility increased with an excessive mixing step by a vortex mixer (Figure 3 D).

For the determination of an overall reproducibility using the robotic workstation, 18 identical insect cell-free reactions were set up. The pipetting of the translation premix and the mRNA on the liquid handling station and the cell extract addition by hand, using a multichannel pipet, resulted in a coefficient of variation (CV) of 2.9% showing the reproducibility of the implementation.

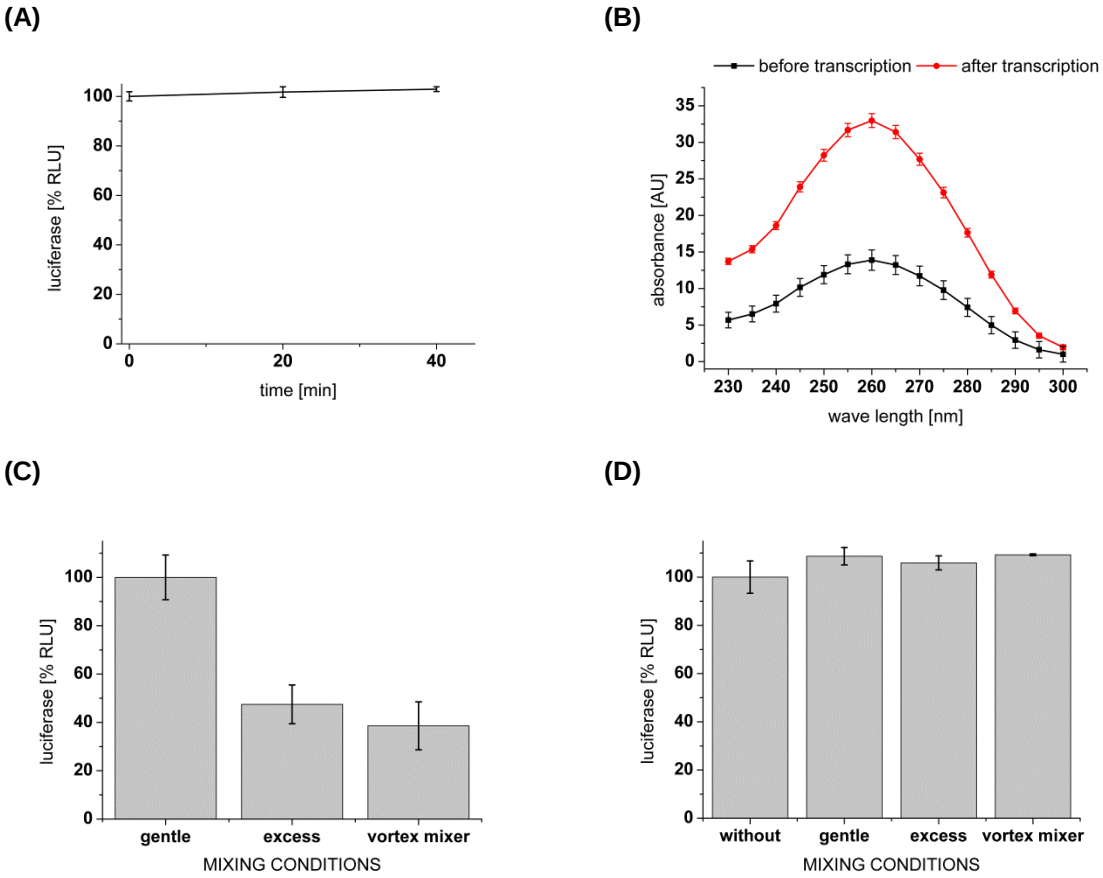


Figure 3:

- **a) Stability tests of mRNA prior to the implementation of insect cell-free expression on the robotic system. The influence of mRNA incubation at 9°C for three different time ranges (0 min, 20 min and 40 min) on the protein expression yield was tested. All samples were translated as described in the methods section.**
- **b) Determination of spectroscopic absorbance [AU] of the mRNA before and after the transcription process to show mRNA accumulation.**
- **c) Influence of cell extract mixing conditions (gentle, excessively, with a vortex mixer) on the expression result before the addition of mRNA and translation premix.**
- **d) Influence of mixing conditions (without, gentle, excessively, with a vortex mixer) on the expression result after the addition of mRNA and insect cell extract to the translation premix.**

3.2 Model-based characterization of insect cell-free reaction

After the reproducible implementation of the insect cell-free expression on the robotic platform, all insect cell-free expression experiments were conducted for both insect cell types, Sf9 and High Five™. The measured luminescence values were calibrated and the set of 566 experiments was analyzed (Figure 2).

A quadratic model function with interaction terms was fitted to the data (see section 2.8). The coefficients of determination (R^2), describing the ratio of the model-explained variance in data compared to the unexplained variance, hence representing a quality measure for the model between 0 and 1, were 0.89 for Sf9 and 0.83 for High Five™,

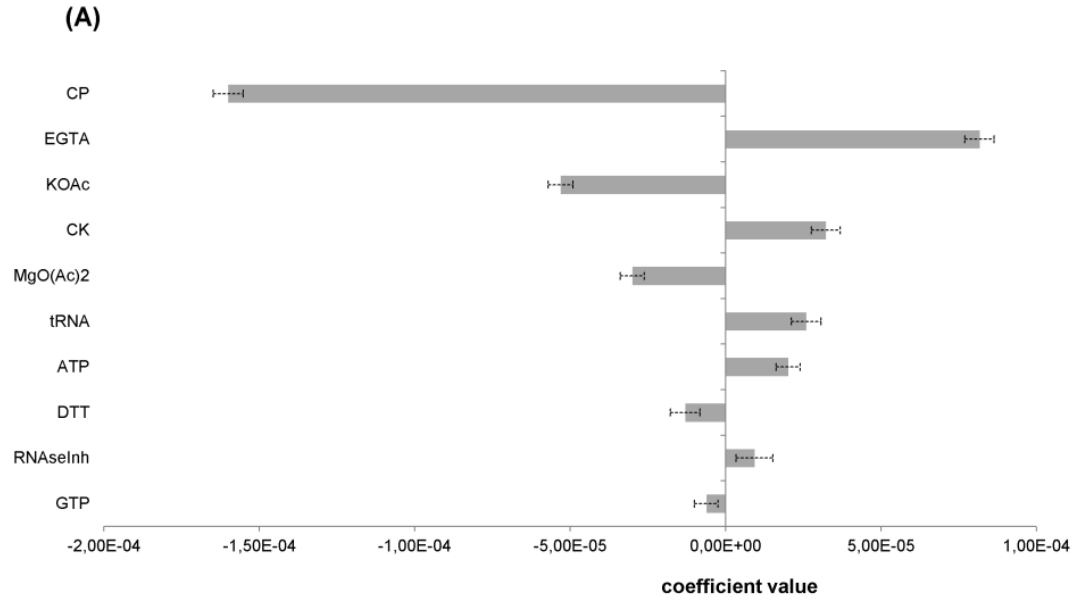
respectively. Also PRESS (Prediction Sum of Squares) was calculated for both regression models by crossvalidation, where observations are removed and the left out values are predicted based on a refitted model. The values of 0.87 (Sf9) and 0.79 (High Five™) indicate that the model is predictive and not over-parameterized. All characteristic values for model building and validation, including the root mean square error (RMSE), the root of the sum of squared distances between real measurements and the model-based predicted values are also given in Table 2.

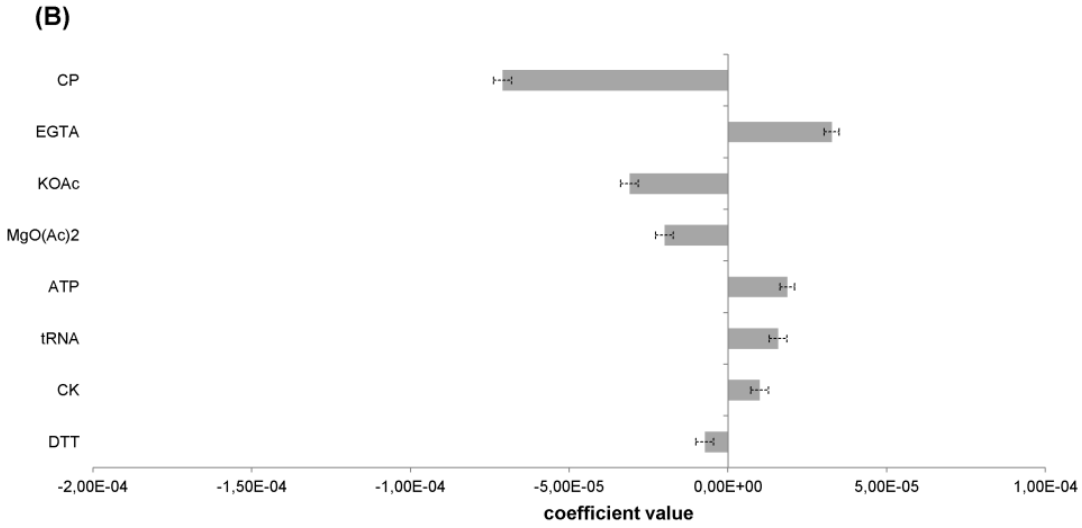
Table 2: Summary of the model-based analysis:

Summary Table	Sf9	High Five™
R^2	0.892	0.828
PRESS R^2 (Q^2)	0.871	0.787
RMSE (root mean square error)	8.753×10^{-5}	5.09×10^{-5}
Validation RMSE	2.288×10^{-4}	8.639×10^{-4}

The linear influences of the different translation premix components present in the model equation (Figure 4 A and B) and the twenty most significant influencing interactions on the protein yield of Sf9 and High Five™ (Figure 4 C and D) are shown as scaled and centered coefficient plots. The bar size represents the influence of a substance on the expression yield, and can be positively or negatively correlated. The 95% confidence intervals were calculated and a t-test with $\alpha = 5\%$ was conducted. With this test, the significance of the coefficients can be determined. The coefficients HEPES-KOH and AA showed no linear influence in both cell systems, whereas the linear effects of RNAse Inhibitor and GTP were not significant in the High Five™ system. Therefore, these terms were not included into the respective model equation. For both cell systems, positive linear effects on the expression performance were determined for EGTA, CK, tRNA and ATP whereas negative effects on the yield were detected for CP, $MgO(Ac)_2$, KOAc and DTT. The two substances additionally affecting the protein yield in the Sf9 system linearly are acting differently; GTP influences the yield negatively whereas an increase of RNAse Inhibitor increases the yield. The coefficient plots of the linear effects are very similar for Sf9 and High Five™, displaying identical effects for all substances; only the strengths of the effects are slightly different (compare Figure 4 A and B). In both systems, CP and EGTA are showing the highest influence on the expression yield of all translation premix substances. Regarding the twenty most significant influencing interaction effects on the expression yield, CP shows the highest quadratic effect on both cell systems (Figure 4 C and D). On the Sf9 system, the interaction of CP and EGTA as well as CP

and KOAc is remarkably strong. The yield of protein from High Five™ is mostly affected by the quadratic effects of EGTA and the influence of the interaction between CP and EGTA. Furthermore, both systems are significantly affected by the interactions of CP and EGTA with other translation premix components.





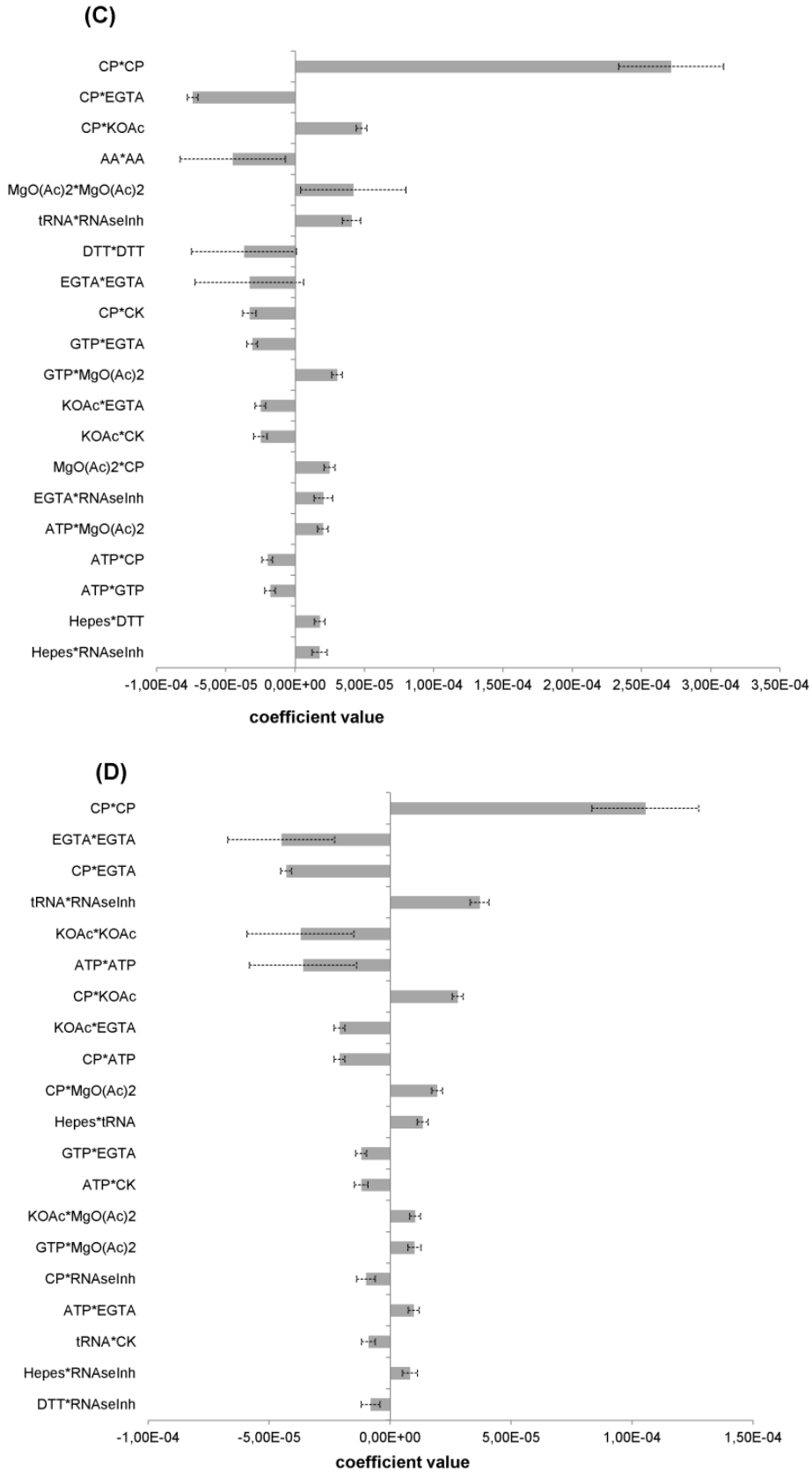


Figure 4: Scaled and centered coefficient plots with confidence intervals (95%) for Sf9 (A: linear influences, C: nonlinear influences) and High Five™ (B: linear influences, D: nonlinear influences)

Figures 5 A (Sf9) and 5 B (High Five™) display the sensitivity of the expression yields to the relevant translation premix substances including confidence intervals of 95%. The general curvature of the sensitivity plots corresponds to the magnitude of effect already explained in Figure 4, but they offer some more information on optimal parameter settings. In detail, the highest expression yield with the Sf9 cell-free system can be reached using a DTT concentration of 2.5 mM and a $MgO(Ac)_2$ concentration of 1.5 mM. Whereas in the High Five™ cell-free expression an ATP concentration of 1 mM, a KOAc concentration of 50 mM and an EGTA concentration of 6 mM provide the highest protein yield. For KOAc in the Sf9 system the lowest applied concentration showed a maximum in expression yield. Therefore, increasing the KOAc concentration would decrease the yield. Similarly, an increasing $MgO(Ac)_2$ or DTT concentration influences the expression result of the High Five™ system negatively (Figure 5 B). Therefore, the respective minimum concentrations were optimal. CK, tRNA, EGTA and ATP concentrations in the SF9 system as well as CK and tRNA concentrations in the High Five™ system show a maximum, therefore the highest expression yield is reachable with these substances highest concentrations applied in this study (Figure 5 A and B). As an exception to the before mentioned results, the cell-free expression system is very sensitive regarding the component CP. Its influence on the expression yield shows an identical minimal turnover (~70 mM CP) in both here applied insect cell extracts. Therefore, the model-based analysis suggests that a lower or a higher substance concentration of CP would both lead to higher protein expression yields. Overall, the translation premix substance concentration influences on the expression yield are very similar for Sf9 and High Five™. Compared to previous studies, the main differences between the results of the here applied model-based approaches and Ezure *et al.* [23] is the detected need for a higher EGTA concentration (~7 mM compared to 0.25 mM [23] for the HighFive™ extract) and a lower Mg^{2+} concentration (1 mM compared to 2 mM for the HighFive™ extract) for optimal yield. Furthermore, the optimum of CP was described by Ezure *et al.* [23] at a concentration of 20 mM, whereas the model-based analysis suggests lower or higher concentrations for a higher expression yield. Thus, the presented approach confirms most of the linear effects determined by Ezure *et al.* [23] as well as the high sensitivity of the system to CP, but the presented methodology provides additional information on factor interactions and is based on highly parallelized and automated HTS experiments reducing the consumption of resources and the experimental effort.

