PHD THESIS

DEVELOPMENT AND APPLICATION OF IN VITRO TRANSLATION VECTORS

VIOLA BARDÓCZY

SUPERVISOR:
DR. TAMÁS MÉSZÁROS
SEMMELWEIS UNIVERSITY
DEPARTMENT OF MEDICAL CHEMISTRY, MOLECULAR BIOLOGY AND PATHOBIOCHEMISTRY

BME DEPARTMENT OF APPLIED BIOTECHNOLOGY AND FOOD SCIENCES

2009

1. Preface

The last decade has brought the renaissance of protein studies and accelerated development of high-throughput methods in all aspects of proteomics. Most protein synthesis systems exploit the capacity of living cells to translate proteins, but their application is limited by several factors. A more flexible alternative method for protein production is cell-free *in vitro* protein translation. Currently there are various *in vitro* translation systems available for high throughput protein production. Among these the wheat germ extract based in vitro translation system seems to be most promising, as it is suitable to synthesize numerous eukaryotic proteins cost-efficiently and in their native folded form. Although currently available vectors for wheat embryo in vitro translation systems ensure high productivity, they do not meet all requirements of state-of-the-art proteomics. Target genes have to be inserted using restriction endonucleases and the plasmids do not encode cleavable affinity purification tags.

The main purpose of my PhD thesis is to design a new family of *in vitro* translational vectors. This should allow easy and efficient cloning of any DNA sequence, its *in vitro* translation and facilitate purification of the resulting protein. To evaluate our improved vectors, a plant mitogen activated protein kinase, AtMPK6 will be cloned and purified.

2. Scientific background

In the last decade, attention focused on functionality and structure of proteins. Accelerated proteomics studies demand high-throughput protein production methods to ensure availability of proteins of interest. Presently, overexpression in *E. coli* cells is the most preferred protein production method. Though this system has been well optimized and is suitable for the simultaneous generation of a panel of proteins, its application is often limited by the insolubility of synthesized eukaryotic proteins. Although different *E. coli* strains and various protein and peptide fusion partners have been developed to increase the solubility of heterologous proteins, these methods are not universal and have to be optimized individually for efficient protein production. Recently, *in vitro* protein translation has emerged as an alternative to cell-based protein synthesis methods. The robustness of the translation apparatus is known since the fifties, and latest technical improvements made cellfree translation systems that approach the efficiency of cell-based systems. Various sources of translation machinery can be used for cell-free *in vitro* translation systems, like *E.coli,* rabbit reticulocyte, and wheat-germ extracts.

Due to its low cost and capacity for synthesizing properly folded, high molecular weight eukaryotic proteins, wheat germ derived protein extract presently seems the most promising choice. The growing popularity of cell-free systems can be explained as many different proteins can be produced within a short period of time, and the reaction parameters can be easily changed within wider limits. The produced proteins can be easily tagged, and modificated.

The procaryotic and eukaryotic translation, its components and its mechanism is a well-known process. The basic components and mechanisms are similar in both cases, as the place of the biosynthesis is the ribosome, tRNAs transport the amino acids to the ribosomes, amino-acid tRNA synthethases provide the connection between tRNAs and amino acids, and addition of other proteins as initation, elongation and termination factors are necessary. Despite the similarities there are many differences between the translation-mechanism of the two organisms, and these differences have practical importance from the aspect of implementation of cell-free protein biosynthesis. The most important difference is between the mRNA molecules appropriate for *in vitro* translation. In prokaryotic cells the transcribed mRNA is not modificated, the transcription and translation goes on simultaneously. From practical aspects it is an advantage, that the DNA vectors used for in vivo *E. coli* expression can be applied in *in vitro E. coli* systems as well. The eukaryotic mRNA is post-transcriptionally modificated, from the aspect of translation, the 5-methyl-guanosine cap and 3' polyadenosine sequence has a great importance, as without these elements the efficiency of translation dramatically decreases. As a result of post-transcriptional modifications, special DNA templates has to be applied in eukaryotic *in vitro* translation systems, to efficiently substitute these elements on the produced mRNA.

Another difference between prokaryotic and eukaryotic systems is based on the localization of transcription and translation within the cells. The two process goes on simultaneously in prokaryotic cells, but in eukaryotes the transcription takes place in the nucleus, and the translation in the cytoplasm. Considering the different ion environment of the nucleus and the cytoplasm, the optimal reaction conditions are different for transcription and translation, and so, the coupling of the two processes is a great challenge.

