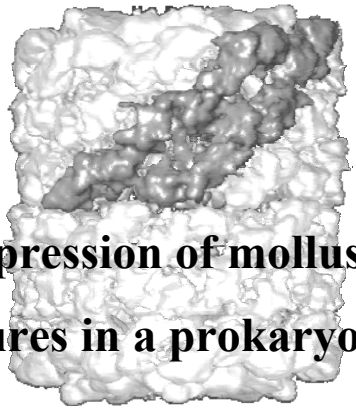


Recombinant expression of molluscan hemocyanin (KLH) substructures in a prokaryotic system: *E. coli*

D i s s e r t a t i o n

Zur Erlangung des Grades

„D o k t o r d e r N a t u r w i s s e n s c h a f t e n“

Am Fachbereich Biologie der Johannes Gutenberg-Universität

in Mainz

Valesca Boisguérin

geb. in Tunis/ Tunesien

Mainz, 2006

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung: 14.09.2006

Index

A Introduction

1 Respiratory proteins

Oxygen-binding molecules are ancient proteins. They probably evolved from enzymes that protected the organism against the inherent toxic activity of the oxygen molecule. As single-celled organisms developed the capacity to use oxygen as an electron acceptor, capturing oxygen and transferring it to the respiratory chain became increasingly important. When multicellular organisms increased in size and complexity, their surface to volume ratios diminished. For any creature having a distance larger than some millimetres between the surrounding respiratory medium and the tissues, transport of oxygen by simple diffusion from the surface is too inefficient to sustain an active life. The development of vascular or coelomic circulatory systems that could move oxygen away from the inner body wall enhanced the oxygen diffusion rate, but the low solubility of oxygen in body fluids was still limiting. For small or relatively inactive creatures, the amount of oxygen that can be dissolved in the circulating blood may prove sufficient for metabolic needs. But larger, complex organisms especially metabolically active ones, demand a more efficient supply. The evolution of simple oxygen-binding proteins into multisubunit, circulating proteins, in combination with the advent of circulatory systems, enabled the transport of oxygen on a significant scale from the periphery of the organism to metabolizing cells in the interior. Thus, it is not surprising that several invertebrate phyla, at about the same time, evolved proteins specialized for oxygen transport. The current estimates date these events to the Precambrian – about 600 to 800 million years ago. Intracellular iron-containing protoheme or myoglobin-type proteins and perhaps copper-containing proteins that could enhance oxygen diffusion and storage began to appear. Interestingly, the evolution of such proteins apparently occurred independently and in parallel in several different groups of organisms. The result is that today we find at least three quite distinct kinds of oxygen transport proteins, each utilized in one or several related phyla (van Holde & Miller, 1995; Terwilliger, 1998).

These three major types of oxygen-transport proteins are those with iron within the prosthetic group (2 types), which reversibly binds to oxygen, and those with copper (1 type). The circulating iron proteins include cellular and extracellular hemoglobins and cellular hemerythrins, where the iron is either coordinately bound to a protoporphyrin or heme group and attached to the protein, or covalently bound to the protein molecule, respectively. The hemoglobins, which serve the vertebrates and also a number of invertebrate

phyla, represent a major class of oxygen-transport proteins. Hemoglobins, especially those found in invertebrates, exhibit an astonishing variety in structural form (Terwilliger, 1998). Hemerythrins are proteins in which oxygen is bound between a pair of iron atoms and, despite their name, do not contain heme.

The hemocyanins are very different from either of the above. In these non-heme proteins, oxygen is bound between a pair of copper atoms. The metal, which is in the cuprous state when the protein is deoxygenated, becomes cupric in the oxy form; the bright blue colour developed on oxygenation has given the protein its name. These copper-based, blue oxygen-transport proteins include molluscan and arthropodan hemocyanins. While the hemocyanins in these two phyla are very distinct in quaternary structure and sequence, their active sites include six highly conserved histidines that bind the two copper atoms; both coppers together bind one oxygen molecule, reversibly. Almost all iron and copper proteins involved in oxygen transport are multisubunit proteins. Consequently, they exhibit cooperative oxygen binding and allosteric modulation of oxygen affinity (van Holde & Miller, 1995; Terwilliger, 1998).

Arthropod hemocyanins are constructed from six 75 kDa subunits, each with one oxygen-binding site, assembled into a ~ 450 kDa hexamer. These hexamers assemble further into 2x6-, 4x6-, 6x6- and 8x6-mers, depending on the class or species (Terwilliger, 1998; for review, see Markl & Decker, 1992).