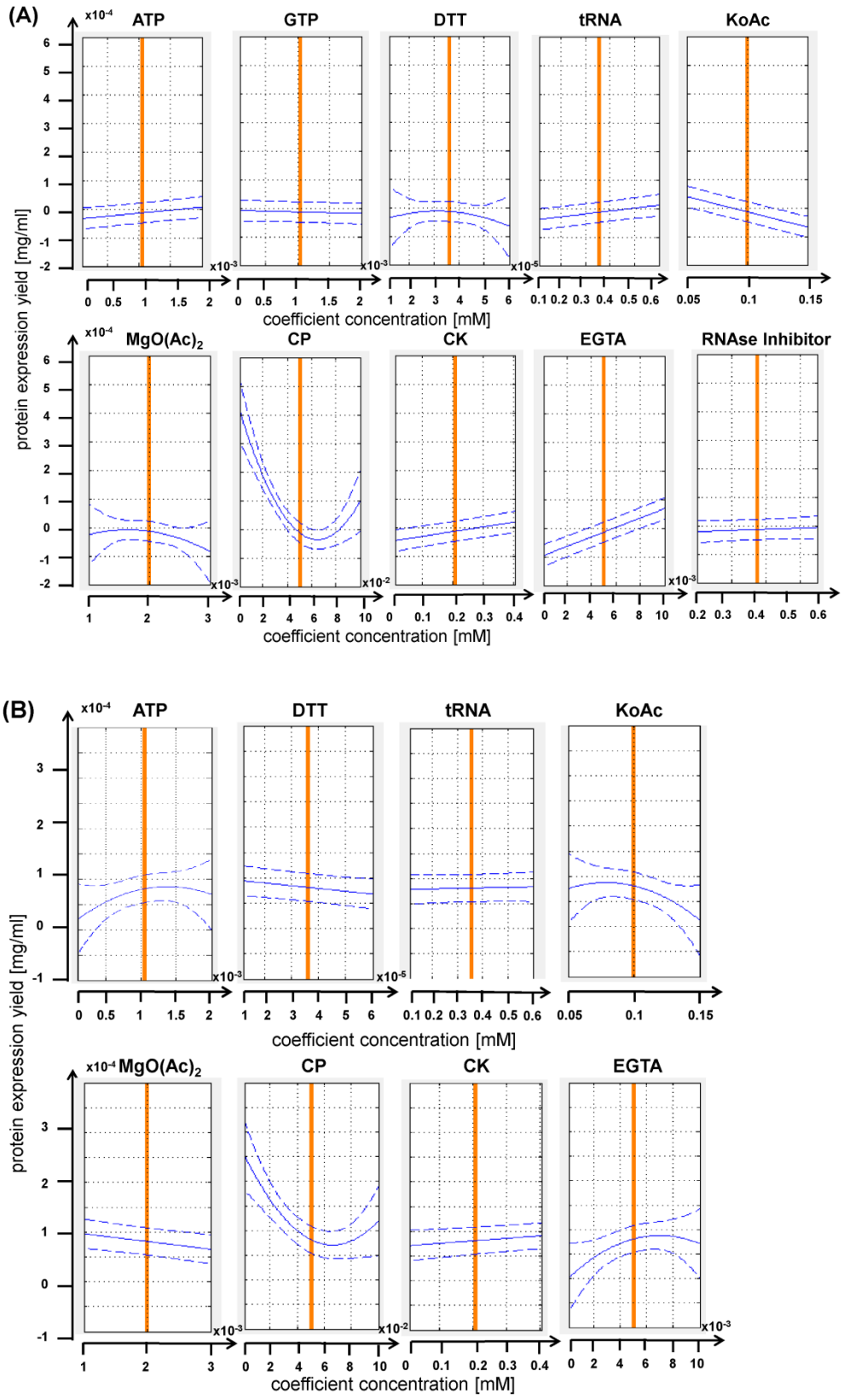


Figure 5:

A) Concentration effects on the Sf9 insect cell-free protein synthesis of the translation premix parameters (ATP, GTP, DTT, tRNA, KOAc, $MgO(Ac)_2$, CP, CK, EGTA, and RNAse Inhibitor). The impact of every substance concentration on the Firefly Luciferase yield [mg/ml] of the model-based results is displayed with the blue line. The dotted blue lines are confidence intervals (95%).

B) Concentration effects on the High Five™ insect cell-free protein synthesis of the translation premix parameters (ATP, DTT, tRNA, KOAc, $MgO(Ac)_2$, CP, CK and EGTA). The impact of every substance concentration on the Firefly Luciferase yield [mg/ml] of the model-based results is displayed with the blue line. The dotted blue lines are confidence intervals (95%).

4 Discussion

In this study, insect cell-free expression on a robotic platform was successfully implemented. This included stability tests of the translation premix, the mRNA and the insect cell-free extract prior to the expression process. Additionally, mixing conditions for the insect cell-free extract and the translation reaction immediately prior to incubation were tested. These results indicated that the insect cell-free extract must be handled carefully and allowed for the development of a stable and reproducible HTS process. Consequently, a model-based characterization of the effects of reaction substance composition to the yield of an insect cell-free system was conducted. A response surface model was established based on DoE-planned experimental data and was used for the characterization of the effects of all twelve translation premix substances on the expression yield. A comparison between the model-based results to a manual one-factor-at-a-time insect cell-free expression characterization by Ezure *et al.* [23] showed that in general both approaches lead to similar results on the single substance effects. Whereas Ezure *et al.* [23] used purified mRNA templates, in this study the transcription mixture was directly pipetted into the translation reaction. Through this step a small amount of Mg^{2+} is additionally added to the translation premix. This is probably the reason for the higher EGTA and lower Mg^{2+} concentration necessary for optimal yield in this study. Divalent Mg^{2+} ions are essential for many biological reactions and one of the most influential factors in cell-free reaction mixture [97]. Their importance and their interaction with EGTA was shown and characterized in the presented model-based analysis.

It has been previously shown, that extremely low CP concentrations reduce the protein expression yield in cell-free expression systems dramatically and that significantly higher concentrations inhibit the complete reaction [98], due to the accumulation of inorganic phosphate [74]. Ezure *et al.* [23] suggested in their publication an optimum of 20 mM CP which is quite opposite to the results of the model-based analysis investigated here. A substantiated explanation of this difference is however currently not at hand. However, both approaches show, that the insect cell-free expression system responds very sensitively to a CP concentration change (Figure 4 and 5).

In vitro protein expression translation premixes consist of three main substance parts, the energy regenerating system, different metal ions and essential substances for the expression process. With the model-based analysis in this study the influences of the

different premix components and their concentrations on the luciferase expression yield of two different cell extracts (Sf9 and High Five™) were observed and characterized. Thus, an improvement of the methodology of investigating cell-free expression by visualizing component interactions and influences on the expression yield was provided. Furthermore, this new approach yields comparable results to previous characterizations of the insect cell-free reactions [23,98] but outperforms them by the efficient transfer to high-throughput experimentation and the possibility to determine interaction effects of premix substances. All in all, we have established a high-throughput tool for further investigations and optimization experiments of cell-free extracts in different cell types, regarding reaction characterization and the corresponding enhancement of protein yield.

Acknowledgements:

We gratefully acknowledge the financial support by the Federal Ministry of Education and Research (BMBF) on the grant number 0315334A. Thanks to Dr. Heike Schwarz and the DIARECT AG for the insect cell supply.

3.2 Soluble full-length expression and characterization of snRNP protein U1-68/70 K

Carolin Richter[a and b)], Thomas Simon[b)], Iris Asen[b)], Gerald Brenner-Weiss[c)], Jürgen Hubbuch[a)]

[a)]Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany

Corresponding author: Juergen.Hubbuch@kit.edu
Phone +49 721 608 42557
Fax +49 721 608 46240

[b)]DIARECT AG, Freiburg, Germany
Carolin.Richter@kit.edu
Thomas.Simon@diarect.com
Iris.Asen@diarect.com
[c)]Karlsruhe Institute of Technology, Campus Nord, Institute of Functional Interfaces, Analytical Biochemistry, Eggenstein-Leopoldshafen, Germany
Gerald.Brenner-Weiss@kit.edu

Keywords

cell-free expression; autoimmunity; snRNP complex; chaperones

Abstract

The autoantigen U1-68/70 K is the dominant diagnostic marker in Mixed Connective Tissue Disease (MCTD) that until recently could not be expressed in its full-length form [Northemann *et al*., 1995]. Using cell-free expression screening, we successfully produced the snRNP protein U1-68/70 K in a soluble full-length form in *Escherichia coli* cells. The protein length and identity was determined by Western Blot and MS/MS analysis. Additionally, its reactivity in the autoimmune diagnostic was confirmed. Establishment of a cell-free expression system for this protein was important for further elucidation of protein expression properties such as the cDNA construct, expression temperature and folding properties; these parameters can now be determined in a fast and resource-conserving manner.

Introduction

The U1-68/70 K protein, a component of the nuclear spliceosomal U1-snRNP particle [99], is a major autoantigen in autoimmune diseases such as Mixed Connective Tissue Disease (MCTD) or Systemic Lupus Erythematodes (SLE; [100,101]). Biotechnological production of U1-68/70 K as a diagnostic autoantigen for detection of autoantibodies in patient sera has so far been difficult, most likely due to the sequence and structural peculiarities of the protein.

U1-68/70 K (total length 437 amino acids) has a long repetitive sequence between amino acids 231 and 393 with a very large proportion of basic (41 % arginine-residues) and acidic (30 %) amino acids [102,103]. Crystal structure analysis of the U1-snRNP particle [104] reveals that the first 60 amino acids of U1-68/70 K wrap around the common heptameric Sm snRNP-core in an unusual extended conformation devoid of regular secondary structure. A helical section (amino acids 61 - 89) and a RNA binding RRM-domain (amino acids 100 - 180) contact the U1-RNA component of the U1-snRNP. The crystal structure analysis [104] however does not include the C-terminal half of U1-68/70 K, which is thought to be unstructured and to provide binding sites for numerous constitutive and alternative splicing factors [105,106] and possibly for RNA in protamine-like fashion [102]; these activities may be regulated by serine phosphorylation [107].

Screening of a large panel of MCTD patient sera identified four major continuous domains within the human U1-68/70 K as autoantibody targets, referred to as regions A', B', C' and [108–111]. Recombinant *E. coli*-based production of U1-68/70 K has been possible only for fragments with the antigenic epitopes, but not for the full-length U1-68/70 K protein [112,113]. Northemann *et al*. [6] detected an inhibitory element within the full-length sequence of U1-68/70 K (sequence X) and proposed that this element interferes with translation. Expression of a protein containing the inhibitory sequence X could inhibit trans-actively the synthesis of other *E. coli* proteins indicating that full-length expression of U1-68/70 K is impossible [6]. Experiments with various deletions of the inhibitory sequence have shown that the number of deletions correlates with the expression level of the truncated U1-68/70 K protein [TS, unpublished data]. Interestingly, the inhibitory element corresponds to part of the charged arginine-rich unstructured region of U1-68/70 K.

To shed some light into this, several expression strategies have been evaluated so far [6]. Truncated forms of U1-68/70 K have been produced in a cell-free wheat germ system for protein-protein interaction studies [114]. In recent years, *in vitro* translation

has become an important tool for rapid and cost reduced screening of different protein expression conditions. Combining *in vitro* expression with high-throughput parameter optimization [2] followed by transfer to cell-based expression is a valid approach for folded and soluble difficult-to-express proteins [115–117]. Variables such as the comparison of eukaryotic or prokaryotic systems, requirements of auxiliary factors (e.g. the addition of chaperones, detergents or cofactors) or protein- and process-specific ones (e.g. temperature, time) can be estimated [118]. However, the full-length expression of the U1-68/70 K autoantigen has not been reported in literature so far.
In this study, we evaluated the potential of an *E. coli* cell-free translation system for producing soluble full-length human U1-68/70 K protein. Furthermore it was elucidated if data from cell-free expression provide useful informations regarding optimizing cell-based expression strategies. Several parameters, which were determined via the *E. coli in vitro* expression system, could be successfully transferred to a cell-based approach. The resulting expression strategy allows now for the first time expression of a soluble, full-length and immunologically active U1-68/70 K autoantigen in *E. coli* cells. This result enables further characterization of the human U1-68/70 K full-length protein in either structural or functional studies.

Material and Methods

DNA template generation: cloning and codon optimization. Work was carried out with the alternatively spliced shorter U1-68 K isoform of human U1-68/70 K (UniProt ID: **P08621** – isoform 2). Full-length as well as the truncated (lacking the 66 AA inhibitory sequence X – Figure 1) cDNA constructs were cloned into different vector systems:

- pET24d (Merck Millipore, Germany) for *E. coli* cell-free and cell-based expression.
- pCDF-Duet-1 (Merck Millipore, Germany) for co-expression with pET24d-based chaperone constructs in *E. coli* cells.

All vectors included C-terminal hexahistidine tags, whereas expression vectors of the original human cDNA sequence additionally included N-terminal hexahistidine tags (see supporting information Figure A1). All constructs were verified by resequencing (Solvias AG, Switzerland). In addition to constructs with the original human cDNA sequence, full-length and truncated U1-68/70 K cDNAs were codon-optimized for *E. coli* expression. Web-based bioinformatic tools were: codon usage (http://www.entelechon.com); GC content (http://www.bioinformatics.org); RNA

secondary structure (www.genebee.msu.su). Gene synthesis was done by Entelechon GmbH (Germany).

Coding regions for the chaperones dnaK, dnaJ and grpE were PCR-amplified from *E. coli* BL21 and cloned into pET24d. For chaperone co-expression, ribosome binding site/chaperone cassettes were combined in a single pET24d construct, with expression of a polycistronic mRNA driven from a single T7 promoter.

```
                                          Inhibitory sequence X [6]
                          250        260        270        280        290
U1-68/70 K           KERERRRSRS RDRRRRSRSR DKEERRRSRE RSKDKDRDRK RRSSRSRERA
U1-68/70 K_truncated KERERRRSRS RDRRRRSRSR ---------- ---------- ----------
                    ********** **********
                          300        310        320        330        340
U1-68/70 K           RRERERKEEL RGGGGDMAEP SEAGDAPPDD GPPGELGPDG PDGPEEKGRD
U1-68/70 K_truncated ---------- ---------- ---------- ----- GPDG PDGPEEKGRD
                                                       **** **********
```

Figure 1: Amino acid sequence comparison of the full-length and truncated U1-68/70 K at the inhibitory sequence X region.

Preparation of bacterial cell-free extract. For cell-free expression the RNaseE-mutant *E. coli* BL21 Star™ (DE3) strain (Life Technologies GmbH, Germany) was transformed with pRARE2 (isolated from *E. coli* Rosetta2, Merck Millipore, Germany). Bacterial cell-free extract was prepared according to Kim *et al.* [54] using the simplified procedure (S12) including slight modifications.

***In vitro* expression method.** *E. coli* cell-free expression was carried out either in 50 µl (analytic) or 500 µl (preparative) reaction volumes. The cell-free reaction mixture (Kim *et al.* [73] with slight modifications) was adapted to the S12 extract. To the cell-free reactions 6 % of DnaK supplement (5 Prime, Germany) was added. Expression temperature was set to a value between 15 – 37°C and reactions were incubated overnight (~12 – 14 h) in a thermo mixer at 300 rpm. Negative controls excluded plasmid DNA and were performed for all cell-free expression methods; control background was analyzed in parallel to the product-containing reactions.

***In vivo* expression method.** BL21Star™(DE3) bacteria transformed with the respective cDNA expression constructs were grown as overnight pre-cultures in MDG media at 37°C. Expression cultures were grown with PepYMD-505 (Studier's ZYM-505 medium with added aspartic acid and NZ-Amine replaced by peptone; [119]). At a bacterial density of OD_{600nm} = 0.6, cultures were induced by addition of 1 mM IPTG.

For protein expression, bacterial cultures were incubated overnight at 25°C at 300 rpm in an incubation shaker (Infors, Switzerland).

Protein analysis. Overnight *E. coli* cell-free reactions were centrifuged (16 000 x g; 5 min), the pellet was washed once with phosphate buffered saline (PBS) (1.5 M NaCl, 37 mM NaH_2PO_4 and 163 mM Na_2HPO_4) and solubilized in 1 % sodium dodecyl sulfate (SDS). *E. coli* cell lysates from *E. coli* cell-based expression cultures were prepared by lysozyme treatment, addition of Triton X-100 to 1 % final concentration and freeze-thaw steps. Lysates were centrifuged (16 000 x g; 5 min) for fractionation of soluble vs. insoluble components. Identical volumes of reducing SDS sample buffer were added to cell-free and cell-based samples. All samples were denatured for 5 min at 95°C and analyzed by SDS-PAGE using 6 % - 20 % gradient gels. SDS gels were either stained with Coomassie Brilliant Blue R250 or used for Western Blot analysis by transferring the proteins onto PVDF membranes (Pall GmbH, Germany).
For sensitive product detection in cell-free reactions, newly-synthesized proteins were fluorescently labeled by addition of 1 µl FluoroTect^{TM} GreenLys label (Promega GmbH, Germany) to a 50 µl *in vitro* reaction. Labeled cell-free reactions were diluted 1:4 in reducing SDS sample buffer, denatured for 2 minutes at 70°C and analyzed by SDS-PAGE. SDS gels were scanned for fluorescence detection of labeled proteins in an Ettan DIGE Imager (GE Healthcare, Germany) using the Cy2 channel.
For immunological detection of His-tagged proteins on Western Blots, anti-pentahistidine antibody (Qiagen GmbH, Germany) followed by anti-mouse IgG antibody-alkaline phosphatase conjugate (Jackson ImmunoResearch, USA) was used. The autoantigenic immunological activity of U1-68/70 K products was verified on Western Blot with an autoantibody-positive patient serum and a secondary goat anti-human IgG antibody-AP (Jackson ImmunoResearch, USA). Primary antibodies were diluted 1:1000, secondary antibodies 1:5000 in 1 % casein / TBS (10 mM Tris, 150 mM NaCl, pH 7.5). The colorimetric detection of AP conjugates was accomplished with BCIP/NBT purple one-component AP membrane substrate solution (Surmodics - BioFX, USA).

Protein purification for MS/MS analysis. U1-68/70 K was purified with Ni-Sepharose High Performance spin columns according to the manufacturer's protocol (GE Healthcare, Germany). Protein samples were adjusted to 6 M guanidine-HCl, 20 mM HEPES pH 7.9 and bound to the column overnight at 4°C. After washing

steps with 6 M urea, 20 mM HEPES pH 7.9, the columns were eluted with 100 mM, 250 mM and 500 mM imidazol in 6 M urea, 20 mM HEPES pH 7.9, respectively.