Recently the most *in vitro* translation systems are based on crude cell extracts, which contain the ribosomes, soluble enzymes, translation factors and tRNA-s. For successful *in vitro* translation, polynucleotide and protein components, and other chemical compounds are necessary. Besides the proper pH, ion concentration and reductive environment amino acids are necessary as well. In most cases the efficiency of translation is improved by the addtion of exogenous tRNAs to the reaction.

The systems detailed above are „batch" type systems, which means that the translation is carried out in a fixed volume reaction vessel. Their application is limited by the short time activity and the low protein yield. In these systems, the translation terminates after 20-60 minutes of incubation, as at this time point, the energy components are run out, and products or by-products reach such critical concentration which inhibit the further reaction. The continous-feed systems eliminate the problems of the classical batch reaction. This improvement radically raised the length of the translation (up two 2 weeks) and protein yield (up to 10 mg/ml in the translation mixture).

The continuous cell free systems can be summarized by the following statements: the substrates (amino acids, energy components) are continuously provided and the products and by-products are removed from the reaction. The continuously fed reactions can be carried out in three different forms. In the CFCF (continuous-flow cell-free) system, a feeding solution containing substrates is pumped into the reaction vessel, where the other components like mRNA, tRNA, translation factors are present, and the vessel is divided into two parts by an ultra-filtration membrane. The membrane retains the higher molecular weight molecules, while the small molecules can pass through the membrane. As a result of the continuous flow, the by-products and products flow out of the reaction vessel.

The CECF system is similar to CFCF from many aspects, the main difference is that instead of ultra-filtration membrane it uses dialysis membrane, and the exhange of the components is carried out by passive diffusion. The chambers containing the feeding solution and the reaction mixture are separated by the membrane. During the reaction, the by-products and substrates exchange place and the products accumulates in the reaction chamber containing the crude cell extract.

The bilayer reaction can be carried out in the simplest way, and the reaction goes on ten times longer than the classical 'batch' reaction. It's protein yield is up to 1 mg/ml.
The higher density translation mixture is carefully pipetted to the bottom of the reaction vessel. During the incubation, the components are difussing through the boundary layer, and at the end of the reaction, a homogenous mixture is the result.

The wheat germ based *in vitro* translation system used in our laboratory was first described by Robertis. In the past three decades, as a result of impovements, the efficiency of the original system became higher. The preparation process of wheat germ extract and the proper plasmids and PCR primers used for *in vitro* transcription were optimized. It was discovered recently, that translation inhibitors like tritin, tionin, ribonucleases, and proteases are present in

the wheat germ extract, mainly originating from the endosperm. With the removal of endosperm, the efficiency of translation becomes higher. It can be further improved by intensive washing of wheat germ embryos before extraction, and the extracted wheat germ based *in vitro* translation apparatus can be stored in frozen or liophilysed state for years without losing activity.

Unlike prokaryotic mRNA, eukaryotic mRNA has to be extensively modified to be an effective translation template. The 5'-cap is essential to translation initiation and has to be introduced to in vitro transcribed mRNAs using RNA polymerase, which incorporates the three modified nucleotides (7-mG-5_-ppp-5_G). The efficiency of incorporation is low, and the excess of free modified nucleotides remaining in the mix dramatically decreases the productivity of translation. The 3'-end poly(A) tail of eukaryotic mRNAs also presents a technical difficulty during *in vitro* translation template preparation, as long polyA/T sequences of plasmids are unstable in host cells. To solve these problems, wheat germ *in vitro* translation vectors have been constructed with a special sequence replacing the cap. In the optimized vectors, the cap structure is substituted by either the tobacco mosaic virus translational enhancer Ω sequence with an additional GAA triplet at the 5'-end (GAA Ω), or an artificial 73 nucleotides containing a leader sequence. The same laboratory also examined the requirements for a poly(A) tail, and found that translation did not depend on the sequence but only on the length of 3'-UTR. An additional benefit of these plasmids is that the produced mRNAs were effective in vitro translation templates in a wider range of concentration than in vitro capped mRNAs.