2 Molluscan hemocyanins

Molluscan hemocyanins have a minimum native molecular mass of ~ 4 million Daltons, with a diameter of about 35 nm and are therefore among the largest proteins known in nature. The basic structure is a homodecamer (Fig. 1b) which can dimerize into didecamers (Fig. 1a) and form multidecamers by consecutive linear addition of decamers to create tube-like multimers. In detail, the decamer consists of five subunit dimers (Fig. 1c), which in most gastropods forms a D5 didecamer, described simply as a hollow cylinder with a central collar complex at both ends (Siezen & van Bruggen, 1974; Miller *et al*, 1990; Lamy *et al*., 1993; Lambert *et al*., 1994, Meissner *et al*., 2000). The subunit/monomer itself consists of a polypeptide chain with a molecular mass of 350 - 450 kDa, which contains seven or eight globular functional units (FUs, Fig. 1d), depending on the species. These FUs are generally termed FU-*a* to FU-*h,* starting from the N-terminus and are inter-connected by short linker peptides of 10 - 15 amino acids (Lang, 1988; Lang & van Holde, 1991). By high resolution cryo-electron microscopy

and 3D-reconstruction (12 Å) the structure of the didecamer of the hemocyanin of *Haliotis tuberculata* isoform 1 (HtH1) has been determined (Fig. 1) by Meissner *et al.* (2000).

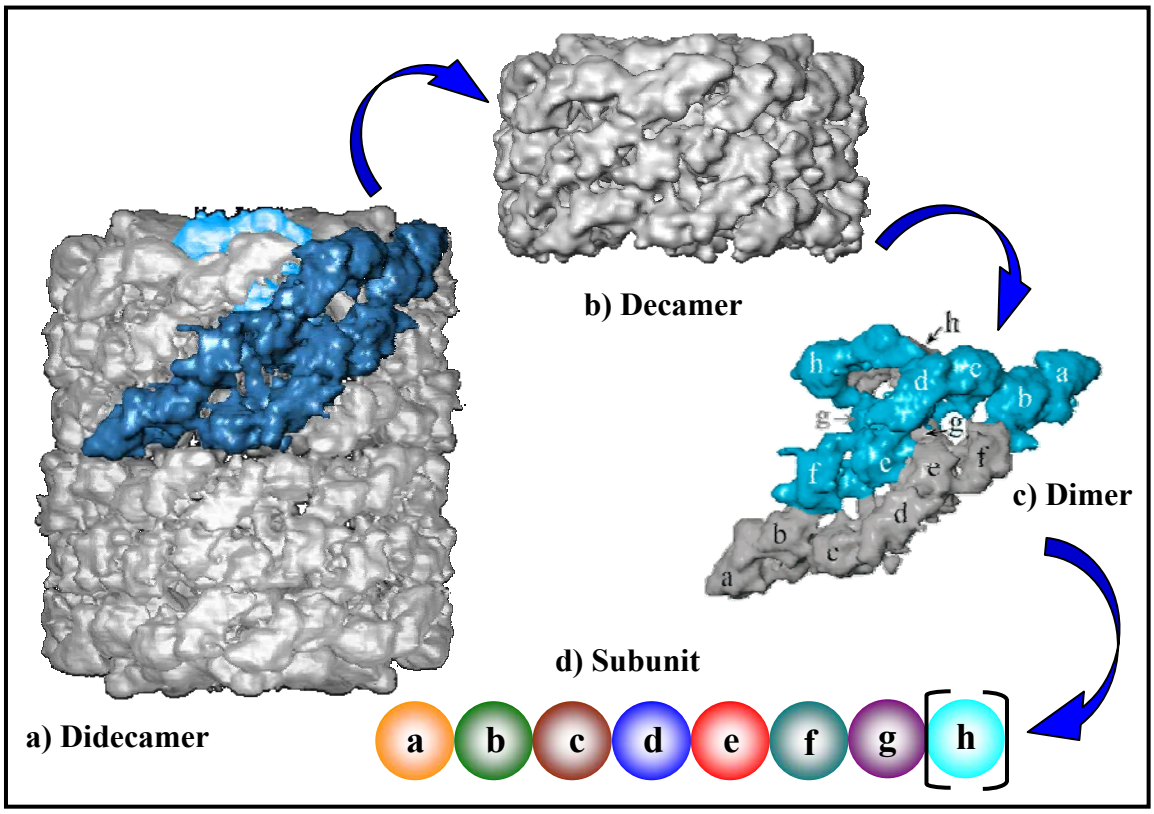


Fig. 1: The structural organization of *Haliotis tuberculata* hemocyanin, isoform 1 (HtH1). **a**) Didecamers are the basic structure for gastropod and bivalve hemocyanin, while the decamer, **b**), occurs only in chitons and cephalopods. **c**) The decamer is formed from five subunit dimers. **d**) One subunit contains seven or eight functional units, termed FU-*a* to FU-*h* (Meissner *et al.*, 2000).