Protein analysis with matrix-assisted laser desorption/ionization-time-of-flight (MS/MS). SDS-PAGE separation was done as described above. Coomassie stained U1-68/70 K protein gel bands were excised and destained by several wash cycles with 50 % acetonitrile in 25 mM ammonium bicarbonate and 25 mM ammonium bicarbonate with 10 minutes in each solution. Gel pieces were dehydrated by washing with 100 % acetonitrile. For trypsin digestion, the dried gel pieces were soaked in 20 µl trypsin (Promega GmbH, Germany) solution (25 ng/µl in 10 mM ammonium bicarbonate), covered with 50 µl ammonium bicarbonate (25 mM) and incubated overnight at 37°C on a thermo mixer. For MS/MS analysis 0.5 µl of the respective peptide sample was mixed with 0.5 µl MALDI-matrix (0.1 % trifluoroacetic acid, 50 % acetonitrile, and 10 mg/ml 4-OH-cinnamic acid) and spotted on a stainless steel MALDI target. Analyses were performed using a MALDI-time-of-flight (TOF)-MS/MS (4800 MALDI-ToF/ToF mass spectrometer; Data Explorer Software 4.0, Applied Biosystems, Framingham, MA). Generated peak lists were calibrated against a peptide standard mixture (des-Arg-Bradykinin m/z: 904.4681; Angiotensin I m/z: 1,296.6853; Glu-Fibrinopeptide B m/z: 1,570.6774; adrenocorticotropic hormone clip 1-17 m/z: 2,093.0867; adrenocorticotropic hormone clip 18–39 m/z: 2,465.1989). Significant peptides were selected for further MS/MS runs to determine the amino acid sequences.

Results

Initial experiments for cell-free synthesis were performed with U1-68/70 K expression constructs (full-length and truncated, Figure 1) derived from the original human cDNA sequence. Expression analysis show a single band of newly synthesized protein with apparent molecular weight of 27 kDa for the full-length and the truncated U1-68/70 K construct (Figure 2a), suggesting that premature termination of protein synthesis occurs before the inhibitory region (here the two constructs diverge). A fragment of the first 225 residues of U1-68/70 K up to the arginine repeats region would have a calculated MW of 26.2 kDa. Unfortunately, in SDS gels interpretation of an exact termination site is complicated because of the aberrant electrophoretic mobility of U1-68/70 K which migrates at around 70 kDa despite an calculated MW of 52 kDa [103]. Analysis of the short synthesis product by mass spectroscopy identified several

peptides covering a total of 63 amino acids between position 2 and 155 of U1-68/70 K, thus confirming the assignment as an N-terminal U1-68/70 K fragment. Premature termination of cell-free U1-68/70 K synthesis could have different explanations: for instance, limitation of arginine-loaded tRNAs in the *E. coli* cell-free system could complicate synthesis of the arginine repeat region, or the twin pair of rare AGG-AGG arginine codons at positions 172-173 causes translational problems (pausing, frame shifting, premature degradation) known from previous *in vivo* expression work [120–122]. At this stage, optimization work was started to firstly increase and verify the protein synthesis capacity of the *E. coli* cell-free system, and secondly optimize the U1-68/70 K plasmid template by a gene synthesis approach.

Therefore, the first change to the cell-free system was the application of the 'Dual Energy' system [73] to S12 bacterial extracts [25]: the ATP in long-term synthesis reactions is regenerated from creatine phosphate via creatine kinase as well as from glucose metabolism, which is enhanced through NAD^+ and coenzyme A addition. Furthermore, an increase of potassium glutamate from 0.09 M [73] to 0.13 M stabilized the consumption of Mg^{2+}. Glutamate is able to bind $Mg^{2+}$and can therefore serve as a sort of buffer [123]. While all these optimization strategies could help to increase the protein synthesis yield, they did not lead to detectable amounts of full-length U1-68/70 K protein (data not shown).

The gene synthesis approach included both full-length and truncated U1-68/70 K constructs which were codon optimized by exchanging all "rare" codons (ATA, CTA, CCC, CGA, CGG, AGA and AGG) with synonymous codons, which are preferred in "Class II" *E. coli* genes with high and continuous expression during exponential growth [124]. Within the U1-68/70 K protein constraints only slight reductions of GC content (fl: 63 % → 60 %; tr: 61 % → 59 %) and free energy content in stem loop structures (fl: -25.4 kcal/mol → -21.9 kcal/mol; tr: -24.2 kcal/mol → -22.9 kcal/mol) could be incorporated into the design.

However, no completed synthesis products were observed when expression constructs with codon-optimized synthetic U1-68/70 K genes were tested in *E. coli* cell-free reactions. Several termination products with apparent molecular weights up to 37 kDa were identified with similar patterns in full-length vs. truncated U1-68/70 K constructs (Figure 2b). As the proteins in Figure 2a and b were produced including FluoroTect™ GreenLys label (Promega GmbH, Germany), the excess lysine-loaded and labelled tRNA are visible in the two figures (band sizes smaller than 25 kDa). The increase of size and band number relative to initial observations of primary U1-68/70 K constructs indicates that the protein synthesis block of the natural cDNA

sequence is cured by codon optimization. Actually a synthesis block at a rare codon cluster, such as position 172-173 seems to limit the *in vitro* synthesis from a natural U1-68/70 K cDNA sequence. The increased size of the termination products from the synthetic gene constructs places the new synthesis block in the first arginine repeat region in front of the inhibitory sequence.

Since chaperones are known to stabilize chain elongation at the ribosome during protein expression [125], an optimization strategy using the addition of DnaK, DnaJ and GrpE in cell-free expression experiments further demonstrated the advantage of *in vitro* expression as an open system. This was an essential improvement for obtaining the U1-68/70 K protein, which can be demonstrated by the now existing difference between the truncated (50 kDa) and the full-length form (68 kDa) (Figure 2c). Both U1-68/70 K protein lengths could be expressed partly soluble, whereas the truncated protein showed a higher content of abortion products (Figure 2c).

However, at the conventional *in vitro* expression temperature of 37°C [68] only truncated U1-68/70 K was expressed (data not shown). Therefore, the following optimization step addressed the temperature dependency of U1-68/70 K protein production in the cell-free expression system. By decreasing the temperature, both full-length and truncated proteins could be expressed with the highest amount of protein at 25°C overnight (Figure 2c). Below this optimum, the truncated and the full-length U1-68/70 K were visible to a lower extent at 20°C (in SDS gels using FluoroTect^{TM} GreenLys labels); no expression could be detected at 15°C. Protein expression at lower temperatures limits aggregation [126] whereas chaperone activity increases with higher temperatures up to their optimum at 30°C [127].

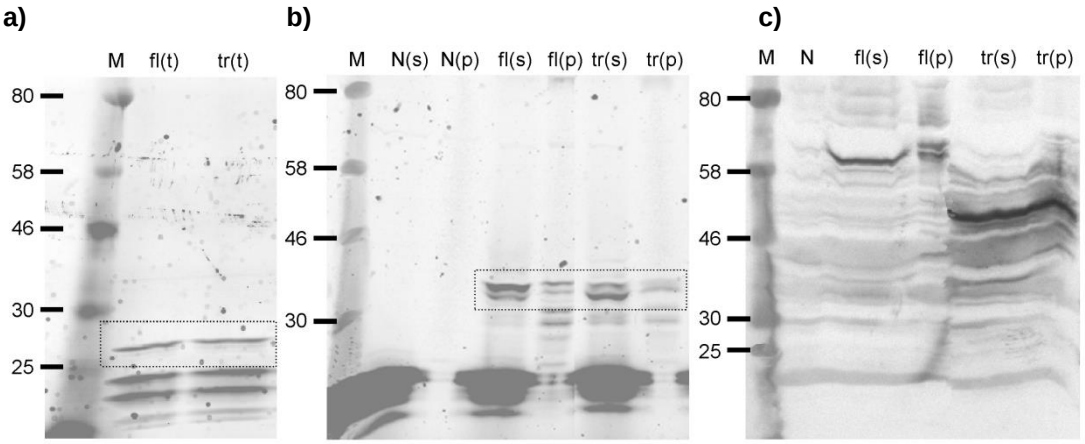


Figure 2: Expression analysis of U1-68/70 K (full-length and truncated) in the *E. coli* cell-free system (M=marker; N=negative control; fl=full-length; tr=truncated; t=total cell-free extract; s=supernatant; p=pellet fraction)

- **a) Detection of FluoroTect™ GreenLys labeled U1-68/70 K expressed with the original human cDNA sequence (proteins are indicated) via SDS-PAGE using a Cy2-filter. Excess lysine-load and labelled tRNA of FluoroTect™ GreenLys label (Promega GmbH, Germany) are also visible (band sizes smaller than 25 kDa).**
- **b) Detection of FluoroTect™ GreenLys labeled U1-68/70 K expressed with the codon optimized cDNA sequence (proteins are indicated) via SDS-PAGE using a Cy2-filter. Excess lysine-load and labelled tRNA of FluoroTect™ GreenLys label (Promega GmbH, Germany) are also visible (band sizes smaller than 25 kDa).**
- **c) Western Blot analysis (using anti-pentahistidine antibody (Qiagen GmbH, Germany) of codon optimized U1-68/70 K including DnaK/DnaJ and GroE supplement**

To demonstrate adaptability of the optimization results from the cell-free expression to a cell-based approach, the important parameters were transferred to an *E. coli* cell-based system. Whereas expression temperature (25°C) and duration (overnight) could easily be adopted, the co-expression of the chaperone system and U1-68/70 K protein required additional plasmid construction. For both proteins the overall expression yields of the natural cDNA and the codon optimized cDNA construct of U1-68/70 K co-expressing the three chaperones did not increase significantly (Figure 3a and b). However, the solubility of both truncated and full-length U1-68/70 K proteins shifted to 100 % with the codon-optimized cDNA and chaperone coexpression. The overexpressed chaperones DnaK (70 kDa), DnaJ (42 kDa) and GrpE (20 kDa) could be detected as prominent bands (Figure 3b).

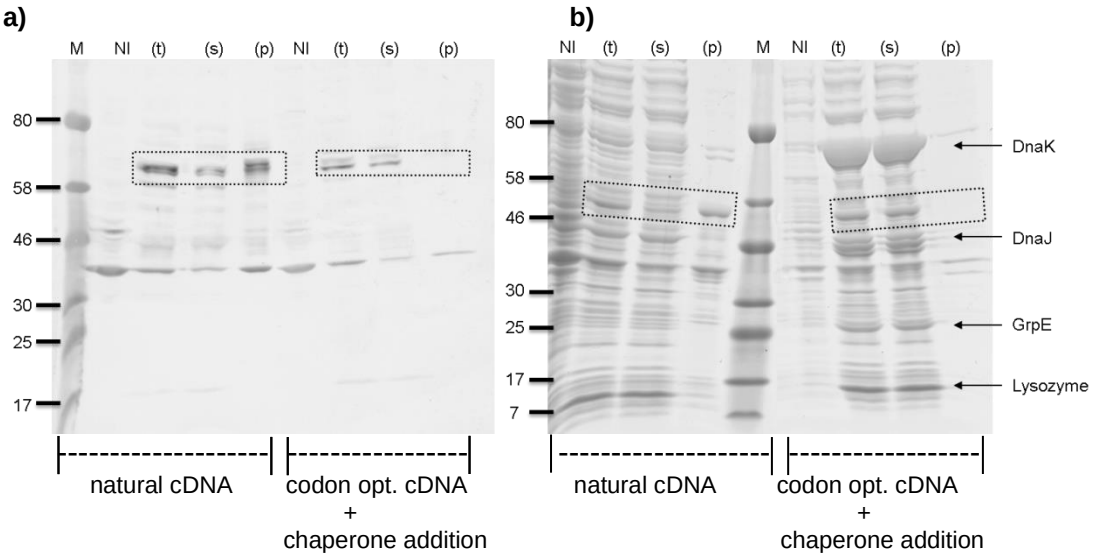


Figure 3: Expression analysis of U1-68/70 K in *E. coli* BL21 Star™ (DE3) before and after the application of the optimization strategies.
(M=marker; NI= not induced control reaction; t=total cell extract; s=supernatant; p=pellet fraction)

- **a) Full-length U1-68/70 K protein expression of the natural cDNA and the codon optimized cDNA, coexpressed with the chaperone system; proteins are indicated (dotted line). Protein detection was accomplished by Western Blot analysis using anti-pentahistidine antibody (Qiagen GmbH, Germany).**
- **b) Truncated U1-68/70 K protein expression of the natural cDNA and codon optimized cDNA, coexpressed with the chaperone system; proteins are indicated (dotted line). Protein detection was accomplished on a SDS-PAGE with Coomassie staining.**

As mentioned previously, U1-68/70 K shows an aberrant performance in gel electrophoresis, hence a mass spectrometry analysis was undertaken to prove the full-length character of the soluble protein. The overall coverage of the U1-68/70 K full-length protein by peptide mass fingerprint analysis was 28 %. The detected peptide sequences, masses and corresponding amino acid sequences are listed in supporting information (Figure A1). Using tandem mass spectrometry (MS/MS) the corresponding amino acid sequences of the four main peptides were determined. This observation of one peptide covering the end of the inhibitory sequence X into the C-terminal part proves that full-length U1-68/70 K was obtained (supporting information Figure A2).

The immunological activity of the U1-68/70 K full-length protein produced in *E. coli* cells was verified by an immunoblot using a U1-68/70 K specific patient serum.

Discussion

The U1-snRNP protein U1-68/70 K has been known for many years as a major autoantigen in MCTD and other autoimmune diseases [100,101]. To date it was not possible to express this protein in its full-length immunologically active form, because

of its inhibitory sequence X [6]. By codon usage improvements on the cDNA level and cell-free expression optimization, the expression of the immunoreactive full-length U1-68/70 K protein was successful and could be proven by MS/MS analysis and immunoblotting. Furthermore, using this *in vitro* approach, expression parameters for this specific protein (temperature, duration and chaperone addition) could be transferred to an *E. coli* cell-based expression system. Overall, the application of the cell-free expression is a simple method, which reduces optimization time and reagent costs.

Northemann *et al.* [6] concluded that the translation inhibition of the “inhibitory sequence X” within the full-length U1-68/70 K is characterized as being trans-active, suppressing protein expression under the same promoter type within one expression plasmid. An extension of these observations was made in this study by comparing *E. coli* cell-based chaperone coexpression experiments of full-length vs. truncated (lacking the inhibitory sequence X) U1-68/70 K protein: with full-length U1-68/70 K protein the coexpression levels of the three chaperones from their separate expression plasmid were partly reduced (data not shown), indicating an inhibitory effect of full-length protein on the translational activity of the cells.

In contrast to Northemann *et al*. we already detected some expression of the full-length U1-68/70 K in *E. coli* cells when using the natural cDNA sequence without chaperone addition. However, a number of differences to Northemann *et al.* can explain our observations of accumulating U1-68/70 K protein: the optimized 25° C expression temperature (instead of 37° C), use of the strong T7 promoter (instead of a tac promoter), and the longer expression duration (overnight vs. 5 h).

Full-length U1-68/70 K's stringent inhibitory effect on translation was uniquely visible in cell-free translation, where no full-length protein was found and the detection of partial products was possible. After removal of a rare codon translation obstacle by codon optimization, a translation block remained for full-length and truncated U1-68/70 K which was mapped to the beginning of the arginine-rich region (despite deletion of the "inhibitory region X" the truncated U1-68/70 K version retains part of the arginine-rich region). Non-specific-causes such as depletion of the arginine amino acid pool in the reaction can be ruled out, since a specific mechanism for the translation block is indicated by the defined nature of the partial product bands and the translation block removal in the presence of DnaK/DnaJ chaperones. In general, the DnaK and DnaJ chaperone system is known to be associated with 5-18% of newly-

synthesized proteins within *E. coli* cells [128]. Furthermore, deletion of the DnaK gene in the *E. coli* chromosome is lethal for the cell because of massive protein aggregation [125]. The peptide binding specificity of DnaK has been mapped [129]: similar to other general chaperones of the Hsp70 family DnaK binds to exposed stretches of hydrophobic peptides such as found in misfolded proteins. In addition, a unique ability of DnaK to bind arginine and lysin-rich basic peptides has been observed [129], with important implications for the present study. A reasonable theory is that the translational block acting on nascent U1-68/70 K protein is removed by binding of DnaK to the arginine-rich region of U1-68/70 K. This can happen during protein synthesis only after the arginine-rich region begins to exit from the ribosomal tunnel and becomes accessible [128]. If not captured by DnaK, the arginine-rich region can be assumed to rebind to the surface of the ribosome, possibly to a RNA component, and thereby to interfere with translational activity. To our knowledge, this is a novel mechanism for translational inhibition of an active ribosome.

Apart from the implications for ribosome biology and biotechnological expression optimization via *in vitro* approaches, the outcome of this study will enable further understanding of the spliceosomal complex including U1-68/70 K and its role in autoimmunity. Furthermore, it shows the importance of chaperones for protein folding in overexpression of heterologous proteins in *E. coli*.

Acknowledgments

We gratefully acknowledge the financial support by the Federal Ministry of Education and Research (BMBF). Thanks to Dr. Armin Neininger for his help with the codon optimization and to Dr. Heinz Haubruck and Prof. Dr. Bodo Liedvogel for their support. Thanks to the IFG Unit of the KIT especially to Dipl.-Ing. Boris Kühl for his help in MS analysis.

3.3 Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera

Carolin Richter[1,2], Kosta Konstantinidis[3], Iris Asen[1], Richard Kneusel[1,3], Jürgen Hubbuch[2]

[1]DIARECT AG, Freiburg, Germany
[2]Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Karlsruhe, Germany
[3]BadenBioTec GmbH, Freiburg, Germany

Corresponding Author: Prof. Jürgen Hubbuch (Juergen.Hubbuch@kit.edu), Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany.

Keywords: Cell-free expression / Diagnostics / Lyme disease / Protein fusion tag / Protein microarray

Abbreviations:

analysis of variance (ANOVA)
antigen (ag)
antigen matrix (am)
complementary DNA (cDNA)
hexahistidine (His)
Human serum albumin (HSA)
Nickel Chelating Sepharose Fast Flow (Ni-CS)
patient sera (ps)
sodium chloride (NaCl)
spot replicate (r)

Abstract

Lyme borreliosis is the most common tick-borne disease in North America and Europe. A two-test approach (an ELISA followed by immunoblots) for testing current and past infection has been adopted in most countries. However, the heterogeneity of *Borrelia* antigens and the semiquantitative character of the immunoblot remain a limitation. By combining a microarray system with cell-free expression, we established a procedure for the expression and subsequent printing of different *Borrelia* antigens onto several multi-well microarray plate surfaces. We successfully expressed and partially purified eleven immunodominant antigens of Lyme borreliosis from different *Borrelia* species in a self-generated *Escherichia coli* cell-free system. Using sera from patients suffering from Lyme disease and different specific monoclonal antibodies, proteins could be reproducibly detected on the microarray plates. To confirm the diagnostic outcome of the new assay, a comparison to the same cell-based expressed, purified and printed *Borrelia* antigens was performed. In summary, this approach serves as a proof of principle for the identification of potential biomarkers and offers the possibility of multiplex protein detection for specific diseases.