Although the optimized vectors improved the productivity of in vitro translation, in order to build high-throughput protein synthesis systems, every step of the procedure must be accelerated, including the cloning of target genes and the purification of translated proteins. Ligation independent cloning (LIC) was developed to facilitate complex cloning and subcloning strategies, and have been applied by many laboratories since then. LIC overcomes important limitations of traditional cloning technologies, since any PCR product can be cloned into LIC compatible vectors without using restriction endonucleases and ligation. The LIC method takes advantage of the 3' exonuclease activity of T4 DNA polymerase to create complementary 12- to 15- nucleotide overhangs in the vector and PCR product. Upon transformation into *E. coli* cells, the host repair enzymes ligate at the vector-insert junction; thus, LIC produces high cloning efficiency with minimal non-recombinant background. A serious bottleneck of high-throughput protein production is the fast and high level purification of target proteins. Generally, the purification step is facilitated by addition of affinity tags to the N- or C- terminus of synthesized proteins. Although the affinity tags aid the purification, it might in many cases alter the *in vivo*

function and structure of proteins; hence, it must be removed by site specific proteases. The Tobacco Etch Virus (TEV) protease is an ideal choice because it cleaves with high specificity at a seven-amino-acid recognition sequence. Furthermore, it is active under a wide range of conditions, such as low temperature and high ionic concentration, and is only mildly sensitive to many protease inhibitors which are used to prevent protein degradation by host proteases.

3. Experiments and results

To improve the applicability of the *in vitro* translation vectors new vector constructions were developed on the basis of the *pEU-E01* and *pEU3-NII* backbone vectors. As a first step the *SspI⁻ pEU-3NII and pEU-E01* plasmids were constructed by site-directed mutagenesis of the *SspI* cleavage site of the original vectors. This step was necessary, because the basis of the further ligation independent cloning is the restriction digestion of the LIC site by *SspI* blunt-end restriction endonuclease.

In the next step a *GST-TEV-LIC* sequence *(Figure 1.)* was inserted into the constructed vectors. The *GST-TEV-LIC* sequence has a great importance from three aspects: 1.) By using the LIC site, the ligation independent cloning of any PCR product with the appropriate LIC overhangs can be implemented. 2.) The GST motif provides the N-terminal tag of the protein, and the possibility of affinity purification on solid phase with reduced glutathione. 3.) Following affinity purification, the N-terminal tag can be cleaved by digestion of the protein by TEV protease. The recognition sequence of TEV protease is *Glu-Asn-Leu-Tyr-Phe-Gln-Gly.*

To construct the *GST-TEV-LIC* sequence the expression vector *pGEX-2T* containing GST motif were used as a template. The sequence was amplified by using the appropriate PCR primers designed in our laboratory. After clean-up of the PCR product, it was digested by *BamHI* restriction endonuclease and ligated into *EcoRV-BamHI digested pEU-3NII* and *pEU-E01 SspI—* vectors. Competent cells were transformed with the ligation reaction and the transformation reaction was spread on LB ampicillin plates. The presence of the DNA fragment was ensured by colony PCR on the grown bacterial colonies. Plasmids were isolated from the positive colonies and the *GST-TEV-LIC* insertion site was sequenced. The expected, correct DNA sequences were obtained.