While the FUs -*a* to -*f* form the cylinder wall, the collar complex is formed from the domains FU-*g* and FU-*h* (Fig. 1, c). Within the functional units, each copper atom (CuA and CuB) is bound by three highly conserved histidines (van Holde *et al.*, 1992; Cuff *et al.*, 1998; Miller *et al.*, 1998). One oxygen molecule can bind reversibly to this active site. Furthermore, one or two binding sites for carbohydrate side chains can be found within several FUs, which are primarily bound by N-glycosylation (N for asparagine) to the polypeptide chain (Keller *et al.*, 1999; Lieb *et al.*, 1999 & 2000) and can be recognized by the binding motives NXS or NXT (where X can be any amino acid, S represents serine and T is threnonine). The quantity of hemocyanin-bound polysaccharides can be up to 9 % of the whole molecular mass (van Holde, 1992; Stoeva *et al.*, 1999) and a variety of glycan side chains has been identified (Kurokawa *et al.*, 2002).

2.1 The crystal structure of FU-g of Octopus dofleini and FU-2e of Rapana thomasiana elucidates the relationship between structure and function

The X-ray analysis of the C-terminal collar domain FU-*g* of *Octopus dofleini* hemocyanin (OdH-g; Cuff *et al*., 1998) has provided a detailed view of the steric arrangement of the polypeptide chain, and due to sequence homologies it can be taken as a model for the other FUs. Following Sander & Schneider (1991), structural homology between two proteins is given when the similarity of the sequences is equal or higher than the threshold of homology of a defined sequence length. With more than 40 % sequence identity when comparing FU-*g* of *Octopus dofleini* to the other seven FU types of other molluscan hemocyanin, this threshold is exceeded. New insights have been provided by the X-ray structural analysis of the hemocyanin wall FU-2*e* of a gastropod, *Rapana thomasiana* (RtH2-*e*; Perbandt *et al*., 2003). Both crystal structures are shown in Fig. 2.

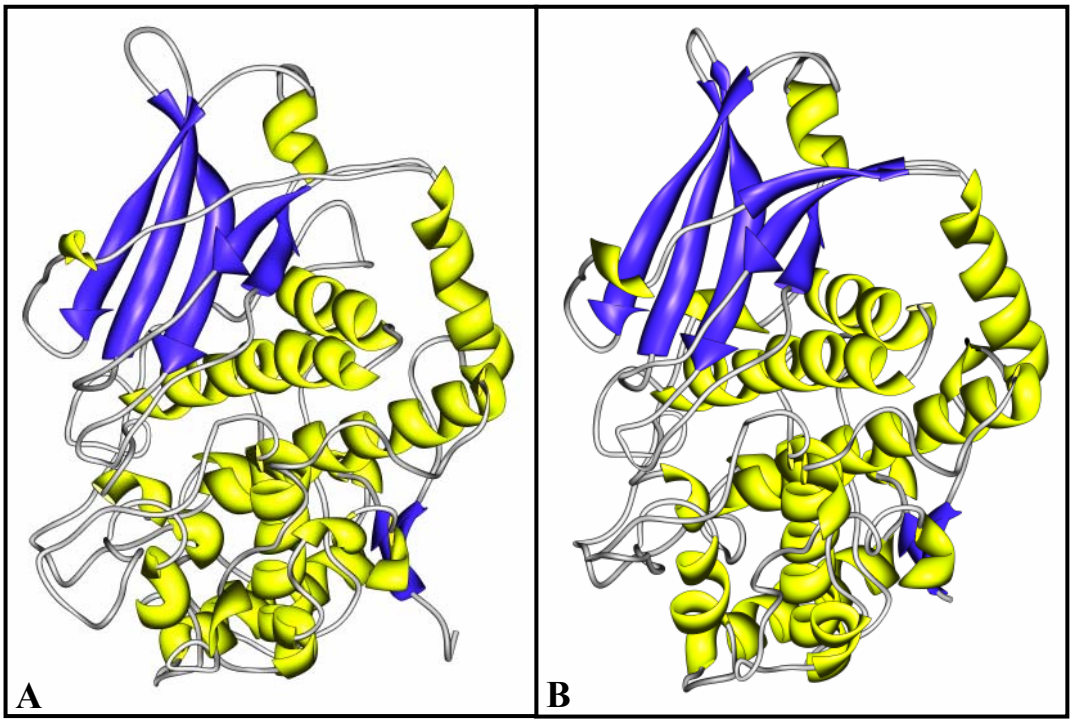


Fig. 2: Crystal structure of RtH2-*e* (*Rapana thomasiana,* **A**), and OdH-*g* (*Octopus dofleini*, **B**). α-helical domains are displayed in yellow, β-sheet-structures in blue.

Both FUs can be divided into two structural domains (Fig. 2A and B): an α-helical copper-binding domain (yellow) and a five-stranded β-sandwich domain (blue). The formed α-helical domain represents the more conserved region of the protein. As already mentioned, there are two copper-binding sites CuA and CuB, in which three histidine residues are present. The second histidine of CuA forms a thioether bridge with a cysteine, which can contribute to the cooperative modulation of the oxygen affinity (Gielens *et*

al., 1997). Furthermore, two disulfide bridges are present close to these histidines. While the first one stabilises a β-β-hairpin-structure near the copper binding site CuA, the second disulfide bridge binds a loop to an α-helix, providing two of the three histidines of CuB. The crystal structure of OdH-*g* exhibits an additional disulfide bridge in the area of the β-sheet domain which, however, is not present in all FUs and therefore may not play an essential role in oxygen binding (Topham *et al*., 1999).