1 Introduction

Lyme borreliosis, caused by spirochetal bacteria from the genus *Borrelia*, is the most common tick-borne disease in North America and Europe [130]. In Germany, 1 000 000 incident cases of this disease are registered per year (*www.borreliose-nachrichten.de, 2011*). Failure to identify and treat Lyme disease early results in later onset of symptoms which may involve the joints, heart and central nervous system [130]. In Europe, five different species of *Borrelia burgdorferi* are considered to cause Lyme disease: *B. burgdorferi* sensu stricto, *Borrelia afzelii, Borrelia garinii*, *Borrelia spielmanii* and the not yet validated species *Borrelia bavariensis* [131]. In most cases, Lyme disease is diagnosed by serological confirmation after the appearance of a red skin rash (erythema migrans). The lack of standardization and poor evaluation of the sensitivity and specificity of diagnostics is caused by the broad heterogeneity of *Borrelia* strains [132]. Therefore, a two-tiered test comprising an initial screen using an ELISA followed by an immunoblot [133,134] is the method of choice. The advantages of this diagnostic approach include high sensitivity and specificity within the first weeks of a *B. burgdorferi* infection and the consistency of results in experienced laboratories [135]. Heterogeneity of the immunodominant antigens remains an issue, because a whole cell lysate immunoblot involves only one strain [136]. Additionally, standardization is difficult due to the differential expression of immunodominant proteins and the semiquantitative character of an immunoblot [135]. A promising alternative to the aforementioned difficulties is the use of recombinant antigens of *B. burgdorferi* sensu lato. The assays could include various antigens of proven diagnostic value from different strains, and they are expected to be more easily standardized and interpretable [132]. Until now different diagnostic *Borrelia* antigens from the strains were discovered and included in immunoassays [137–139]. To improve, accelerate and reduce costs of conventional diagnostic methods, protein microarrays came into the picture. As Ekins *et al.* [140] stated in their ambient analyte theory in the 1990s, a very small spot of macromolecules e.g. antibodies can provide better sensitivity than conventional immunoassays. Although DNA microarray technology became very important in gene expressing profiling [141–145], protein microarrays are more and more widely used in proteomic research [146,147]. They facilitate the identification and quantitation of proteins as well as their function in biological processes and the proteome. Furthermore, protein microarrays are predestined for measuring the amounts of high numbers of proteins in complex mixtures, as in multiplex immunoassays.

Traditionally, proteins for microarray applications are expressed in cell-based systems and subsequently purified and immobilized on the respective surface. These are laborious and time-consuming steps, which can be eliminated by producing different proteins in small amounts using cell-free expression systems [1]. Common cell-free expression systems are *E. coli*, wheat germ extract and the rabbit reticulocyte lysate [1]. The combination of cell-free expression and protein microarrays is to date accomplished by DNA printing and a subsequent protein expression on the microarray surface. This is currently used successfully for a range of applications, such as biomarker detection in cancer and autoimmune diseases, immunological studies, vaccine development, protein-protein interactions and toxin detection [148,149]. The comparability to protein microarrays is challenging, because the difference of printing DNA molecules instead of proteins is evident. Furthermore, the reproducibility of on-chip expression may limit the use of this technology in commercial diagnostics. By printing cell-free expressed proteins on microarray surfaces, the development of diagnostic microarrays for production may be improved. Faster and cost-reducing biomarker screening and the possibility of improving the protein conditions for printing may be feasible. This said the question of comparability between arrays developed on the basis of cell-free expression and those manufactured by the large-scale compatible cell-based route is still an open question.

In this study, we compared a diagnostic protein microarray that includes eleven cell-based and purified *Borrelia* antigens to a new approach, using cell-free expressed and crudely purified HaloTag® fusion *Borrelia* antigens. Both protein microarray approaches comprising the same antigens expressed using identical complementary DNA (cDNA) sequences, were printed on two different plate surfaces. The results of these investigations provide a proof of principle for the identification of potential biomarkers using cell-free expression and a multiplex protein microarray to expand the assay portfolio for the diagnosis of Lyme disease.

2 Materials and methods

2.1 Preparation of expression plasmids

Eleven different cDNA sequences originating from different *Borrelia* species (Table 1) were synthesized (Entelechon GmbH, Germany), cloned into pET24d vectors (Merck Millipore, Germany) and cell-free expressed in *E. coli* lysate. Vectors included either a N-terminal hexahistidine (His)-tag or a HaloTag7 (Promega GmbH) [5] followed by a

His-tag. The cDNA constructs of p41 *B. burgdorferi* sensu stricto and p100 *B. burgdorferi* sensu stricto has a C-terminal His-tag.

Table: Eleven *Borrelia* proteins used in this investigation and their originating species.

No.	Protein	*Borrelia* species	Molecular weight [kDa]	Molecular weight of antigen with HaloTag® [kDa][a]
ag 1	p100	*B. burgdorferi* sensu stricto	77	110
ag 2	VlsE1	*B. garinii*	27	60
ag 3	p41	*B. burgdorferi* sensu stricto	36	67
ag 4	DbpB	*B. burgdorferi* sensu stricto	19	52
ag 5	DbpA	*B. burgdorferi* sensu stricto	17	52
ag 6	DbpA	*B. afzelii*	17	50
ag 7	BmpA	*B. burgdorferi* sensu stricto	36	69
ag 8	BmpA	*B. afzelii*	36	69
ag 9	OspC	*B. spielmanii*	21	54
ag 10	OspC	*B. burgdorferi* sensu stricto	21	54
ag 11	OspA	*B. afzelii*	29	62

[a] For comparison of the gel-based cell-free expression analysis in supporting information Figure 1.

2.2 Antigens, antibodies and patient sera

Specific cell-based produced (*E. coli* or baculovirus/insect cell expression system) *Borrelia* antigens (*ag*) were kindly provided by DIARECT AG (Freiburg im Breisgau, Germany). Monoclonal antibodies (*mAb*) and patient sera (*ps*) were supplied by Dr. med. Volker Fingerle (National Reference Center for *Borrelia*, Oberschleißheim, Germany) and ravo Diagnostika GmbH (Freiburg im Breisgau, Germany), respectively. Positive patient sera (ps) were from patients diagnosed with Lyme disease by the standard two-tier method.

2.3 Preparation of bacterial cell-free extract

For cell-free expression the RNaseE-mutant *E. coli* BL21 Star™ (DE3) strain (Life Technologies GmbH, Germany) was transformed with pRARE2 (isolated from *E. coli* Rosetta2, Merck Millipore, Germany). Bacterial cell-free extract preparation was conducted according to Kim *et al.* [54] employing the simplified procedure (S12).

2.4 *In vitro* protein expression and purification of antigens

E. coli cell-free expression was carried out in 500 µl reaction volume. The composition of the cell-free reaction (Kim *et al.* [73]) was modified slightly to adapt the reaction mixture to the S12 extract. Expression temperature was set to 30°C and reactions were incubated for 6 h in a thermo mixer at 300 rpm.

The *in vitro* expression supernatants (100 x g; 5 min) were purified with spin columns containing Nickel Chelating Sepharose Fast Flow (Ni-CS) (GE Healthcare, Germany). For this, supernatants were concentrated with Vivaspin columns (Sartorius, Germany). Protein samples were adjusted to 20 mM HEPES, 500 mM sodium chloride (NaCl), 20 mM imidazol, 1% Triton X-100, pH 8.0 and bound to the columns over night at 4°C. After washing steps with 20 mM HEPES, 20 mM imidazol, pH 8.0 the columns were eluted stepwise with an increasing Imidazol concentration up to 500 mM in 20 mM HEPES, pH 8.0.

2.5 Protein analysis of antigens

To determine the binding of the HaloTag® to its ligand, unpurified and IMAC purified cell-free expressed fusion antigens (ag) were analyzed via SDS-PAGE. A negative control reaction including the HaloTag® protein without a fusion partner was analyzed in parallel. After the expression, the cell-free cultures were centrifuged (100 x g; 5 min) and equal volumes of *in vitro* reaction supernatant and previously purified fusion protein solution (15 µl) were each incubated with 3 µl of fluorescently labeled HaloTag® TMR Ligand at 37°C for 15 min. Equal volumes (30 µl) of reducing SDS sample buffer were added and samples were denatured for 5 min at 95°C and analyzed by 6–20% gradient SDS-PAGE. Gels were scanned for fluorescence detection of the HaloTag® fusion antigens specifically bound to the HaloTag® TMR Ligand in an Ettan DIGE Imager (GE Healthcare, Germany) using the Cy2 filter.

2.6 Microarray performance

For microarray printing two plate formats were used. MaxiSorp™ 96 strip well plates (Nunc™, Thermo Scientific) were used for total protein binding and mainly for antigens expressed in cells, while HaloLink™ 96 strip well plates (Promega GmbH, Germany) were used for a specific link of the cell-free expressed antigens containing a HaloTag®. A scheme of the HaloLink™ system is given in Figure 1A. Prior to the printing process, samples were adjusted to carbonate buffer (pH 9.0) by buffer exchange using Zeba Spin Desalting Columns 7K MWCO (Thermo Scientific) according to the manufacturer's protocol. Each strip of the HaloLink™ 96 strip well

plates and MaxiSorp™ 96 strip well plates was fixed in an adapter for the printing process with a contact printer (OmniGrid 100, GeneMachines, USA) using four pins. Seven replicates of each *Borrelia* antigen (Figure 1B) were printed on the two different plate type surfaces (HaloLink™ and MaxiSorp™).

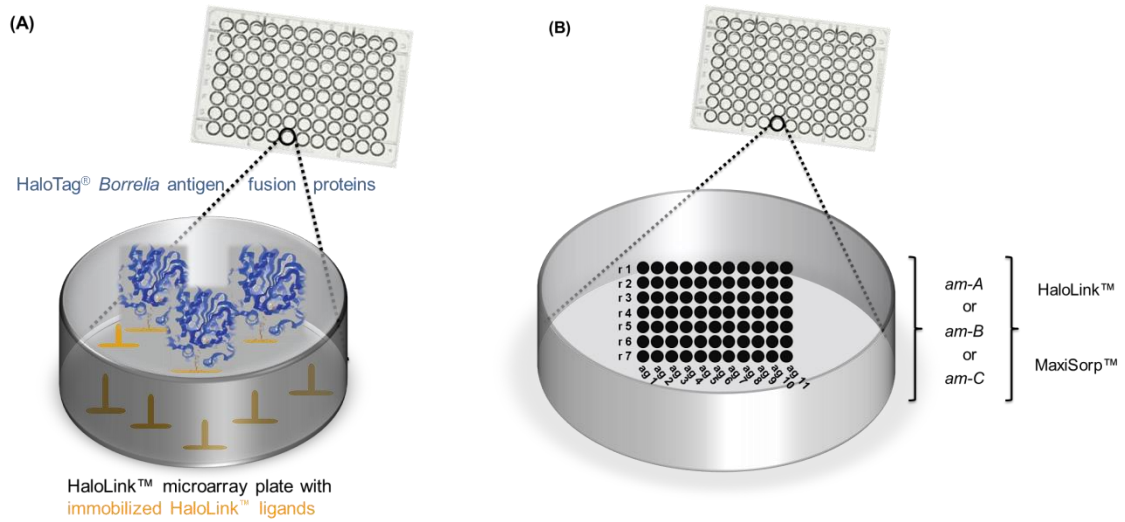


Figure 1: (A) Principle of HaloTag® binding to a HaloLink™ surface. (B) Antigen matrices A, B and C (*am-A*: unpurified cell-free expressed *Borrelia* proteins; *am-B*: purified cell-free expressed *Borrelia* proteins; *am-C*: purified cell-based expressed *Borrelia* proteins) as applied in the multiplex protein microarrays. Seven replicate spots of *Borrelia* antigens were printed on two different plate type surfaces (HaloLink™ and MaxiSorp™).

2.7 Antigen matrix for printing

Three different antigen matrices (am) based on the different expression and purification approaches were used for the printing process: am-A, cell-free expressed HaloTag® fusion antigens, unpurified; am-B, cell-free expressed and Ni-CS purified HaloTag® fusion antigens; and am-C, cell-based expressed and purified *Borrelia* antigens (Figure 1B).

2.8 On-print controls

For evaluation and validation of the printing process, human IgG and IgM (DIANOVA GmbH, Hamburg, Germany) were spotted as positive controls in each microarray pattern. Human serum albumin (HSA) (Sigma, Germany) and crude *E. coli* lysate, which was also used for cell-free expression, were printed as negative controls.

2.9 Microarray processing

The dried microarray spots on the plate surfaces (2 h at 37°C) were blocked with StableGuard Choice (Surmodics, USA). To determine the diagnostic performance of

the distinct array types, nine different monoclonal antibodies (*mAb1 – mAb9*) against defined *Borrelia* antigens and sera from 14 different patients suffering from Lyme disease (*ps1 – ps14*) were used. For negative control reaction, sera from individual healthy blood donors were applied. The monoclonal antibodies and patient sera were diluted 1:100, secondary antibodies (anti-human IgG antibody-Cy5, anti-human IgM antibody-Cy5, anti-mouse IgG antibody-Cy5 [DIANOVA GmbH, Hamburg, Germany]) 1:1000 in StableGuard Choice. Following incubation with the primary (60 min) and secondary (30 min) antibodies, the wells were washed three times with PBST (3.7 mM NaH_2PO_4, 16.3 mM Na_2HPO_4, 0.15 M NaCl, 0.05 % Tween20) followed by three washes with PBS buffer. The developed microarrays were scanned in a Fluorescence Array Imaging Reader "FLAIR" (Sensovation, Germany).

2.10 Microarray data analysis

Data normalization was performed to compensate specific effects arising from different plate types, different protein purities and the presence or absence of tags. Furthermore, microarrays in general show systematic effects due to array characteristics which need to be compensated. The raw fluorescence microarray data was normalized as shown in Equation (1), starting with a $\log_2$ transformation of the raw fluorescence data and followed by scale normalization including identically processed antigens (Beissbarth *et al.*, Recommendations for normalization of microarray data. 2005 [www.science.ngfn.de]), [150]. The factor k (specifically k=65535) was introduced due to instrument specific data generation.

$$\text{normalized data} = \log_2(\text{fluorescence data} * k) - \text{median}[\log_2(\text{fluorescence data} * k)\ \text{antigens identically processed}] \quad (1)$$

Spot reproducibility was defined to evaluate the information value of the printing process. It was determined by using the raw fluorescence data of the seven spot replicates (*r1 – r7*) developed by the 14 patient sera and calculating the coefficient of variation (CV) as outlined in Equation (2).

$$CV = SD(\text{replicates})/\text{median}(\text{replicates}) * 100\ [\%] \quad (2)$$

For an overall comparison of the three different antigen matrices, the CV frequencies of defined percentage ranges (0-5%, 6-10%, 11-15%, 16-20% and higher than 20%) were compared. The lower the CV value, the higher the reproducibility of the spot replicates.

To ensure the comparability of the different antigen matrices and the different plate types after data normalization, an evaluation of the on-print controls was necessary. Therefore, an analysis of variance (ANOVA) was performed on the four different on-print controls; two positive controls (IgG and IgM) and two negative controls (HSA and the cell-free *E. coli* extract). The null-hypothesis in the ANOVA proposed no significant difference between the examined observations. As a result of this evaluation, p-values were obtained, describing the probability of a significant difference. In general p-values lower than a level of 0.05 describe significant effects of plates or matrices while p-values greater than 0.05 indicate no significant effects and therefore confirm the comparability of the antigen matrices on the different plate types.

3 Results and Discussion

3.1 Cell-free expression of active HaloTag® fusion antigens

The overriding prerequisite for the given study is the success of *Borrelia* fusion antigen cell-free expression and interaction of the HaloTag®–HaloLink™ system. Eleven antigens originating from different *Borrelia* species (Table 1) were expressed in an *E. coli* based cell-free expression system. By including the HaloTag® sequence to the cDNA of the antigens, specific fusion protein detection with the HaloTag® TMR-ligand was possible. Previous experiments indicated that the HaloTag® TMR-ligand only binds covalently to a soluble and active HaloTag® fusion protein (data not shown). Therefore, activity of the HaloTag® fusion protein is shown by the gels presented in supporting information Figure 1 since band detection was realized through fluorescence originating from the bound HaloLink™ TMR-ligand. *Borrelia* antigens, with exception of p100 *B. burgdorferi* sensu stricto*,* were successfully expressed in their soluble and active form with the *E. coli* cell-free expression system (supporting information Figure 1A) and purified by Ni-CS spin columns (supporting information Figure 1B). The HaloTag® (32 kDa) was also expressed cell-free in relatively high amounts (red dotted line, supporting information Figure 1A). After the Ni-CS purification almost no tag contamination was visible (supporting information Figure 1B) in contrast to the nonpurified samples in supporting information Figure 1A.

3.2 Development of microarray system based on cell-free expressed antigens

Prior to a detailed comparison of microarrays based on cell-free and cell-based expressed antigens it was necessary to develop and optimize the cell-free expression and subsequent binding of the respective antigens. The CV values obtained with the

14 patient sera were tested for the three antigen matrices and the two plate formats. When comparing CV frequencies as determined from the raw fluorescence data of spot replicates (Figure 2) a distinct performance could be observed. The CV values within the 15% range were for *am-A*, representing the cell-free expressed, unpurified antigens, 89% on the HaloLink™ and 97% on the MaxiSorp™ plate (Figure 2A). For both plate types *am-B*, the partially purified cell-free expressed antigens, showed for 92% of all spot replicates a maximum CV of 15% (Figure 2B). This indicates high spot reproducibility for both antigen matrices on both plate types.