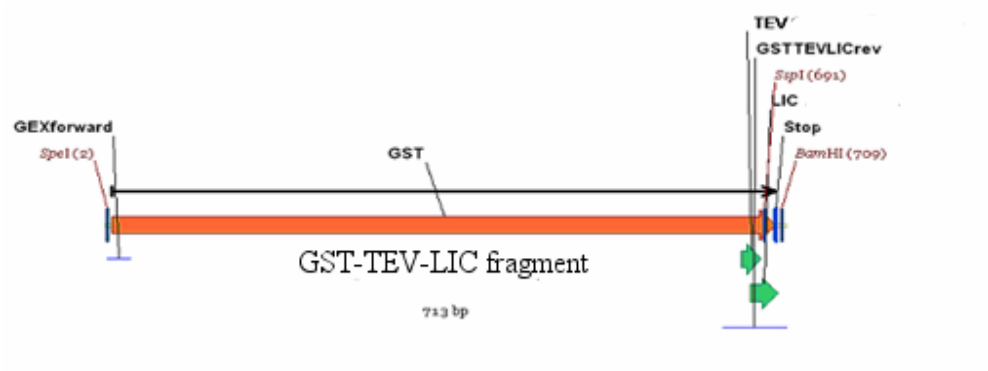

Figure 1: The GST-TEV-LIC sequence

In the next step the aim was to prove the applicability of the constructed *pEU-3NII* and *pEU-E01 GST-TEV-LIC in vitro* translation vectors. The cDNA of the *At*MPK6 protein kinase was inserted into the vectors by ligation independent cloning. The cDNA of the protein was amplified from an *At*MPK6 Sheen vector by primers containing the appropriate LIC overhangs. *E.coli* DH5α competent cells were transformed with the ligation independent cloning reaction mixture, and the colonies grown on LB ampicillin plate were tested for the presence of the DNA fragment by colony PCR. The PCR product was analysed on agarose gel, and according to the result of the analysis, the correct-size DNA fragment was present.

mRNA templates were produced by either T7 or SP6 bacteriophage RNA polymerases from *pEU3-NII-GST-TEV-MPK6*, and *pEU-E01-GST-TEV-MPK6* by *in vitro* transcription, which harbor T7 and SP6 promoters, respectively. After agarose gel analysis of the quality and the adequate quantity of the produced mRNA products, these were used as templates in *in vitro* translation reactions.

Protein biosynthesis was carried out by using bilayer *in vitro* translation system. The synthesized proteins were analysed on 12% polyacrylamide gel, and detected by Coomassie-blue staining. The GST-MPK6 protein can be well distinguished from the proteins of the wheat germ control. The produced proteins were detected by Western-blot, using GST-specific antibody. The Western-blot reaction proved obviously, that the GST-MPK6 protein was present in the translation mixture.

To test the funcionality of the GST tag motifs of our vectors, the GST tagged proteins were purified by using reduced glutathione containing solid phase resin. The non-specific binders were removed by washing, then the GST-MPK6 protein was eluted with buffer containing reduced glutathione. The experiments produced high purity MPK6 protein in one affinity purification

step, showing that fusion of the protein and the GST tag coded by the vectors simplifies the isolation of the synthesized proteins.

In the further experimental steps applicability of the TEV recognition sequence coded in the vectors was examined. The eluted GST-MPK6 protein was digested with TEV protease, removing with this step the N-terminal GST tag. Considering these results, it can be concluded that the protease cleavage site coded in our vectors function properly, in accord with our expectations.

Having shown that the designed vectors are indeed useful for *in vitro* protein synthesis, the translation process was optimised by varying several parameters. According to some bibliographic references the efficieny of *in vitro* translation can be increased using linearization of the DNA vector template. In contrast to data in the literature, comparative study of GST-MPK6 protein synthesis indicated that translation proceeds with approximately identical efficiency using cyclic and linearized DNA. These results suggest that linearization of vectors is not necessary; so this expensive and lenghty procedure can be avoided.

The next step in the optimization process was the evaluation, if the GST tag could be substituted by a much simpler histidine tag, both of which allow affinity purification. The new DNA casette (*His-TEV-LIC)* contains 6 Histidine signals, which is much more shorter, than the GST. As this sequence is very short, it doesn't alter the properties of the protein so much, so the cleavage of this is not inevitably necessary. For testing this hypothesis *pEU-3-NII-His-TEV-AtMPK6* and *pEU-E01-His-TEV-AtMPK6* vectors were created. First the *His-TEV-LIC* casette was created. After designing the adequate DNA sequences, the two strands of the casette were synthesized, and the hybridization reaction was carried out using appropriate reaction conditions.

The so produced double-stranded DNA was digested with *SpeI* and *BamHI* enzymes, and inserted into the *pEU-3NII-GST-TEV-AtMPK6* and *pEU-E01-GST-TEV-AtMPK6* vectors treated with the same restriction endonucleases. The ligation reaction was transformed into competent cells, and the presence of the fragments was verified by *SspI* restriction endonuclease treatment of the resulted colony PCR product. Plasmids were isolated from the positive colonies, and the sequence of the construction was verified by sequencing of the inserts. From the His-TEV-LIC vectors, with the methods detailed above, we constructed the *pEU-3NII-His-TEV-AtMPK6* and *pEU-E01-HIS-TEV-AtMPK6* plasmids. Transcription and translation was carried out on the vector templates, and the protein marking of the *in vitro* translation reaction was analysed by polyacrylamide gel electrophoresis. Similarly to the GST constructs, the produced *At*MPK6 proteins can be detected by Coomassie-staining. The practical utilization of the His tag was

examined with the affinity purification of the kinase proteins on nickel-ion containing solid phase. High purity proteins were obtained by imidazole-containing buffer elution from the affinity–purification resin. According to these results, the cleavage of the protein with TEV protease from the affinity purification resin results in higher purity proteins.