When comparing the two FU crystal structures, prominent differences can be defined. First, FU-*g* of *Octopus dofleini* forms dimers during crystal formation, while FU-2*e* of *Rapana thomasiana* creates a regular cylindrical structure. This different molecular arrangement within the crystal can be explained by the location of the FUs in the structural organisation of the native hemocyanin molecule. Furthermore, a tunnel is present within RtH-2*e*, which runs from the surface to the active site. The structural analysis of RtH-2*e* was performed under deoxygenation, while OdH-*g* was crystallized in its oxygenated form, so that the tunnel formation may possibly be related to oxygen binding (Perbandt *et al*., 2003). An additional difference relates to the location of the glycosylation site in the two FUs.

2.2 The hemocyanin of most gastropods forms didecamers and sometimes multidecamers

In general, the hemocyanins of gastropods form didecamers with a molecular mass of ~ 8 million Daltons. The decamer is a hollow cylinder, wall of which contains ten copies of the FUs *abcdef* and an internal collar region contains the FUs *g* and *h* (Lieb *et al*., 1999). In contrast to cephalopods, the collar is found at the edge of each decamer and therefore gives the hemocyanin decamer asymmetry, thus the decamer has a collar-side and an open-side (Harris *et al*., 1992; Orlova *et al*., 1997; Meissner *et al*., 2000). Didecamers result from a face-to-face assembly of the open-sides of two decamers (van Holde & Miller, 1982, van Holde *et al*., 1992). Tubular multidecamer formation can be found for some marine gastropods, where several decamers successively assemble with their open-side around a central didecamer (Herskovitz *et al*., 1991; Markl *et al*., 1991b, Lieb *et al*., 2004).

Comparing the different species, it is clear that one, two or three different hemocyanin isoforms can be found in gastropods. While the hemocyanin of the sea hare *Aplysia californica* (AcH) is represented by only one isoform (Herskovits & Hamilton, 1991; Lieb *et al*., 2004), two different isoforms, KLH1 and KLH2, are present in the giant keyhole

limpet *Megathura crenulata* (Senozan *et al*., 1981; Senozan & Briggs, 1989; Gebauer *et al*., 1994; Swerdlow *et al*., 1996; Söhngen *et al*., 1997; Harris & Markl, 1999), in the abalone *Haliotis tuberculata* (HtH1 and HtH2; Keller *et al*., 1999; Lieb *et al*., 1999) and also in *Rapana thomasiana* (RtH1 and RtH2; Idakieva *et al*., 1993; Gebauer *et al*., 1999). For *Helix pomatia* (HpH) and *Helix aspersa* (HaH) it has been claimed that three hemocyanin isoforms occur in the hemolymph (van Breemen *et al*., 1975; Gielens *et al*., 1987; Gielens *et al*., 1990). The sequence identity between the two hemocyanin isoforms of *Megathura crenulata* is around 65 %. Comparing the sequences of two vetigastropodan hemocyanins containing two different isoforms, like *Haliotis tuberculata* and *Megathura crenulata*, it is apparent that the highest sequence identities are found between corresponding isoforms of the different animals (HtH1 ↔ KLH1: ~ 67 %), while different isoforms of one animal have lower sequence identities (KLH1 ↔ KLH2: ~ 61 %). This result leads to the hypothesis that the splitting up into the two hemocyanin isoforms within a common has occurred before the dispartment of the genera *Haliotis* and *Megathura* and the families of *Fissurelidae* and *Haliotidae*, respectively (Gebauer *et al*., 1994; Keller et a., 1999; Lieb *et al*., 1999; Awenius *et al*., 2000).

3 The hemocyanin of the Keyhole Limpet *Megathura crenulata*

As long as 30 years ago, the **K**eyhole **L**impet **H**emocyanin, **KLH**, was found to have a strong immuno-stimulatory function and is therefore today widely used in basic immunological research and in some immunotherapeutic clinical applications. KLH is the study material of the present work and therefore the biochemical, structural and immunological properties will now be covered in detail.

In several studies it was shown that two immunological different KLH isoforms exist (Senozan *et al*., 1981; Savel-Niemann *et al*., 1990; Markl *et al*., 1991a, b). Due to their different migration during native PAGE, they were termed KLH1 and KLH2 (Gebauer *et al*., 1994). The molecular mass determination by SDS-PAGE showed, compared to marker proteins, that the subunit of KLH1 was ~ 390 kDa, while the KLH2 was ~ 360 kDa (Van Holde & Miller, 1995; Söhngen *et al*., 1997). It could later be shown that in KLH2, a C-terminal peptide is sometimes lost; both isoform have therefore an average molecular mass of ~ 400 kDa (Gebauer *et al*., 1999b). Transmission electron microscopy showed that three aggregation forms can be observed for KLH: didecamers, clusters of didecamers and multidecamers. It was shown that the didecamers were always built up

as homo-oligomers (Gebauer *et al*., 1994; Harris *et al*., 1992). Furthermore, it was shown that multidecamer formation occurred only for KLH2 (Fig. 3B, white arrow; black arrow: didecamer), while didecameric aggregation (Fig. 3A, black arrow: side view, white arrow: top view) was found for KLH1 (Markl *et al*., 1991b).