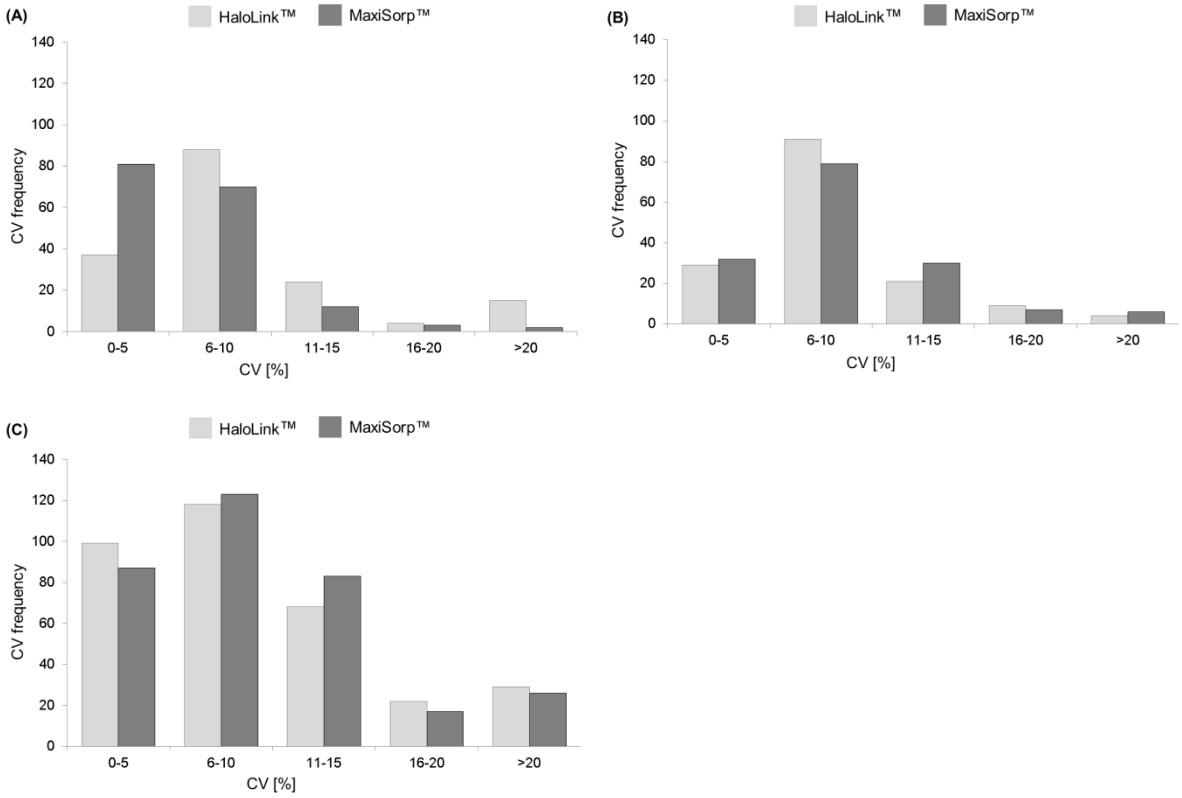


Figure 2: Frequency distribution of coefficients of variation (CVs), *ag* detected with the 14 patient sera. (A) cell-free expressed, unpurified (*am-A*) and (B) cell-free expressed, purified *Borrelia* antigens (*am-B*) and (C) cell-based expressed and purified *Borrelia* antigens (*am-C*) on both plate types (MaxiSorp™ and HaloLink™).

To ensure the comparability of the different antigen matrices and plate types of the cell-free expression system an ANOVA evaluation of the on-print controls was performed. It showed that both positive controls (IgG and IgM) and the HSA negative control generate a p-value higher than 0.05 (supporting informationTable 1) implying no significant difference between the three antigen matrices and the two plate types considering the controls. In contrast, the ANOVA of the *E. coli* lysate negative control showed a p-value less than 0.05 (supporting informationTable 1). Due to these significant differences of the lysate control *am-A* and *am-B* cannot be compared.

Therefore, a comparison to the cell-based system (*am-C*) was only performed for *am-B* – the partially purified cell-free expressed antigens.

For an evaluation of the suitability of the different plate types (HaloLink™ and MaxiSorp™) for the cell-free expressed *am-B* all antigens *ag1* – *ag11* were printed and incubated with patient serum eleven (*ps11*). The distinct microarray pattern of this analysis is shown in Figures 3A and B. The normalized fluorescence signals of the eleven *Borrelia* antigens developed with *ps11* show high similarities. The mean scattering is slightly higher when using MaxiSorp™ plates. The print pictures of the seven spot replicates verify the fluorescence data. Similarly, when printing the cell-based expressed and purified His-tagged antigens on both plate types, a statistical comparison with N-Way ANOVA shows that the null hypothesis is not rejected ($p<0.05$) implying no statistical significant difference between the two plate types.

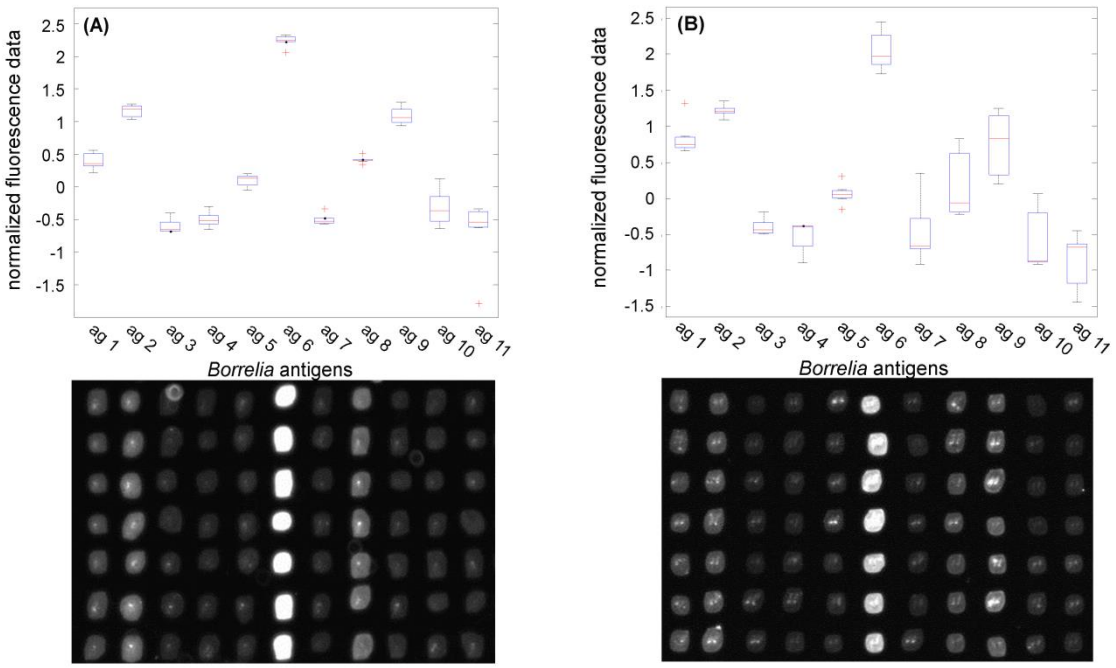


Figure 3: Boxplots of the normalized fluorescence data of the eleven *Borrelia* antigens (*am-B* - cell-free expressed HaloTag® fusion proteins) printed on (A) HaloLink™ and (B) MaxiSorp™ plates, after detection with *ps11*. The corresponding seven replicates of the eleven *Borrelia* antigens as visible after microarray printing are illustrated below. (Numbering according to Table 1)

3.3 Development of microarray system based on cell-based expressed antigens

Analogous to the evaluation of comparability in the cell-free system the cell-based system – *am-C* – and plate format combinations were analyzed. After development with patient sera, spot reproducibility with regard to the different plate types was determined and the CV frequencies from the raw fluorescence data of the spot replicates (Figure 2C) were compared. 85% of the frequency distributions of all spot

replicates were found with a maximum CV of 15% on the HaloLink™ plates. For MaxiSorp™ an even higher value of 87% within the before mentioned CV range could be observed. This is an advantage, because all following comparisons of *am-C* are conducted on MaxiSorp™ plates. It indicates again high spot reproducibility for the *am-C* on both plate types. The frequency distributions of the CVs from the cell-free expressed, partially purified *Borrelia* antigens (Figure 2B; *am-B*) and the cell-based expressed *Borrelia* antigens (Figure 2C; *am-C*) on both plate types (MaxiSorp™ and HaloLink™) of all spot replicates with a maximum CV of 15% show an overall percentage of about 90. Therefore, the requirement of high spot reproducibility in all approaches and on both plate types is met.

3.4 Microarray performance: Cell-free vs. cell-based expressed antigens

In order to compare the feasibility of our approach and to evaluate the comparability of the microarrays produced with an antigen matrix based on cell-free synthesis and cell-based expression we performed three interaction studies comprising mAb2 interaction with *ag1 – ag11* (case 1), *ps11* interaction with *ag1 – ag11* (case 2) and *ps1 – ps14* interaction with *ag6*. We finally analyzed the complete matrix of interactions comprising *ag1 – ag11*, *mab1 – mAb9* and *ps1 – ps14* (case 4). All statistical comparisons between the plate type (HaloLink™ and MaxiSorp™), the application of monoclonal antibodies (*mAb1 – mAb9*) or patient sera (*ps1 – ps14*), and the eleven *Borrelia* antigens applied *ag1 – ag11* (Table 1), were performed by N-Way ANOVA.

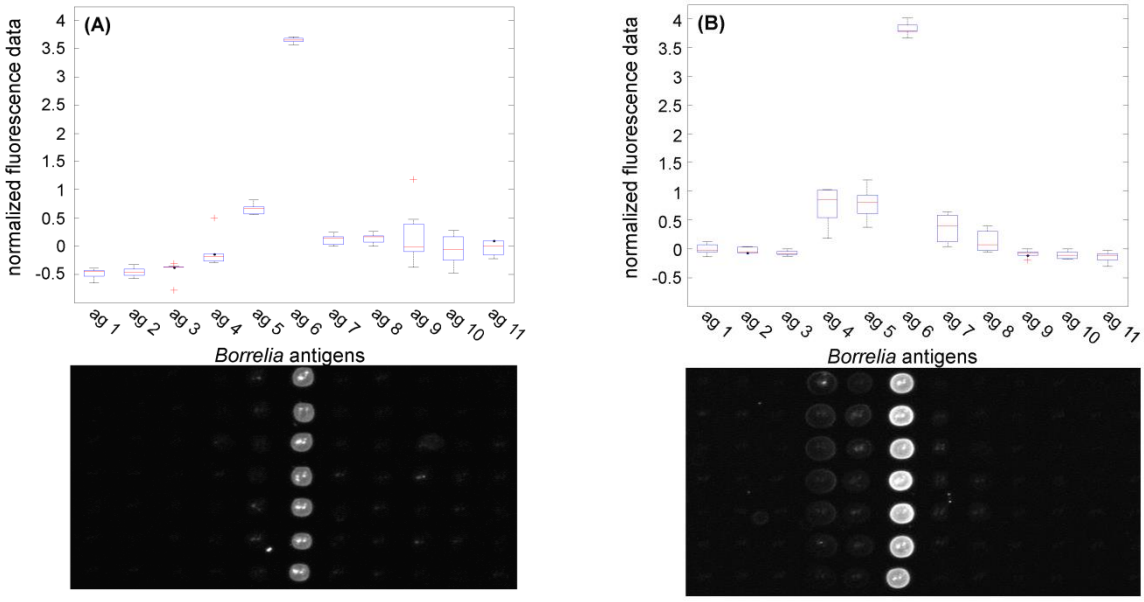
3.4.1 Case 1: *mAb2* interaction with *ag1 – ag11*

For comparison of the two microarray approaches (*am-B* and *am-C*), boxplots for every monoclonal antibody or patient serum and the eleven *Borrelia* antigens of one plate type were generated. For arrays with *am-B* on HaloLink™ (Figure 4A) and *am-C* on MaxiSorp™ (Figure 4B) developed with *mAb2*, high similarities are apparent with specific detection of *ag6* (DbpA, *Borrelia afzelii*). Equivalent results were found for three other proteins specifically detected by monoclonal antibodies, namely *ag1* (p100, *B. burgdorferi* sensu stricto) (protein was produced in an extra and further optimized cell-free expression), *ag11* (OspA, *B. afzelii*) and *ag8* (BmpA, *B. afzelii*) (data not shown). The mAbs detect the respective antigen specifically, showing almost no interaction with any other antigens. In addition, the detection intensities are very high, including very low standard deviations. The performance of the cell-free approach *am-B* is highly similar to the cell-based expressed antigens in *am-C*,

showing the applicability of the cell-free system for microarray screenings when mAbs are used.

3.4.2 Case 2: *ps11* interaction with *ag1 – ag11*

Developing the arrays with patient sera generally led to a detection of more than only one antigen. By comparing the normalized fluorescence data of the eleven *Borrelia* antigens after development with *ps11* in a boxplot and the corresponding microarray printing profiles in Figure 4C and D, the proteins *ag2* (VlsE1, *B. garinii*), *ag6* (DbpA, *B. afzelii*) and *ag9* (OspC, *B. spielmanii*) display the highest normalized fluorescence values indicating specific detection in both approaches (Figs. 4C and D). No significant difference between the matrix *am-B* (Figure 4C; HaloLink™) and *am-C* (Figure 4D; MaxiSorp™) as well as the respective plate type is observable. The variation between the seven print replicates is slightly lower when comparing the cell-free expressed antigens on the HaloLink™ plate to *am-C*. However, the MaxiSorp™ plate already showed a slightly higher mean scattering before in the plate comparison experiments when only *am-B* was printed. The specific detection of both systems is comparable and therefore, the applicability of *am-B* in contrast to *am-C* could be shown for *ps11*.



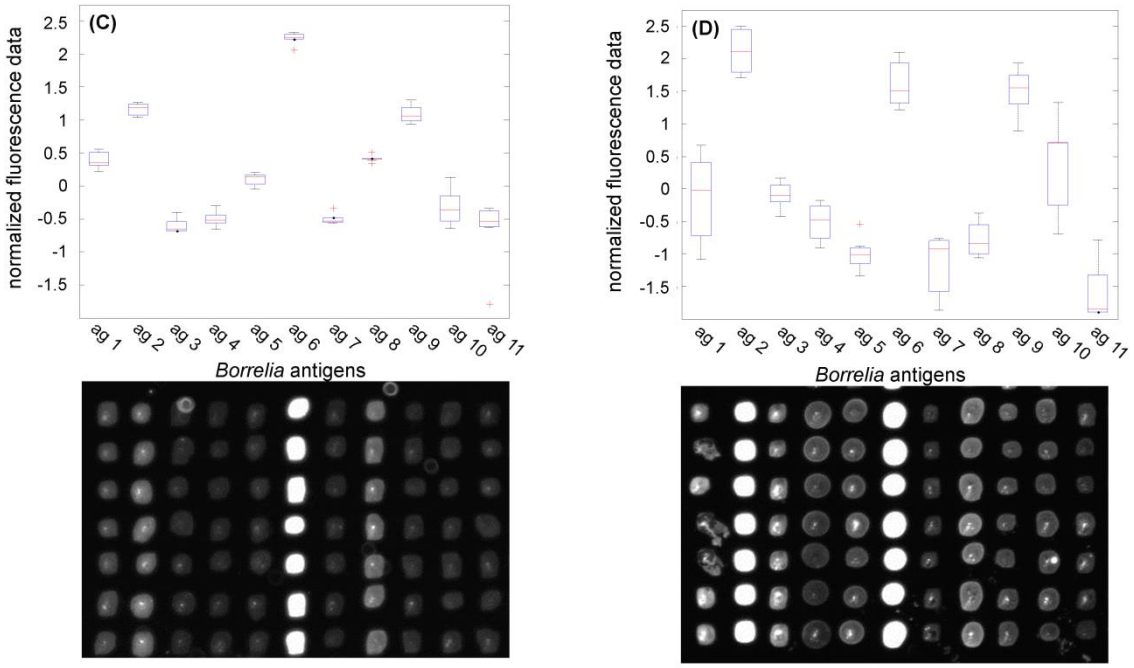


Figure 4: Boxplots of the normalized fluorescence data of the eleven *Borrelia* antigens after detection with *mAb2* (A and B) and *ps11* (C and D). The corresponding seven replicates of the eleven *Borrelia* antigens as visible after microarray printing are illustrated below. (A and C) *am-B* (cell-free expressed *Borrelia* HaloTag® fusion proteins purified by Ni-CS) on HaloLink™ plates. (B and D) *am-C* (cell-based expressed *Borrelia* antigens IMAC-purified) on MaxiSorp™ plates. (Numbering according to Table)

3.4.3 Case 3: *ps1* – *ps14* interaction with *ag6*

In addition to the comparison of the two microarray plates (HaloLink™ and MaxiSorp™) after development by one certain patient serum, Figure 5 illustrates fluorescence data of *ag6* (DbpA from *B. afzelii*) after incubation with the 14 patient sera and the negative control (blood donor serum). The bar plots depict the normalized fluorescence data with the different patient sera on the microarray matrix *am-B* on HaloLink™ (Figure 5A) and *am-C* on MaxiSorp plates ™ (Figure 5B). The latter two microarray types show high similarities for the 14 patient sera and the blood donor. The data indicates that *ag6* (DbpA, *B. afzelii*) detection with seven of the patient sera was just as small as with the negative control serum. In all these cases the normalized fluorescence data are close to zero (Figure 5). However, seven patient sera react positively with *ag6* (DbpA, *B. afzelii*) in both approaches, a finding that is reproducible for the seven spots respectively (Figs. 5A and B). Overall, eight out of eleven *Borrelia* antigens reacted similarly on the different plate types in both approaches. Since the interactions of the cell-free expressed antigens are similar to the cell-based expressed antigens over the patient sera used in this study, specific detection is possible with cell-free expressed antigens and the applicability of this system could be shown.