In order to compare the *in vitro* kinase activities of proteins produced in *E.coli* and in a cell-free system, the His tagged AtMPK6 cassette was transferred to a pET11a prokaryotic expression vector. The kinase was isolated by metal chelate affinity chromatography either from *E. coli* or from the *in vitro* translation reaction mixture. Equal amounts of His-AtMPK6 were used to determine the in vitro kinase activities of overexpressed and translated proteins using myelin basic protein (MBP) as a substrate. According to autoradiography results, phospholabelling of MBP was hardly detectable when bacterially overexpressed His-*At*MPK6 was tested, while the *in vitro* translated protein kinase displayed a clearly visible activity. The high kinase activity of translated *At*MPK6 indicates proper folding of the kinase domain, although posttranslational modification(s) of the translated protein are also likely to be responsible for elevated kinase activity; indeed, proper phosphorylation of MAP kinases is essential to gain their full kinase activity and kinases present in wheat germ extract could perform these phosphorylations.

After testing our vector constuctions we proved the wide-range applicability of our *pEU-E01-His-TEV-LIC* vector with the production of four poultry disease related virus proteins of the *La Sota* and *NDVDE* virus strains. The cDNAs of the proteins were amplified from cDNA library with ligation independent cloning compatible ends, and then inserted into the vector. The vectors coding the proper proteins were used as templates for in vitro transcription and translation, as detailed above, and high purity proteins were isolated from the reactions. The produced proteins met the requirements of aptamer selection from the aspects of quantity and quality as well.

4. Summary

In the course of my PhD studies I have constructed a set of four vectors to facilitate cloning and purification of *in vitro* protein translation systems based on wheat germ extract. The presented vectors eliminate the traditional cloning steps and aid the purification of translated proteins by incorporation of a LIC site and a TEV cleavable affinity tag, respectively. Success of using the new vectors was shown by synthesis and purification of a plant mitogen activated protein kinase. These vectors allow fast and parallel cloning and protein purification, and hence represent useful molecular tools for *in vitro* translation of eukaryotic proteins not only for our laboratory, but for other protein-producing facilities as well.

5. Thesis

1. *GST-TEV-LIC* and *HIS-TEV-LIC DNA* sequences were designed, which improve properties of the conventional *pEU-E01* and *pEU-3NII in vitro* translation vectors. Using a variety of molecular biological methods the *pEU-3NII-GST-TEV-LIC and pEU-E01-GST-TEV-LIC*, *pEU-3NII-His-TEV-LIC* and *pEU-E01-His-TEV-LIC in vitro* translation vectors were produced. [1]

2. Ligation independent cloning primers were designed for cloning the *At*MPK6 protein. This important plant protein responsible for signal transduction was successfully amplified using an *AtMPK6 Sheen* vector as a template [3].

3. The cDNA of the *At*MPK6 protein was successfully ligated into the constructed *in vitro* translation vectors. The so constructed plasmids were used as templates for *in vitro* transcription, and the resulted mRNAs were used as templates for *in vitro* translation. The AtMPK6 protein was present in all of the translation reactions, and the protein was successfully purified using the appropriate protein affinity purification tags [1].

4. It was shown that a circular DNA template can be as efficient template for *in vitro* translation as a linear DNA template [1].

5. Using an *in vitro* kinase reaction it was shown that *At*MPK6 protein produced by *in vitro* translation shows higher kinase activity than the protein produced in a prokaryotic *in vivo* system [1] [2].

6. To show the wide applicability of the vectors discussed above four viral proteins were cloned into the constructed vectors by ligation independent cloning. The DNA templates were used in transcripction, and then the produced mRNAs in *in vitro* translation. The produced proteins were successfully affinity purified and stored for further aptamer selection experiments [4].