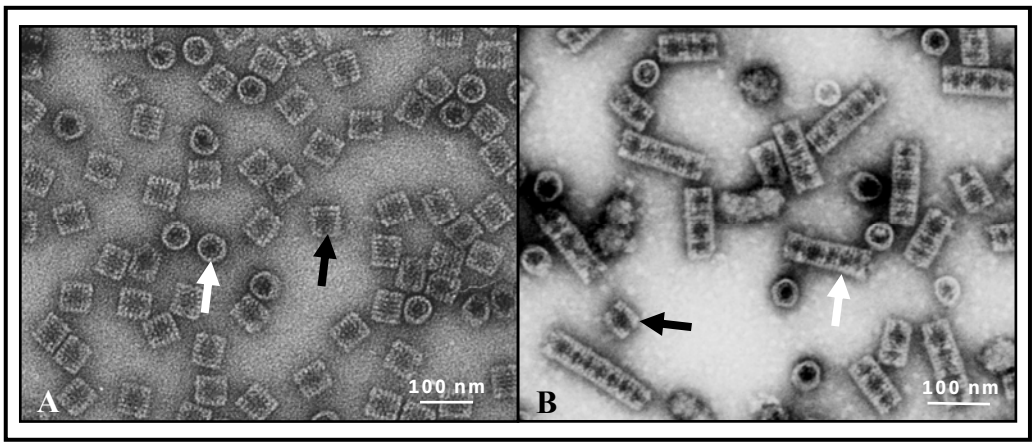


Fig. 3: Transmission electron micrographs of keyhole limpet hemocyanin (Bergman, S., doctoral thesis, 2004). A: KLH1 didecamers (black arrow: side view; white arrow: top view). B: KLH2 didecamers and mutlidekamers (black arrow: didecamer; white arrow: multidecamer).

The polypeptide chains of the subunits of both isoforms each contain eight functional units (FUs), termed FU-*a* to FU-*h* from the N-terminus to the C-terminus. The FU molecular masses range from ~ 40 kDa to ~ 60 kDa. Specific cleavage by incubation of native KLH with the enzyme elastase resulted in the evidence of eight individual and immunological different FUs from each KLH isoform (Söhngen *et al*., 1997, Gebauer *et al*., 1999a).

4 Clinical relevance of KLH

Hemocyanins from molluscs and arthropods are useful for various types of immunological studies because of their unique physical and immunochemical properties. These respiratory proteins are potent immunogens, which induce the synthesis of large amounts of specific antibodies (Herscowitz *et al*., 1972). From the 1960s medical interest grew around the hemocyanin of *Megathura crenulata*, as experiments with mammalian animals showed the immune stimulatory properties of KLH (Weigle, 1964; Curtis *et al*., 1970; Herscowitz *et al*., 1972). In 1967, the first immune competence tests were done in humans by Swanson & Schwartz. The main field of application of native KLH is for the treatment of superficial bladder carcinoma (Olson *et al*., 1974, Jurincic *et al*., 1988; Flamm *et al*., 1990; Sargent & Williams, 1992). Several studies concerning the immunotherapeutic action of KLH showed that this molecule contains oligosaccharides with a

terminal galactose(β1-3)-N-acetylgalactosamine-epitope. This epitope leads to specific antibody formation *in vivo*, which cross-reacts with the Thomsen-Friedenreich (T) antigen (i.e. there is an equivalent antigenic epitope on the surface of bladder carcinoma cells) and therefore results in enhanced immune stimulation followed by reduction of tumour growth. KLH is also especially suited for tumour treatment due to its safety in humans and to the fact that compared to other medicines or therapies (e.g. chemotherapy), no side-effects have been found (Jurincic *et al*., 1990).

The large mass of the KLH molecule is apparently not the reason for its strong antigenicity, because the immune response also occurs after dissociation of the didecamers into single 400 kDa subunits. The significant factor is likely to be the multi-antigenicity, due to the eight immunological different functional units found in each isoform. There can be more than 100 KLH xenogen epitopes, challenging the immune system all at once, which could lead to this massively enhanced immune response. The strong immuno-modulatory properties of KLH induce cytotoxic T lymphocytes and lead to macrophage invasion around the application area. KLH can therefore influence the innate as well as the cell-mediated immunity, and it also interacts with monocytes and polymorphonuclear lymphocytes (Tzianabos, 2000). Furthermore, KLH is also used in immunological laboratories and clinics as a hapten-carrier for low molecular weight compound molecules, to enhance antibody production by improved antigen presentation.