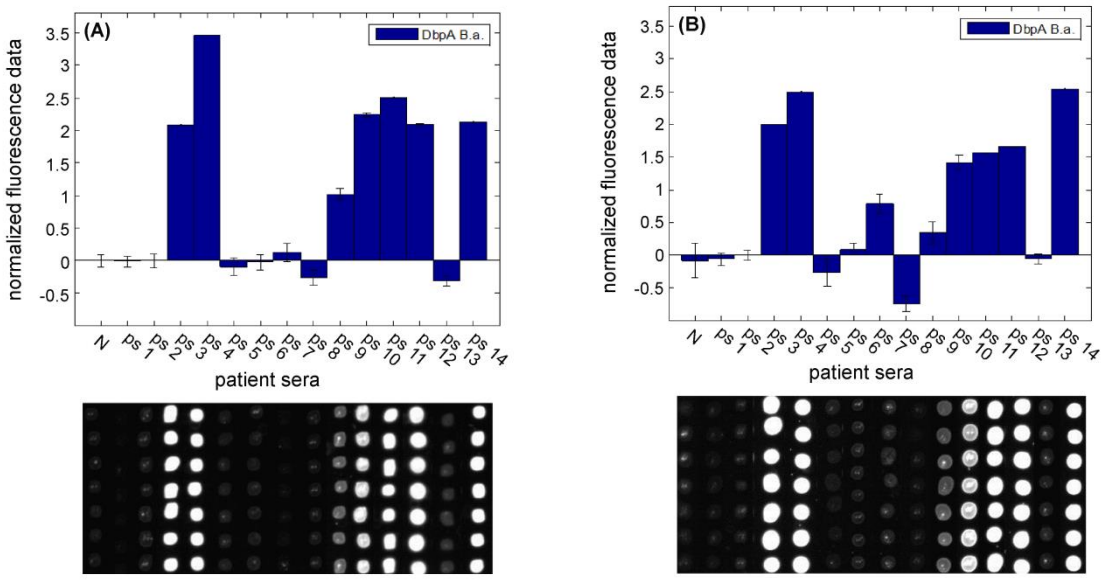


Figure 5: Bar plots of the normalized fluorescence data of *ag6* DbpA from *B. afzelii* after detection with the 14 different patient sera (N = blood donor serum). The error bars represent the standard deviation. The corresponding seven replicates of DbpA from *B. afzelii* after the particular developments as visible after microarray printing are illustrated below. (A) *am-B* (cell-free expressed *Borrelia* HaloTag® fusion proteins purified by Ni-CS) on HaloLink™ plates. (B) *am-C* (cell-based expressed *Borrelia* antigens IMAC-purified) on MaxiSorp™ plates.

3.4.4 Case 4: Overall data evaluation – *ag1-ag11* / *mAb1-mAb9* / *ps1-ps14*

By comparing the antigen matrices B and C on both plate types, four out of eleven *Borrelia* antigens were statistically different from the others to positive detection by the monoclonal antibodies (Figure 6A). For the other seven *Borrelia* antigens it was not possible to reject the null hypothesis and therefore no significant effect can be assumed when compared to the negative on-print control (HSA). Likewise for three out of the eleven *Borrelia* antigens the null hypothesis could not be rejected after incubation with the 14 patient sera and hence could not be detected with the sera pool of this study (Figure 6B). However, *ag11* (OspA, *B. afzelii*) was detectable by *mAb6*. Only two of the eleven antigens – *ag3* and *ag7* - remained undetermined after measurement. Therefore, 82% of the *Borrelia* antigens examined were detected by the different patient sera and monoclonal antibodies applied, illustrating the suitability of the investigated antigens. From the results obtained it is evident that printing cell-free expressed *Borrelia* antigens is reproducible, even on different plate types. With the on-print controls the comparability between the different approaches (e.g. cell-based and cell-free expressed proteins, usage of different plate types etc.) is indicated and viable. In addition, the results obtained in the different approaches, including different microarray plate surfaces and antigen matrices, show – based on statistical evaluations – a high similarity. The results are thus a successful demonstration of the feasibility to apply a time- and resource-saving combination using cell-free expressed

proteins for printing, successfully to a recombinant antigen system for the diagnosis of Lyme borreliosis. HaloTag® fusion proteins are to date pipetted in gaskets onto the HaloLink™ surfaces, which differs from the printing process basically in reagent volume and sample number [151]. In this study we printed the cell-free expressed HaloTag® fusion proteins, increasing the number of possible experiments from 50 per slide in conventional methods up to a spot density of 196 dots per well of a microtiter plate.

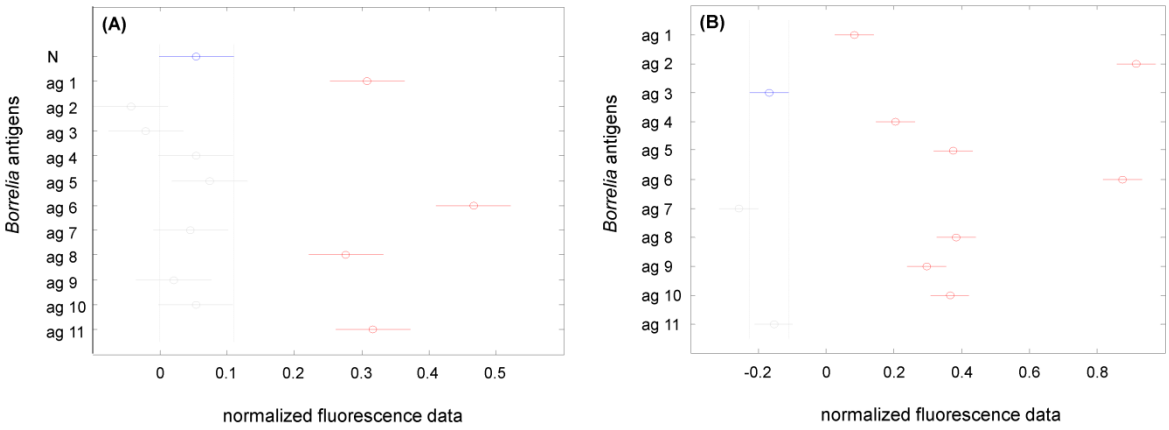


Figure 6: Multcompare plots illustrating the eleven *Borrelia* antigens as well as the negative control (N = HSA) after detection with the nine *mAbs* (A) and the 14 patient sera (B) as calculated by N-Way ANOVA. (Numbering according to Table 1)

4 Concluding remarks

Previously, proteins for microarray applications were expressed in cell-based systems, which is time and material-consuming. New applications using cell-free expression for a cheap and fast alternative printed the DNA molecules, but not the proteins itself. Therefore, the comparability to protein microarrays is challenging. This study provides a proof of principle for the identification of potential new biomarkers using proteins from cell-free expression combined with multiplex protein microarrays, representing a fast and cost-reducing approach compared to cell-based expression and conventional analytic methods. Furthermore, the comparability to cell-based expressed protein microarrays could be reproducibly shown and hence the improvement of printing conditions is now possible. Future applications may expand the diagnostic potential to simultaneous analyses of an almost arbitrarily high number of e.g. biomarkers in a relatively short time.

Acknowledgements

We gratefully acknowledge the financial support by the Federal Ministry of Education and Research (BMBF) on the grant number 0315334A. The authors have declared no conflict of interest.

4 Conclusion and Outlook

The research outcome of this study contributes to help simplifying and accelerating the biochemical protein production and analysis. Recently, a high number of genes from different species were identified. However, conventional cell-based expression is limited in expressing these gene sequences for functional classification and validation of the corresponding proteins. To overcome these limitations, cell-free expression can be used to produce these proteins which are of high value in the personalized medicine and pharmaceutical research.

Through the application of DoE and model-based analysis the influences of the different premix components and their concentrations on the luciferase expression yield of two different insect cell extracts (Sf9 and High Five™) were observed and characterized. The use of the high-throughput method enables many data points to be generated. To the experimental data quadratic response surface models were fitted and model predictivity was validated successfully. The characterization of the complete *in vitro* translation process included quantification and visualization of the parameter influences on the expression yield and the robustness of the systems. Furthermore, the results extend and simplify previous insect cell-free optimizations. Therefore, the establishment of a new method for further investigations and optimization experiments of cell-free extracts in different cell types has been successfully accomplished. In the future, using this comprehensive system characterization based on optimally designed high-throughput screenings on robotic systems a further enhancement of protein yield as well as cell-free system optimizations regarding single protein expression improvements can be established.

Furthermore, this thesis showed that cell-based protein expression optimization using cell-free expression systems is possible. The process parameters established during the cell-free approach could be successfully transferred to the cell-based system. Additionally, this research showed the importance of protein folding for the overexpression of heterologous proteins in *E. coli*. For the first time, a soluble, full-length and immunologically active U1-68/70 K autoantigen was expressed in *E. coli* cells. Further understanding of the human U1-68/70 K full-length protein and the spliceosomal complex including U1-68/70 K and its role in autoimmunity is now possible.

We established a procedure for the cell-free expression and subsequent printing of different *Borrelia* antigens onto several multi-well microarray plate surfaces. This is a fast and cost-saving approach compared to cell-based expression and conventional

analysis methods. Furthermore, the statistical analysis showed that cell-free expression is reproducibly comparable to cell-based expressed protein microarrays. Furthermore, the improvement of printing conditions, depending on expression techniques, is now possible. Therefore, this approach serves as a proof of principle for the identification of potential biomarkers. In the future, it offers the possibility of multiplex protein detection for specific diseases and may therefore expand the diagnostic biomarker portfolio in a relatively short time.

Overall the research in this thesis shows different types of applications of cell-free expression. Despite the recently significantly improved reaction yields in *in vitro* protein production systems the potential for optimization is not yet exhausted. Especially linking cell-free expression to protein analysis needs further examination. In the future, cell-free expressions with integrated folding proteins or other additives will expand the portfolio of desired protein for personalized medicine and pharmaceutical research.

5 Abbreviations

adenosine triphosphate (ATP)
amino Acids (AA)
analysis of variance (ANOVA)
antigen (ag)
antigen matrix (am)
base pair (bp)
calcium chloride ($CaCl_2$)
coefficient of variation (CV)
complementary DNA (cDNA)
coenzym A (CoA)
creatine kinase (CK)
creatine phosphate (CP)
cytosin triphosphate (CTP)
deoxyribonucleic acid (DNA)
design of experiments (DoE)
Dithiothreitol (DTT)
ethylene glycol tetraacetic acid (EGTA)
enzyme-linked immunosorbent assay (ELISA)
Escherichia coli (E. coli)
green fluorescent protein (GFP)
guanosine triphosphate (GTP)
hexahistidine (His)
high-throughput (HT)
high-throughput screening (HTS)
human serum albumin (HSA)
immunglobulin G (IgG)
immunglobulin M (IgM)
kilo base pair (kbp)
magnesium acetate ($MgO(Ac)_2$)
mass spectrometry (MS)
matrix-assisted laser desorption/ionization (MALDI)
messenger ribonucleic acid (mRNA)
mixed connective tissue disease (MCTD)
monoclonal Antibody (mAb)
multivariate data analysis (MVDA)
Nickel Chelating Sepharose Fast Flow (Ni-CS)

nicotinamide adenine dinucleotide (NAD)
probability value (p-value)
patient sera (ps)
polymerase chain reaction (PCR)
potassium acetate (KOAc)
prediction sum of squares (PRESS)
ribonucleic acid (RNA)
root mean square error (RMSE)
small nuclear ribonucleic particles (snRNP)
sodium chloride (NaCl)
sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)
spot replicate (r)
tandem mass spectrometry (MS/MS)
transfer ribonucleic acid (tRNA)
uracil triphosphate (UTP)
untranslated region (UTR)

6 References

[1] M. He, Cell-free protein synthesis: applications in proteomics and biotechnology, N. Biotechnol. 25 (2008) 126–132.

[2] M. He, Y. He, Q. Luo, M. Wang, From DNA to protein: No living cells required, Process Biochem. 46 (2011) 615–620.

[3] E.D. Carlson, R. Gan, C.E. Hodgman, M.C. Jewett, Cell-free protein synthesis: applications come of age., Biotechnol. Adv. 30 (2011) 1185–1194.

[4] T. Suzuki, T. Ezure, M. Ito, M. Shikata, An Insect Cell-Free System for Recombinant Protein Expression Using cDNA Resources, in: H. Koga (Ed.), Methods Mol. Biol., Humana Press, Totowa, NJ, 2009: pp. 97–108.

[5] R.F. Ohana, K.Q. Zhao, L.P. Encell, D. Simpson, M. Slater, M. Urh, *et al.*, HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification, Protein Expr. Purif. 68 (2009) 110–20.

[6] W. Northemann, H. Berg, G. Stahnke, M. Walter, N. Hunt, S. Fenning, Identification of an inhibitory element within the human 68-kDa (U1) ribonucleoprotein antigen, Protein Expr. Purif. 6 (1995) 748–756.

[7] F. Katzen, T.C. Peterson, W. Kudlicki, Membrane protein expression: no cells required., Trends Biotechnol. 27 (2009) 455–60.

[8] L. Kaiser, J. Graveland-Bikker, D. Steuerwald, M. Vanberghem, K. Herlihy, S. Zhang, Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses., Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 15726–31.

[9] H.-P. Phan, T. Ezure, M. Ito, T. Kadowaki, Y. Kitagawa, T. Niimi, Expression and Chain Assembly of Human Laminin-332 in an Insect Cell-Free Translation System, Biosci. Biotechnol. Biochem. 72 (2008) 1847–1852.

[10] K.A. Calhoun, J.R. Swartz, Energizing cell-free protein synthesis with glucose metabolism., Biotechnol. Bioeng. 90 (2005) 606–13.

[11] M.C. Jewett, K. a Calhoun, A. Voloshin, J.J. Wuu, J.R. Swartz, An integrated cell-free metabolic platform for protein production and synthetic biology., Mol. Syst. Biol. 4 (2008) 220.

[12] H.-C. Kim, T.-W. Kim, D.-M. Kim, Prolonged production of proteins in a cell-free protein synthesis system using polymeric carbohydrates as an energy source, Process Biochem. 46 (2011) 1366–1369.

[13] J.Watson, F.H. Crick, Molecular structure of nucleic acids, Nature. 171 (1953) 737–738.

[14] S.H. Eom, J. Wang, T.A. Steitz, Structure of Taq polymerase with DNA at the polymerase active site., Nature. 382 (1996) 278–81.

[15] D.H. Persing, MINIREVIEW Polymerase Chain Reaction : Trenches to Benches, 29 (1991) 1281–1285.

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

[17] A.M. Jackson, J. Boutell, N. Cooley, M. He, Cell-free protein synthesis for proteomics, 2 (2004).

[18] T.U. Yaeta Endo, Kazuyuki Takai, Cell-Free Protein Production Methods and Protocols, 2009.

[19] S. Rungpragayphan, H. Nakano, T. Yamane, PCR-linked *in vitro* expression: a novel system for high-throughput construction and screening of protein libraries, FEBS Lett. 540 (2003) 147–150.

[20] A. Sachs, P. Sarnow, M. Hentze, Starting at the Beginning, Middle, and End: Translation Initiation in Eukaryotes, Cell. 89 (1997) 831–838.

[21] M. Mie, S. Shimizu, F. Takahashi, E. Kobatake, Selection of mRNA 5'-untranslated region sequence with high translation efficiency through ribosome display., Biochem. Biophys. Res. Commun. 373 (2008) 48–52.

[22] T. Suzuki, M. Ito, T. Ezure, S. Kobayashi, M. Shikata, K. Tanimizu, *et al.*, Performance of expression vector, pTD1, in insect cell-free translation system., J. Biosci. Bioeng. 102 (2006) 69–71.

[23] T. Ezure, S. Kobayashi, T. Suzuki, S. Higashide, E. Shintani, K. Endo, *et al.*, Cell-Free Protein Synthesis System Prepared from Insect Cells by Freeze-Thawing, Biotechnol. Prog. 22 (2006) 1570–1577.

[24] T. Ezure, T. Suzuki, M. Shikata, M. Ito, E. Ando, T. Utsumi, *et al.*, Development of an insect cell-free system, Curr. Pharm. Biotechnol. 11 (2010) 279–284.

[25] G. Kumar, G. Chernaya, Cell-free protein synthesis using multiply-primed rolling circle amplification products., Biotechniques. 47 (2009) 637–9.

[26] J.-H. Ahn, H.-S. Chu, T.-W. Kim, I.-S. Oh, C.-Y. Choi, G.-H. Hahn, *et al.*, Cell-free synthesis of recombinant proteins from PCR-amplified genes at a comparable productivity to that of plasmid-based reactions, Biochem. Biophys. Res. Commun. 338 (2005) 1346–52.

[27] D.-M. Kim, C.-Y. Choi, J.-H. Ahn, I.-S. Oh, T.-W. Kim, C.-G. Park, Development of a rapid and productive cell-free protein synthesis system, Biotechnol. Bioprocess Eng. (2006) 235–239.

[28] H.Z. Chen, G. Zubay, Prokaryotic coupled transcription-translation., Methods Enzymol. 101 (1983) 674–690.

[29] A.S. Spirin, V.I. Baranov, L.A. Ryabova, S. YU., Y.B. OVODOV, ALAKHov, A Continuous Cell-Free Translation System Capable of Producing Polypeptides in High Yield, Science (80-.). (1988).

[30] B.G. Erickson AH, Cell-free translation of messenger RNA in a wheat germ system., Methods Enzym. 96 (1983) 38–50.

[31] Y. Endo, High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system, Biotechnol. Adv. 21 (2003) 695–713.

[32] H.T. Jackson RJ, Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA, Methods Enzym. 96 (1983) 50–74.

[33] R. Hempel, J. Schmidt-Brauns, M. Gebinoga, F. Wirsching, a Schwienhorst, Cation radius effects on cell-free translation in rabbit reticulocyte lysate., Biochem. Biophys. Res. Commun. 283 (2001) 267–72.

[34] E. Gasior, F. Herrera, I. Sadnik, C.S. McLaughlin, K. Moldave, The preparation and characterization of a cell-free system from Saccharomyces cerevisiae that translates natural messenger ribonucleic acid., J. Biol. Chem. 254 (1979) 3965–9.

[35] P. Sarnowl, N. Iizuka, L. Najita, A. Franzusoff, Cap-Dependent and Cap-Independent Translation by Internal Initiation of mRNAs in Cell Extracts Prepared from Saccharomyces cerevisiae, 14 (1994) 7322–7330.