6. Publications related to the thesis:

1. Bardoczy,V., V. Geczi, T. Sawasaki, Y. Endo, and T. Meszaros. 2008. A set of ligation-independent in vitro translation vectors for eukaryotic protein production. *BMC Biotechnol* 8:32. IF: 2,36; I:7

2. Sonkoly, B., V. Bardoczy and T. Meszaros (2009) Expression and Purification of Active Protein Kinases from Wheat Germ Extracts *Methods in Molecular Biology, Plant Kinase Protocols (Schnittger, A., ed.) Humana Press (accepted) Book chapter*

3. Meszaros, T., A. Helfer, E. Hatzimasoura, Z. Magyar, L. Serazetdinova, G. Rios, V. Bardóczy, M. Teige, C. Koncz, S. Peck, and L. Bögre. 2006. The Arabidopsis MAP kinase kinase MKK1 participates in defence responses to the bacterial elicitor flagellin. *Plant J* 48:485-498. IF: 6,49 I:

4. Bardoczy,V., Meszaros,T. 2009 Aptamerek-az antitestek lehetséges alternatívái. *Élelmiszervizsgálati Közlemények* 2: 105. IF: 0,02

Other publication:

5. Aranyi, T., M. Ratajewski, V. Bardoczy, L. Pulaski, A. Bors, A. Tordai, and A. Varadi. 2005. Identification of a DNA methylation-dependent activator sequence in the pseudoxanthoma elasticum gene, ABCC6. *J Biol Chem* 280:18643-18650. IF: 5,85

Oral lectures

1. Bardóczy Viola: Bioanalitikai vizsgálatok rekombináns fehérjék jellemzésére. *2008 Május 22 –Magyar Kémiai Egyesület Richter Gyári Csoport*

2. Bardóczy Viola, Mészáros Tamás: *In-vitro* transzlációs vektorok fejlesztése. XIII. Nemzetközi Vegyészkonferencia, *2007 November 9- Kolozsvár*

3. Bardóczy Viola, Mészáros Tamás: *In-vitro* transzlációs vektorok fejlesztése *2007 - Az Oláh György Doktori Iskola konferenciája*

4. Bardóczy Viola, Mészáros Tamás: Aptamerek szelekciója makromolekulákra és kismolekulákra, *2006-Az Oláh György Doktori Iskola konferenciája*

Posters, abstracts

1. Gergely Lautner, Júlia Szűcs, Róbert E.Gyurcsányi, Zsófia Balogh, Viola Bardóczy, Tamás Mészáros, Beata Komorowska: Selective detection of of plant virus coat proteins by aptamer based biochips, International Conference on Electrochemical Sensors, *Dobogókő, 5-10 th of October,2008*

2. Gergely Lautner, Júlia Szűcs, Róbert E. Gyurcsányi, Zsófia Balogh, Viola Bardóczy, Beata Komorowska, Tamás Mészáros: Selective detection of of plant virus coat proteins by aptamer based biochips, *A Magyar Biokémiai Egyesület Vándorgyűlés, Szeged, 2008 augusztus 31. – szeptember 3*

3. Bardóczy Viola , Mészáros Tamás: *In-vitro* transzlációs vektorok fejlesztése. Development of *in-vitro* translation vectors, XIII. Nemzetközi Vegyészkonferencia, *2007 November 9- Kolozsvár*

4. Bardóczy Viola, Géczi Viktória, Mészáros Tamás: *In vitro* transzlációs vektorok fejlesztése *2007- A Magyar Biokémiai egyesület Vándorgyűlése, Debrecen*

5. Júlia Szűcs, Gergely Lautner, Róbert E. Gyurcsányi, Viola Bardóczy, Tamás Mészáros, Klára Tóth: Development of biochips for surface plasmon resonance imaging detection of aptamer – ligand interactions; *Chemistry conference, Pardubice, 2007 August*

6. T. Mészáros; A. Helfer; S. Peck, V. Bardóczy ; L. Bögre: MAPKs modules in plant pathogen perception *2005- FEBS Conference, Budapest*

7. T. Arányi, M. Ratajewski,;V. Bardóczy, L. Pulaski, A. Bors, A. Tordai ; A. Váradi Epigenetic Regulation of Transcription *2005- FEBS Conference, Budapest*