Knowledge of the cross-reaction between a carbohydrate epitope on the surface of larvae of *Schistosoma mansoni* and one found in KLH (Dissous *et al*., 1986; Ko & Harn, 1987) resulted in an additional clinical interest in the hemocyanin of *Megathura crenulata*. *S. mansoni* is a human pathogen, a trematode worm which causes the infectious disease Bilharziosis. The common epitope was shown to be a terminal fucose residue, which is bound in a (α1→3)-position to N-acetyl-galactosidase (Wuhrer *et al*., 1999; Kantelhardt *et al*., 2002). Injection of KLH during Bilharziosis infection produces a cross-reaction of anti-KLH-antibodies against the *Schistosoma*-glycolipids (Markl *et al*., 1991a). These antibodies can additionally be used directly for Bilharziosis diagnosis by ELISA testing.

The broad clinical approaches using KLH seems to be almost endless when considering the flood of clinical data produced during the past three decades. In addition, the preliminary results of Kippel (1991) showed positive effects in the field of renal cell carcinoma, with remission and reduced tumour progression. KLH had also been compared against mitomycin C chemotherapy in patients and was found to be superior in preventing bladder tumour recurrence with no adverse local or systemic side effects

(Jurincic-Winkler *et al*., 1996 & 2000). In 2002, the *in vitro* anti-cancer effects of KLH were tested in multiple cancer cell lines and resulted in a significant cancer cell growth inhibition for estrogen-dependent and -independent breast cancer, pancreatic and prostate cancer (Riggs *et al*., 2002). Their latest results (2005) focussed on KLH's action on breast and pancreas cancer cells via apoptotic or cytokine pathways. Significant growth inhibition was observed in all cell lines at all KLH concentrations tested as well as significant changes in cytokine production in all cell lines. It was also reported that KLH inhibited the growth of human Barrett's esophagal adenocarcinoma *in vitro, via* apoptotic and non-apoptotic mechanisms (McFadden *et al*., 2003). They hypothesize that KLH reduces the growth of Barrett's cancer cells by altering protein expression profiles. These results indicated that this was accompanied by a cellular stress response and attenuation of metabolic processes. Investigations of Vona-Davis *et al*. in 2005 showed that KLH also inhibits the growth of skin melanoma cell lines *in vitro,* up to 50 % of the control. They showed that KLH also elicits an anti-proliferative response and results in changes in late apoptosis. So, KLH has had a considerable impact in clinical research, especially throughout the range of different anti-cancer therapies.

5 Carbohydrates of KLH

The knowledge of the oligosaccharide content of KLH is growing, due to several studies. It is thought that the carbohydrate components of this highly immunogenic glycoprotein are widely involved in the immune stimulatory properties. Therefore, there is a need to further this knowledge. The carbohydrate moiety is about 4 % of the KLH molecular mass (van Kuik *et al*., 1990; Kamerling & Vliegenthart, 1997) and a variety of structural glycosylation motives had been identified (Stoeva *et al*., 1999; Kantelhardt *et al*., 2002; Kurokawa *et al*, 2002; Wuhrer *et al*., 2004). Due to the amino acid sequence, determined within our working group, conclusions can be drawn about the available N- and O-glycosylation sites.

There are three known possibilities concerning carbohydrate association with proteins:

- N-glycosylation

Here, the linkage of the oligosaccharide occurs by covalent binding to the nitric amide of an asparagine residue. The glycosylation motives consist of the sequences NXT, NXS or NXC (Foster & Davie, 1984), whereas X can be substituted by any amino acid, except proline. The N-glycosylation *via* NXC can only take place if the cysteine (C) is not involved in the formation of a disulfide bridge (S - S). Although N-linked oligosaccharides

are generally quite different, a common core structure can be defined. This uniform element consists of a penta-saccharide, containing three mannose and two N-acetylglucosamine (GlcNAc) residues.

- O-glycosylation

This type of glycosylation occurs at the oxygen atom of the amino acid residues of serine or threonine. A specific peptide sequence binding motif is not known, but the sugar side chains contain mostly N-acetylglucosamine or N-acetylgalactosamine. The linkage can result in eight different core structures. The most common one, Core-1, is also termed Thomsen-Friedenreich(T)-antigen, which is also found as the KLH-equivalent antigen on the surface of bladder cancer cells. Core-2 is only found occasionally and the remaining six sugars are even more infrequent.

- C-glycosylation

The presence of C-glycosylation has only recently been found; it was discovered in human RNase 2 in 1998 by Vliegenthart and Casset. Here, mannosylation occurs at the C2-atom of a tryptophan residue, incorporated into the peptide sequence motif WXXW. Again, X can be any amino acid. This sequence motif was also found in several other mammalian proteins, wherefore C-glycosylation is thought to be also present in other proteins (Krieg & Hartmann, 1998). In KLH, however, although this sequence motif is found several times, the linkage of carbohydrates to a C-atom had yet not been detected.