[36] M.R. Swerdel, a M. Fallon, Cell-free translation in lysates from Spodoptera frugiperda (Lepidoptera: Noctuidae) cells., Comp. Biochem. Physiol. B. 93 (1989) 803–6.

[37] S. Mureev, O. Kovtun, U.T.T. Nguyen, K. Alexandrov, Species-independent translational leaders facilitate cell-free expression., Nat. Biotechnol. 27 (2009) 1–20.

[38] S. Mikami, T. Kobayashi, S. Yokoyama, H. Imataka, A hybridoma-based *in vitro* translation system that efficiently synthesizes glycoproteins., J. Biotechnol. 127 (2006) 65–78.

[39] S. Mikami, M. Masutani, N. Sonenberg, S. Yokoyama, H. Imataka, An efficient mammalian cell-free translation system supplemented with translation factors, Protein Expr. Purif. 46 (2006) 9.

[40] F. Gebauer, D.F. Corona, T. Preiss, P.B. Becker, M.W. Hentze, Translational control of dosage compensation in Drosophila by Sex-lethal: cooperative silencing via the 5’ and 3' UTRs of msl-2 mRNA is independent of the poly(A) tail., EMBO J. 18 (1999) 6146–54.

[41] A.W. Murray, Chapter 30 Cell Cycle Extracts, in: B.K.K. and H.B.P.B.T.-M. in C. Biology (Ed.), Xenopus Laevis Pract. Uses Cell Mol. Biol., Academic Press, 1991: pp. 581–605.

[42] T. Endoh, T. Kanai, Y.T. Sato, D. V Liu, K. Yoshikawa, H. Atomi, *et al.*, Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile., J. Biotechnol. 126 (2006) 186–95.

[43] K. Sitaraman, D.K. Chatterjee, High-Throughput Protein Expression Using Cell-Free System, in: High Throughput Protein Expr. Purif., Humana Press, 2009: pp. 229–244.

[44] F. Katzen, G. Chang, W. Kudlicki, The past, present and future of cell-free protein synthesis, Trends Biotechnol. 23 (2005) 150–156.

[45] H.-C. Chang, C.M. Kaiser, F.U. Hartl, J.M. Barral, De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria., J. Mol. Biol. 353 (2005) 397–409.

[46] P. Braun, J. LaBaer, High throughput protein production for functional proteomics, Trends Biotechnol. 21 (2003) 383–388.

[47] H. Ding, Molecular cloning, expression, purification, and characterization of soluble full-length, human interleukin-3 with a baculovirus–insect cell expression system, Protein Expr. Purif. 31 (2003) 34–41.

[48] Y. Endo, T. Sawasaki, Cell-free expression systems for eukaryotic protein production., Curr. Opin. Biotechnol. 17 (2006) 373–80.

[49] F. Katzen, W. Kudlicki, Efficient generation of insect-based cell-free translation extracts active in glycosylation and signal sequence processing, J. Biotechnol. 125 (2006) 194–197.

[50] T. Sawasaki, T. Ogasawara, R. Morishita, Y. Endo, A cell-free protein synthesis system for high-throughput proteomics, Natl. Acad. Sci. USA. 99 (2002) 14652–14657.

[51] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, *et al.*, Cell-free translation reconstituted with purified components., Nat. Biotechnol. 19 (2001) 751–5.

[52] J. Swartz, Developing cell-free biology for industrial applications., J. Ind. Microbiol. Biotechnol. 33 (2006) 476–85.

[53] J.F. Zawada, Cell-Free Production of Pharmaceutical Proteins, in: Compr. Biotechnol., Second Edi, Elsevier B.V., 2011: pp. 391–398.

[54] T.-W. Kim, J.-W. Keum, I.-S. Oh, C.-Y. Choi, C.-G. Park, D.-M. Kim, Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system, J. Biotechnol. 126 (2006) 554–561.

[55] J. Pratt, Coupled Transcription-Translation in Prokaryotic Cell-free Systems, in: Transcr. Transl. — A Pract. Approach, 1984: pp. 179–209.

[56] G. Zubay, *In vitro* synthesis of protein in microbial systems., Annu. Rev. Genet. 7 (1973) 267–87.

[57] D.-M. Kim, T. Kigawa, C.-Y. Choi, S. Yokoyama, A highly efficient cell-free protein synthesis system from *Escherichia coli*., Eur. J. Biochem. 239 (1996) 881–6.

[58] J.R. Swartz, M.C. Jewett, K. a Woodrow, Cell-free protein synthesis with prokaryotic combined transcription-translation., Methods Mol. Biol. 267 (2004) 169–82.

[59] T. Kigawa, T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, A. Tanaka, et al., Preparation of Escherichia coli cell extract for highly productive cell-free protein expression., J. Struct. Funct. Genomics. 5 (2004) 63–8.

[60] D. Schwarz, C. Klammt, A. Koglin, F. Löhr, B. Schneider, V. Dötsch, *et al.*, Preparative scale cell-free expression systems: new tools for the large scale preparation of integral membrane proteins for functional and structural studies., Methods. 41 (2007) 355–69.

[61] E. Seki, N. Matsuda, T. Kigawa, Multiple inhibitory factor removal from an *Escherichia coli* cell extract improves cell-free protein synthesis., J. Biosci. Bioeng. 108 (2009) 30–5.

[62] M.C. Jewett, A.M. Voloshin, J.R. Swartz, Prokaryotic systems for *in vitro* Expression., in: M.P. Weiner, Q. Lu (Eds.), Gene Cloning Expr. Technol., Eaton Publishing, Westborough, MA, 2002: pp. 391–411.

[63] E. Seki, N. Matsuda, S. Yokoyama, T. Kigawa, Cell-free protein synthesis system from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation, Anal. Biochem. 377 (2008) 156–161.

[64] D.-M. Kim, J.R. Swartz, Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of *Escherichia coli*., Biotechnol. Bioeng. 85 (2004) 122–9.

[65] D.-M. Kim, C.-Y. Choi, A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane., Biotechnol. Prog. 12 (1996) 645–9.

[66] H. Nakano, T. Shinbata, R. Okumura, S. Sekiguchi, M. Fujishiro, T. Yamane, Efficient coupled transcription/translation from PCR template by a hollow-fiber membrane bioreactor., Biotechnol. Bioeng. 64 (1999) 194–9.

[67] T. Sawasaki, Y. Hasegawa, M. Tsuchimochi, N. Kamura, T. Ogasawara, T. Kuroita, *et al.*, A bilayer cell-free protein synthesis system for high-throughput screening of gene products, FEBS Lett. 514 (2002) 102–105.

[68] D.-M. Kim, J.R. Swartz, Prolonging cell-free protein synthesis with a novel ATP regeneration system, Biotechnol. Bioeng. 66 (1999) 180–188.

[69] D.-M. Kim, J.R. Swartz, Prolonging cell-free protein synthesis by selective reagent additions, Biotechnol. Prog. 16 (2000) 385–90.

[70] D.-M. Kim, J.R. Swartz, Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis, Biotechnol. Bioeng. 74 (2001) 309–316.

[71] T.-W. Kim, J.-W. Keum, I.-S. Oh, C.-Y. Choi, H.-C. Kim, D.-M. Kim, An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source, J. Biotechnol. 130 (2007) 389–93.

[72] T.-W. Kim, H.-C. Kim, I.-S. Oh, D.-M. Kim, A highly efficient and economical cell-free protein synthesis system using the S12 extract of *Escherichia coli*, Biotechnol. Bioprocess Eng. 13 (2008) 464–469.

[73] T.-W. Kim, I.-S. Oh, J.-W. Keum, Y. Kwon, J. Byun, K.-H. Lee, *et al.*, Prolonged Cell-Free Protein Synthesis Using Dual Energy Sources: Combined Use of Creatine Phosphate and Glucose for the Efficient Supply of ATP and Retarded Accumulation of Phosphate, Biotechnol. Bioeng. 97 (2007) 1510–1515.

[74] H.-C. Kim, D.-M. Kim, Methods for energizing cell-free protein synthesis, J. Biosci. Bioeng. 108 (2009) 1–4.

[75] M.C. Jewett, J.R. Swartz, Rapid Expression and Purification of 100 nmol Quantities of Active Protein Using Cell-Free Protein Synthesis, Biotechnol. Prog. (2004) 102–109.

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

[77] F. Lottspeich, J.W. Engels, eds., Bioanalytik, 3rd ed., Spektrum Akademischer Verlag, 2012.

[78] S. Bettati, E. Pasqualetto, G. Lolli, B. Campanini, R. Battistutta, Structure and single crystal spectroscopy of Green Fluorescent Proteins., Biochim. Biophys. Acta. 538 (2010).

[79] T. Nakatsu, S. Ichiyama, J. Hiratake, A. Saldanha, N. Kobashi, K. Sakata, *et al.*, Structural basis for the spectral difference in luciferase bioluminescence, Nature. 440 (2006) 372–376.

[80] C. Langlais, B. Guilleaume, N. Wermke, T. Scheuermann, L. Ebert, J. LaBaer, *et al.*, A systematic approach for testing expression of human full-length proteins in cell-free expression systems., BMC Biotechnol. 7 (2007) 64.

[81] C.E. Hodgman, M.C. Jewett, Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis., Biotechnol. Bioeng. 110 (2013) 2643–54.

[82] V. V Zeenko, C. Wang, M. Majumder, A.A. Komar, M.D. Snider, W.C. Merrick, *et al.*, An efficient *in vitro* translation system from mammalian cells lacking the translational inhibition caused by eIF2 phosphorylation., RNA. 14 (2008) 593–602.

[83] H. Tarui, S. Imanishi, T. Hara, A novel cell-free translation/glycosylation system prepared from insect cells, J. Biosci. Bioeng. 90 (2000) 508–514.

[84] H. Tarui, M. Murata, I. Tani, S. Imanishi, S. Nishikawa, T. Hara, Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment, Appl. Microbiol. Biotechnol. 55 (2001) 446–453.

[85] Y. Sato, K. Aizawa, T. Ezure, E. Ando, N. Uozumi, A simple fed-batch method for transcription and insect cell-free translation., J. Biosci. Bioeng. 114 (2012) 677–679.

[86] S.P. Chambers, S.E. Swalley, High Throughput Protein Expression and Purification, Humana Press, Totowa, NJ, 2009.

[87] B.A. Sherf, K. V Wood, Firefly Luciferase Engineered for Improved Genetic Reporting, Promega Notes Mag. (1994) 14–21.

[88] J.L. Vaughn, R.H. Goodwin, G.J. Tompkins, A. Mccawley, V.E.T. Al, The Establishment of two Cell Lines from the Insect Spodoptera fruciperda (Lepidoptera ; Noctuidae), 13 (1977) 0–4.

[89] R.R. Granados, A.C.G. Derksen, K.G. Dwyer, Replication of the Trichoplusia ni Granulosis and Nuclear Polyhedrosis Viruses in Cell Cultures, J. Virol. 152 (1986) 472–476.

[90] V. V Gurevich, I.D. Pokrovskaya, T. a Obukhova, S. a Zozulya, Preparative *in vitro* mRNA synthesis using SP6 and T7 RNA polymerases, Anal. Biochem. 195 (1991) 207–213.

[91] D.C. Montgomery, Design and analysis of experiments, 8th ed., John Wiley & Sons, 2012.

[92] T. Ezure, S. Higashide, T. Suzuki, M. Ito, K. Endo, Preparation method of insect cell extract solution for cell-free protein synthesis, the insect cell extract solution and cell-free synthesis method of protein using the insect cell extract solution, US Patent 7323332, 2008.

[93] K. Treier, P. Lester, J. Hubbuch, Application of genetic algorithms and response surface analysis for the optimization of batch chromatographic systems, Biochem. Eng. J. 63 (2012) 66–75.

[94] P.S. Wierling, R. Bogumil, E. Knieps-gru, High-Throughput Screening of Packed-Bed Chromatography Coupled With SELDI-TOF MS Analysis : Monoclonal Antibodies Versus Host Cell Protein, 98 (2007) 440–450.

[95] M. Bensch, B. Selbach, J. Hubbuch, High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems, Chem. Eng. Sci. 62 (2007) 2011–2021.

[96] K. Treier, S. Hansen, C. Richter, P. Diederich, J. Hubbuch, P. Lester, High-throughput methods for miniaturization and automation of monoclonal antibody purification processes., Biotechnol. Prog. 28 (2012) 723–732.

[97] A.S. Spirin, J.R. Swartz, Cell-Free Protein Synthesis, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2007.

[98] J.R. Swartz, Cell-Free Protein Expression, Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2003.

[99] D.A.P. Krummel, K. Nagai, C. Oubridge, Structure of spliceosomal ribonucleoproteins, F1000 Biol. Rep. 2 (2010) 1–5.

[100] B. Combe, M. Rucheton, H. Graafland, V. Lussiez, C. Brunel, J. Sany, Clinical significance of anti-RNP and anti-Sm autoantibodies as determined by immunoblotting and immunoprecipitation in sera from patient with connective tissue disease, Clin. Exp. Immunol. 75 (1989) 18–24.

[101] H.J. Netter, H.H. Guldner, C. Szostecki, H.J. Lakomek, H. Will, A recombinant autoantigen derived from the human (U1) small nuclear RNP-specific 68-kd protein, Arthritis Rheum. 31 (1988) 616–622.

[102] H. Theissen, M. Etzerodt, R. Reuter, C. Schneider, F. Lottspeich, P. Argos, *et al.*, Cloning of the human cDNA for the U1 RNA-associated 70K protein, EMBO J. 5 (1986) 3209–3217.

[103] R. Spritz, K. Strunk, C. Surowy, S. Hoch, D. Bartion, U. Francke, The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding, Nucleic Acids Res. 15 (1987) 10373–10391.

[104] D.A. Pomeranz Krummel, C. Oubridge, A.K.W. Leung, J. Li, K. Nagai, Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution., Nature. 458 (2009) 475–80.

[105] B.R. Graveley, Sorting out the complexity of SR protein functions, RNA J. 6 (2000) 1197–1211.

[106] C.W. Smith, J. Valcárcel, Alternative pre-mRNA splicing: the logic of combinatorial control, Trends Biochem. Sci. 25 (2000) 381–388.

[107] M.M. Sander, G.H. Dixon, The Biosynthesis of Protamine in Trout Testis, IV. Sites of Phosphorylation, J. Biol. Chem. 247 (1972) 851–855.

[108] H.H. Guldner, H.J. Netter, C. Szostecki, H.J. Lakomek, H. Will, Epitope mapping with a recombinant human 68-kDa (U1) ribonucleoprotein antigen reveals heterogeneous autoantibody profiles in human autoimmune sera, J. Immunol. 141 (1988) 469–475.

[109] H.J. Netter, H.H. Guldner, C. Szostecki, H. Will, Major autoantigenic sites of the (U1) small nuclear ribonucleoprotein-specific 68-kDa protein, Scand. J. Immunol. 32 (1990) 163–176.

[110] D. Cram, N. Fisicaro, R. Coppel, S. Whittingham, L. Harrison, Mapping of multiple B cell epitopes on the 70-kilodalton autoantigen of the U1 ribonucleoprotein complex, J. Immunol. 145 (1990) 630–635.

[111] Y. Takeda, U. Nyman, A. Winkler, K.S. Wise, S.O. Hoch, I. Pettersson, *et al.*, Antigenic domains on the U1 small nuclear ribonucleoprotein-associated 70K polypeptide - A comparison of regions selectively recognized by human and mouse autoantibodies and by monoclonal antibodies, Clin. Immunol. Immunopathol. 61 (1991) 55–68.

[112] B. Frorath, M. Scanarini, H.J. Netter, C.C. Abney, B. Liedvogel, H.J. Lakomek, *et al.*, Cloning and expression of antigenic epitopes of the human 68-kDa (U1) ribonucleoprotein antigen in *Escherichia coli*, Biotechniques. 11 (1991) 364–371.

[113] H. Berthold, M. Scanarini, C.C. Abney, B. Frorath, W. Northemann, Purification of Recombinant Antigenic Epitopes of the Human 68-kDa (U1) Ribonucleoprotein Antigen Using the Expression System pH6EX3 Followed by Metal Chelating Affinity Chromatography, Protein Expr. Purif. (1992) 50–56.

[114] R.L. Nelissen, C.L. Will, W.J. van Venrooij, R. Lührmann, The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins, EMBO J. 13 (1994) 4113–4125.

[115] D. Busso, R. Kim, S.-H. Kim, Using an *Escherichia coli* cell-free extract to screen for soluble expression of recombinant proteins, J. Struct. Funct. Genomics. 5 (2004) 69–74.

[116] T. Lamla, S. Hoerer, M.M.T. Bauer, Screening for soluble expression constructs using cell-free protein synthesis, Int. J. Biol. Macromol. 39 (2006) 111–121.

[117] D. Busso, Expression of soluble recombinant proteins in a cell-free system using a 96-well format, J. Biochem. Biophys. Methods. 55 (2003) 233–240.

[118] H. Singh, S.-I. Makino, Y. Endo, Y. Li, A.N. Stephens, G. Nie, Application of the wheat-germ cell-free translation system to produce high temperature requirement A3 (HtrA3) proteases, Biotechniques. 52 (2012) 23–28.

[119] F.W. Studier, Protein production by auto-induction in high density shaking cultures, Protein Expr. Purif. 41 (2005) 207–234.

[120] J.F. Kane, Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*, 8 (1995) 494–500.

[121] U. Brinkmann, R.E. Mattes, P. Buckel, High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product, 85 (1989) 109–114.

[122] G.F. Chen, M. Inouye, Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes, Nucleic Acids Res. 18 (1990) 1465–1473.

[123] A.S. Spirin, J.R. Swartz, Cell-free Protein Synthesis Systems:Historical Landmarks, Classification, and General Methods, in: A.S. Spirin, J.R. Swartz (Eds.), Cell-Free Protein Synth., WILEY-VCH Verlag GmbH & Co.KGaA, Weinheim, 2008: pp. 1–34.

[124] Henaut, Danchin, Analysis and Predictions from *Escherichia coli* sequences, in: Neidhardt (Ed.), *Escherichia Coli* Salmonella Typhimurium Cell. Mol. Biol., American Society for Microbiology, 1996: pp. 2047–2066.

[125] F.U. Hartl, M. Hayer-Hartl, Molecular chaperones in the cytosol: from nascent chain to folded protein, Science. 295 (2002) 1852–1858.

[126] H.P. Sørensen, K.K. Mortensen, Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*, Microb. Cell Fact. 4 (2005) 1–8.

[127] A. Mogk, M.P. Mayer, E. Deuerling, Mechanisms of protein folding: molecular chaperones and their application in biotechnology, Chembiochem. 3 (2002) 807–14.

[128] B. Bukau, E. Deuerling, C. Pfund, E.A. Craig, Getting Newly Synthesized Proteins into Shape, Cell. 101 (2000) 119–122.

[129] A. de Crouy-Chanel, M. Kohiyama, G. Richarme, Specificity of DnaK for arginine/lysine and effect of DnaJ on the amino acid specificity of DnaK, J. Biol. Chem. 271 (1996) 15486–15490.

[130] W. Burgdorfer, A.G. Barbour, S.F. Hayes, J.L. Benach, E. Grunwaldt, J.P. Davis, Lyme disease-a tick-borne spirochetosis, Science. 216 (1982) 1317–1319.

[131] G. Margos, S.A. Vollmer, M. Cornet, M. Garnier, V. Fingerle, B. Wilske, *et al.*, A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes, Appl. Environ. Microbiol. 75 (2009) 5410–5416.

[132] G. Goettner, U. Schulte-Spechtel, R. Hillermann, G. Liegl, B. Wilske, V. Fingerle, Improvement of Lyme borreliosis serodiagnosis by a newly developed recombinant immunoglobulin G (IgG) and IgM line immunoblot assay and addition of VlsE and DbpA homologues, J. Clin. Microbiol. 43 (2005) 3602–3609.

[133] S.M. Engstrom, E. Shoop, R.C. Johnson, Immunoblot interpretation criteria for serodiagnosis of early Lyme disease, J. Clin. Microbiol. 33 (1995) 419–427.

[134] F. Dressler, J.A. Whalen, B.N. Reinhardt, A.C. Steere, Western Blotting in the Serodiagnosis of Lyme Disease, J. Infect. Dis. 167 (1993) 392–400.

[135] B.J.B. Johnson, Chapter 4: Laboratory Diagnostic Testing for *Borrelia* burgdorferi Infection, in: J.J. Halperin (Ed.), Lyme Dis. An Evidence-Based Approach (Advances Mol. Cell. Biol. Ser., CAB International, 2011: pp. 73–88.

[136] U. Hauser, G. Lehnert, R. Lobentanzer, B. Wilske, Interpretation criteria for standardized Western blots for three European species of *Borrelia* burgdorferi sensu lato, J. Clin. Microbiol. 35 (1997) 1433–1444.

[137] M. Becker, J. Bunikis, B.D. Lade, J.J. Dunn, A.G. Barbour, C.L. Lawson, Structural investigation of *Borrelia burgdorferi* OspB, a bactericidal Fab target, J. Biol. Chem. 280 (2005) 17363–17370.

[138] D.A. Jobe, S.D. Lovrich, R.F. Schell, S.M. Callister, C-Terminal Region of Outer Surface Protein C Binds Borreliacidal Antibodies in Sera from Patients with Lyme Disease, 10 (2003) 573–578.

[139] L.A. Magnarelli, M. Lawrenz, S.J. Norris, E. Fikrig, Comparative reactivity of human sera to recombinant VlsE and other *Borrelia* burgdorferi antigens in class-specific enzyme-linked immunosorbent assays for Lyme borreliosis, J. Med. Microbiol. 51 (2002) 649–655.

[140] R.P. Ekins, Multi-analyte immunoassay, J. Pharm. Biomed. Anal. 7 (1989) 155–168.

[141] A.C. Pease, D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, S.P.A. Fodor, Light-generated oligonucleotide arrays for rapid DNA sequence analysis, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 5022–5026.

[142] M. Schena, D. Shalon, R.W. Davis, P.O. Brown, Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, Science (80-.). 270 (1995) 467–470.

[143] J.L. DeRisi, V.R. Iyer, P.O. Brown, Exploring the metabolic and genetic control of gene expression on a genomic scale, Science. 278 (1997) 680–686.

[144] E.E. Schadt, S.A. Monks, T.A. Drake, A.J. Lusisk, N. Chek, V. Colinayok, *et al.*, Genetics of gene expression surveyed in maize, mouse and man, Nature. 422 (2003) 297–302.

[145] M. Morley, C.M. Molony, T.M. Weber, J.L. Devlin, K.G. Ewens, R.S. Spielman, *et al.*, Genetic analysis of genome-wide variation in human gene expression, Nature. 430 (2004) 743–747.

[146] G. MacBeath, Protein microarrays and proteomics, Nat. Genet. 32 Suppl (2002) 526–532.

[147] M.F. Templin, D. Stoll, J.M. Schwenk, O. Pötz, S. Kramer, T.O. Joos, Protein microarrays: promising tools for proteomic research, Proteomics. 3 (2003) 2155–66.

[148] H. Chandra, P.J. Reddy, S. Srivastava, Protein microarrays and novel detection platforms, Expert Rev. Proteomics. 8 (2011) 61–79.

[149] H. Chandra, S. Srivastava, Cell-free synthesis-based protein microarrays and their applications, Proteomics. 10 (2009) 1–14.

[150] D.P. Kreil, R.R. Russell, There is no silver bullet--a guide to low-level data transforms and normalisation methods for microarray data., Brief. Bioinform. 6 (2005) 86–97.

[151] N. Nath, R. Hurst, B. Hook, P. Meisenheimer, K.Q. Zhao, N. Nassif, *et al.*, Improving protein array performance: focus on washing and storage conditions., J. Proteome Res. 7 (2008) 4475–82.

7 Appendix

Supporting information to Research Article

Soluble full-length expression and characterization of snRNP protein U1-68/70 K

Carolin Richter, Thomas Simon, Iris Asen, Gerald Brenner-Weiss, Jürgen Hubbuch

Correspondence: Prof. Jürgen Hubbuch (Juergen.Hubbuch@kit.edu), Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany.

```
                  (His-tag)
                   -8        1       10         20         30         40
U1-68/70 K          (MHHHHHHA) MTQFLPPNLL ALFAPRDPIP YLPPLEKLPH EKHHNQPYCG
U1-68/70 K_truncated (MHHHHHHA) MTQFLPPNLL ALFAPRDPIP YLPPLEKLPH EKHHNQPYCG
                             ********** ********** ********** **********
                            50         60          70         80        90
U1-68/70 K          IAPYIREFED PRDAPPPTRA ETREERMERK RREKIERRQQ EVETELKMWD
U1-68/70 K_truncated IAPYIREFED PRDAPPPTRA ETREERMERK RREKIERRQQ EVETELKMWD
                    ********** ********** ********** ********** **********
                          100        110        120        130        140
U1-68/70 K          PHNDPNAQGD AFKTLFVARV NYDTTESKLR REFEVYGPIK RIHMVYSKRS
U1-68/70 K_truncated PHNDPNAQGD AFKTLFVARV NYDTTESKLR REFEVYGPIK RIHMVYSKRS
                    ********** ********** ********** ********** **********
                          150        160        170        180        190
U1-68/70 K          GKPRGYAFIE YEHERDMHSA YKHADGKKID GRRVLVDVER GRTVKGWRPR
U1-68/70 K_truncated GKPRGYAFIE YEHERDMHSA YKHADGKKID GRRVLVDVER GRTVKGWRPR
                    ********** ********** ********** ********** **********
                          200        210        220        230        240
U1-68/70 K          RLGGGLGGTR RGGADVNIRH SGRDDTSRYD ERDRDRDRER ERRERSRERD
U1-68/70 K_truncated RLGGGLGGTR RGGADVNIRH SGRDDTSRYD ERDRDRDRER ERRERSRERD
                    **************************************************
                                                 Inhibitory sequence X [16]
                          250        260        270        280        290
U1-68/70 K          KERERRRSRS RDRRRRSRSR DKEERRRSRE RSKDKDRDRK RRSSRSRERA
U1-68/70 K_truncated KERERRRSRS RDRRRRSRSR ---------- ---------- ----------
                    ********** **********
                          300        310        320        330        340
U1-68/70 K          RRERERKEEL RGGGGDMAEP SEAGDAPPDD GPPGELGPDG PDGPEEKGRD
U1-68/70 K_truncated ---------- ---------- ---------- ----- GPDG PDGPEEKGRD
                                                        **** **********
```

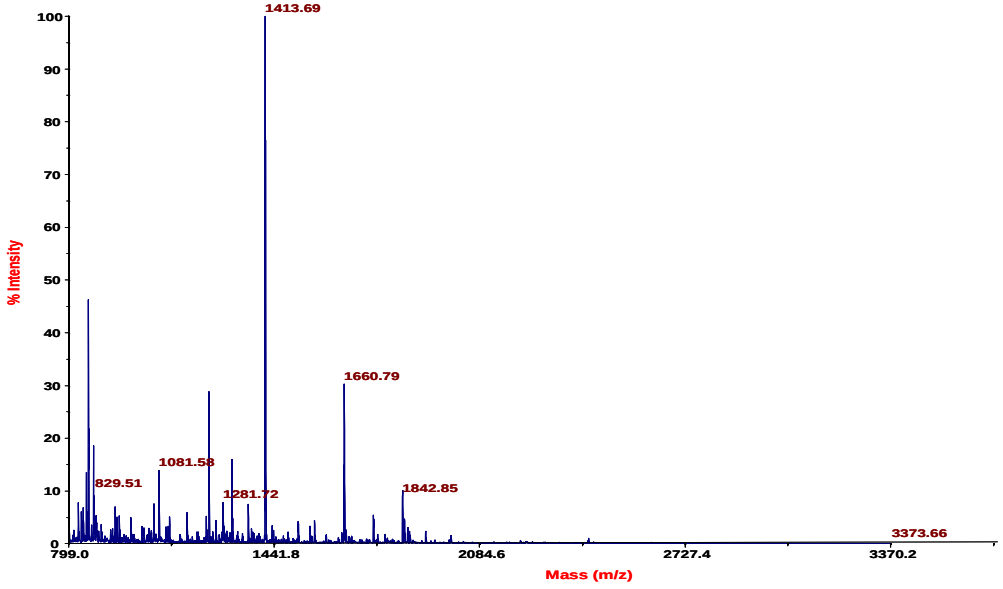
```
                            350        360        370        380        390
U1-68/70 K            RDRERRRSHR SERERRRDRD RDRDRDREHK RGERGSERGR DEARGGGGGQ
U1-68/70 K_truncated RDRERRRSHR SERERRRDRD RDRDRDREHK RGERGSERGR DEARGGGGGQ
                     ********** ********** ********** ********** **********

                                                          His-tag
                            400        410        420      428
U1-68/70 K            DNGLEGLGND SRDMYMESEG GDGYLAPENG YLIEAAPEHH HHHH
U1-68/70 K_truncated DNGLEGLGND SRDMYMESEG GDGYLAPENG YLMEAAPEHH HHHH
                     ********** ********** ********** ********** ****
```

Figure A.1: Comparison of the amino acid sequences of C-terminally His-tagged full-length and truncated U1-68/70 K.

Table 1: MS-determined peptide masses of the U1-68/70 K full-length protein. The four main fragments marked with (*) were confirmed by MS/MS analysis.

Mass	Position	Peptide sequence
3373.3920	302-335	GGGGDMAEPSEAGDAPPDDGPPGE LGPDGPDGPEEK*
1842.7864	88-103	MWDPHNDPNAQGDAFK*
1659.7917	385-402	GGGGGQDNGLEGLGNDSR*
1413.6433	145-155	GYAFIEYEHER*
1281.7089	17-27	DPIPYLPPLEK
1103.5579	79-87	QEVETELK
1081.5564	122-130	EFEVYGPIK
1056.4844	110-118	VNYDTTESK
829.4778	174-180	VLVDVER

Figure A.2: Peptide mass fingerprint of U1-68/70 K full-length protein.



Supporting information to Research Article

Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera

Carolin Richter, Kosta Konstantinidis, Iris Asen, Richard Kneusel, Jürgen Hubbuch

Correspondence: Prof. Jürgen Hubbuch (Juergen.Hubbuch@kit.edu), Karlsruhe Institute of Technology, Institute of Process Engineering in Life Sciences, Section IV: Biomolecular Separation Engineering, Engler-Bunte-Ring 1, 76131 Karlsruhe, Germany.

Table 1: ANOVA p-values of on-print controls.

on-print control	p-value
Human serum albumin	0.29
Escherichia coli lysate	0.00
IgG	0.37
IgM	0.92

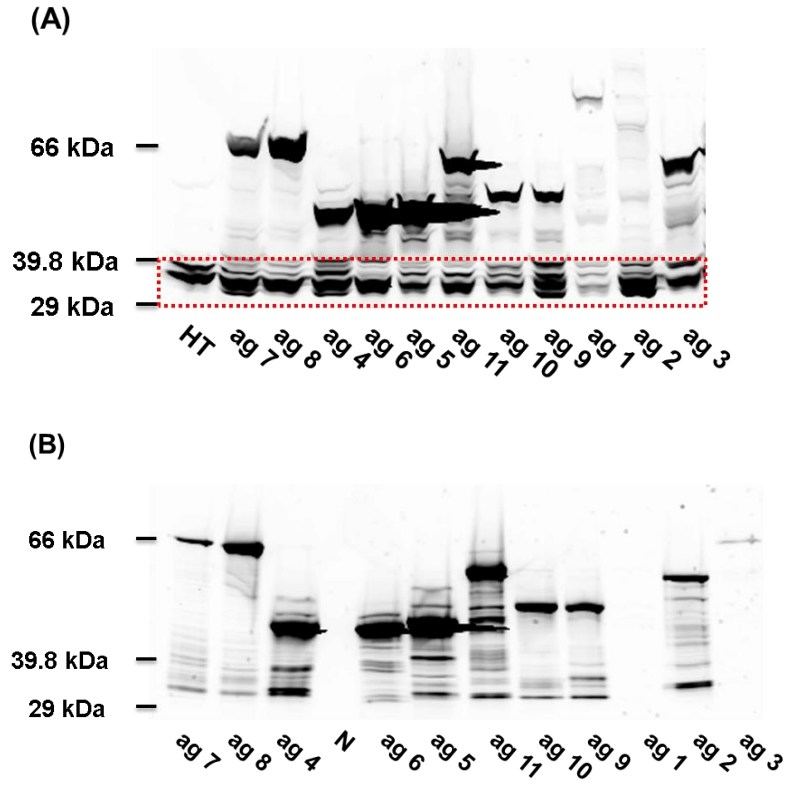


Figure 1: SDS-Page analysis of eleven *in vitro* expressed *Borrelia* HaloTag® fusion antigens (A) unpurified and (B) Ni-CS purified; detected with the HaloTag® TMR ligand, measured in a fluorescence scanner using the Cy2 channel. HT: HaloTag® protein, N: negative control (*E. coli* lysate). The HaloTag® (32 kDa) was expressed in relatively high amounts (red dotted line). (Numbering according to Table 1 in the research article)

8 Curriculum Vitae

Professional Experience

09/2013 Management Trainee at Pfizer Manufacturing, Freiburg im Breisgau

---

07/2012 – 12/2012 PhD student at Karlsruhe Institute of Technology

---

07/2009 – 06/2012 PhD student at DIARECT AG, Freiburg

10/2006 – 02/2007 Internship at Sanofi Aventis, Frankfurt am Main

---

Education

10/2003 – 06/2009 Life Science Engineering at Karlsruhe Institute of Technology

Majors: Technical Biology, Food Process Engineering

Graduated as Diplom-Engineer with a grade of 1.6

11/2008 – 06/2009 Diploma thesis at Karlsruhe Institute of Technology

Institute of Process Engineering in Life Sciences

Section IV: Biomolecular Separation Engineering (Prof. Jürgen Hubbuch)

„Establishment of an industrial monoclonal antibody purification process on a robotic workstation”

04/2007 – 08/2007 Undergraduate thesis at Macquarie University, Sydney

Department of Biological Sciences,

Division of Environmental and Life Science

„Morphology and molecular analysis of *Eimeria* species from

the tammar wallaby (*Macropus eugenii*)“

9 Publications

Papers

Michelle L Power, Carolin Richter, Samantha Emery, Jasmine Hufshmid and Ian Beveridge; “Phylogeny of *Eimeria trichosuris*, a marsupial coccidium, based on 18S rDNA sequences” (Experimental Parasitology 122 (2009) 165-168)

Nichola J Hill, Carolin Richter, Michelle L Power; “Pinning down a polymorphic parasite: New genetic and morphological descriptions of *Eimeria macropodis* from the Tammar wallaby (*Macropus eugenii*).” (Parasitology International 61 (2012) 461–465)

Katrin Treier, Sigrid Hansen, Carolin Richter, Patrick Diederich, Jürgen Hubbuch, Philip Lester; „High-throughput methods for miniaturization and automation of monoclonal antibody purification processes.” (Biotechnol. Prog., 2012, Vol. 28, No. 3, 723-732)

Carolin Richter, Kosta Konstantinidis, Richard Kneusel, Jürgen Hubbuch; „Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera“ (Life Science Engineering, 2014, DOI: 10.1002/elsc.201300109)

Carolin Richter, Fabian Bickel, Anna Osberghaus, Jürgen Hubbuch; „High-throughput characterization of an Insect cell-free expression” (submitted to Life Science Engineering)

Carolin Richter, Thomas Simon, Iris Asen, Gerald Brenner-Weiss, Jürgen Hubbuch; “Soluble full-length expression and characterization of snRNP protein U1-68/70 K” (submitted to Protein Expression and Purification)

Presentations

Carolin Richter, Kosta Konstantinidis, Richard Kneusel, Jürgen Hubbuch; „Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera“ (“International Workshop on New and Synthetic Bioproduction Systems”, Dezember 2012, Hamburg-Harburg)

Carolin Richter, Kosta Konstantinidis, Richard Kneusel, Jürgen Hubbuch; „Cell-free expression of recombinant antigens of *Borrelia burgdorferi* and microarray-based multiplex detection using different patient sera“ (“Cell-free Protein Synthesis – 3rd Public Status Seminar”, Januar 2013, Fraunhofer Institut Potsdam)