6 Goal of the present work

For many years, the studies performed within our research group have focused on the molluscan hemocyanins. These macromolecules are impressive, not only because of their enormous mass but also because of their use in broad ranging studies. Investigations on hemocyanin organisation, from quaternary to primary structure are coming increasingly to the fore. Due to its large size, the hemocyanin gene can provide an enormous amount of sequence information. The basis for numerous phylogenetic analyses is provided by the comparative study of the protein subunit substructures, the functional units (FUs). On one hand, the FUs are marked by recurrent and highly conserved characteristics, whereby the sequences can be compared to each other. On the other hand, variable areas are present around the copper-binding site, whose sequence variances can serve as evolutionary markers.

In addition to the interesting structural and phylogenetical aspects, KLH, the hemocyanin of the giant keyhole limpet *Megathura crenulata*, has proven its importance within

immunological and clinical research. Facing the numerous application possibilities of such a potent immune stimulator and vaccine carrier such as KLH, a cost-effective, easy biosynthesis and purification of this molecule is desired. To date, the available KLH for clinical applications, produced by companies such as *Biosyn*, is still extracted and purified from of the hemolymph of the marine limpet *Megathura crenulata*. The animals are captured, part of their hemolymph is withdrawn and the animals are returned to their natural habitat. In contrast, to achieve the biotechnological purification of KLH is much more complicated. As already mentioned, the naturally occurring KLH isoforms are based on a 400 kDa polypeptide and due to this large mass the production of high-purity hemocyanin is possible only in combination with high development costs, time- and effort-investments. Since there is commercial interest in the development of novel immunotherapeutic treatments, a procedure has to be found to enable a fast, cost-effective and simple way of KLH-production using present-day biotechnological approaches.

The main interest on recombinant hemocyanin of our working group, however, lays on investigations on the biological activity of the protein as well as on structural and functional analyses.

The goal of the present work was, therefore, the establishment of recombinant expression of KLH in our working group, using the Gateway™ technology from Invitrogen, Karlsruhe. This will initially be attempted in a prokaryotic system: *E. coli*. For this purpose, the generation of KLH1- and KLH2-encoding cDNAs of single FUs as well as the cloning of these FUs into this expression system was required. Biochemical investigations of the recombinant expression constructs were done and compared to their native counterparts, isolated from the keyhole limpet. For further recombinant expression studies in an eukaryotic expression system, namely insect cells, the generation and cloning of cDNAs encoding for larger substructures, up to the whole subunit, encompassing all eight paralogous functional units, will be attempted by SOE-PCR.

B Materials and Methods

1 Animals

The marine mollusc *Megathura crenulata* is one of the largest known keyhole limpets (commonly known as the giant keyhole limpet). It belongs to the phylum Mollusca and is classified into the class Gastropoda, subclass Prosobranchia. Within the order Archaegastropoda, it is a member of the family Fissurellidae.

The species is present in the Pacific Ocean from Monterey Bay to the Isla Asuncion off Baja California in a living space from low-tide line to shallow depths, but also common on breakwaters.

Fig. 4: The giant keyhole limpet *Megathura crenulata* from the top view. (From: www.goldenstateimages.com/livereef.htm, left picture, and www.diver.net, right picture)

The outer mantle tissue may be of cream or grey colour with stripes and spots of darker brown or grey pigment, or can also be of a solid black (Fig. 4). The large muscular flattened yellow foot acts like a strong suction disk, with which the animal adheres to the rock surface. Limpets move about in search of food, drawing in water around the foot, and they live either as scrapers of small detritus or as herbivores (seaweed, algae, tunicates, sponges). For feeding purposes they use their radula, a sharp, sand-papery-like tongue. Although the shell is relatively small, up to 8 cm long, the animal can extend itself up to 13 cm. The mantle completely covers the shell in the living mollusc, leaving only the keyhole visible (Fig. 4). This single large opening in the centre of the shell's radiating ribs is the source of the name and it allows waste products to exit without re-entering the gills or mouth area (Fig. 5). The eggs and sperm also exit this way. The in-

ternal digestive organs of keyhole limpets are also modified to permit an exit channel through the body, leaving at the top.

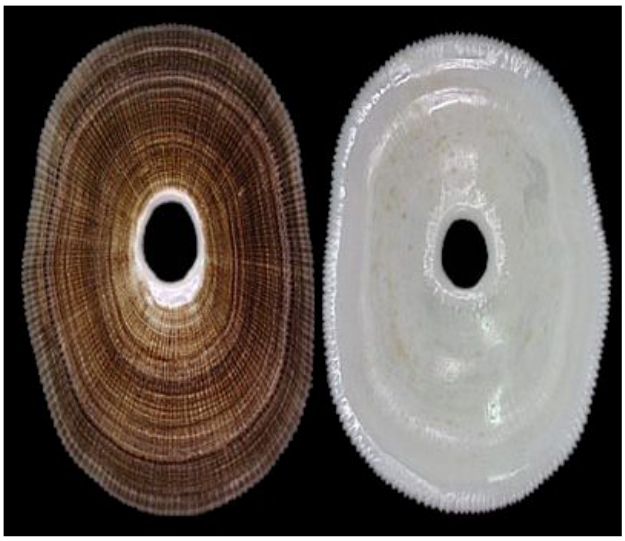


Fig. 5: Inner and outer view of the shell of the giant keyhole limpet *Megathura crenulata* (From: http://www.gastropods.com/2/Shell_4152.html).

The animals used in this work for RNA preparation were kindly provided by the Biosyn Company, Carlsbad, California. They were stored in the Institute of Zoology, University of Mainz, at - 80 °C.

Furthermore, small 1-year-old snails were made available by the courtesy of S.M.E.L. (Syndicat Mixte d' Equipment du Littoral), Blainville sur Mer, France. While the RNA from these animals had been extracted by [redacted], the mRNA enrichment was done by [redacted] and was kindly placed at the disposal of the present work for all investigations concerning KLH 1.

The RNA used for the research work on KLH 2 was extracted and kindly made available by [redacted].

2 Applied chemicals and equipments

The chemicals were obtained in analytical quality and purchased from Roche (Mannheim), Fluka (Buchs, Schweiz), Gibco Life Technologies (Karlsruhe), ICN Biomedicals (Meckenheim), Merck (Darmstadt), Roth (Karlsruhe), Serva (Heidelberg) and Sigma (Deisenhofen). Commercial Petri dishes were from Greiner (Frickenhausen) and plastic tubes (termed „Falcons") from Falcon (Heidelberg). Standard laboratory equipment was obtained from Labotec (Wiesbaden). Unless otherwise mentioned, all centrifugation steps were done in a microliter bench-centrifuge (Centrifuge 5417 R; Eppendorf, Hamburg) or for larger volumes in a Sorvall® RC-5B bench centrifuge (Du Pont Instruments,

Bad Homburg). Falcons were centrifuged in a chilled Megafuge 1.0 R (Heraeus, Fulda). For the evaluation of agarose gels UV-Transilluminators at 366 nm and 254 nm wavelengths, from Bachofer (Reutlingen), were used. A BioPhotometer from the Eppendorf Company (Hamburg) was used for spectrophotometric analysis. An Ultra-Turax (purchased from Jahnke and Kunkel, Staufen in the Breisgau) was employed for tissue preparation. All the chemicals and equipment used will also be mentioned in the appropriate sections, below.

3 General precautions

For all molecular or microbiological handling autoclaved equipment was used as well as autoclaved and demineralised water (1 bar overpressure, 120 °C, 30 minutes). To prevent contamination with RNases, bidistilled, (diethyl-pyrocarbonate) DEPC-treated water was used for all experiments with RNA, as well as filter tips and latex cloves. For additional safety, when dealing with ethidium bromide-contaminated equipment or gels, nitrile gloves were used.

4 Common microbiological methods

4.1 Medium and Agar plates

LB-Medium (per litre)	5 g yeast extract
	10 g NaCl
	10 g Tryptone
	pH 7,0
LB-Agar (per litre)	20 g Agar-Agar
	at 1 l LB-Medium

Both, medium and Agar-plates were produced following their respective protocol and were autoclaved at 120°C for 30 minutes at 1 bar overpressure. After cooling to 55 °C, the LB-Agar was poured into commercial Petri dishes and for the selective growth of specific bacterial strains the following quantities of the appropriate antibiotics were added:

Ampicillin	100 mg/l
Kanamycin	50 mg/l

The Petri dishes were then stored at 4 °C in a cold room and the medium was kept in the dark at room temperature.

4.2 Bacterial strains and vectors

For cloning and expression the following bacterial strains and vectors were used:

Bacterial strains Invitrogen, Karlsruhe:

***E. coli* DH5α**

***E. coli* BL21-AI**

***E. coli* BL21Star™(DE3)plysS**

***E. coli* DB3.1**

Vectors from Invitrogen, Karlsruhe:

pDONR201™: 4,470 bp, kanamycin resistance

pDONR221: 4,762 bp, kanamycin resistance

pDEST14™: 6,422 bp, ampicillin resistance

pDEST15™: 7,013 bp, ampicillin resistance, N-terminal GST-tag

pDEST17™: 6,354 bp, ampicillin resistance, N-terminal 6xHis-tag

pENTR-gus™: 3,841 bp, kanamycin resistance

pENTR/D/TOPO™: 2,580 bp, kanamycin resistance, compatible with Gateway™ Technology due to flanking *att*L-sites

PCR-2.1-TOPO: 3,900 bp; kanamycin and ampicillin resistance, lacZα gen (subsequently mentioned as TOPO-TA)

PCR-XL-TOPO: 3,500 bp; kanamycin and zeocin resistance, lacZα gen, specialized for the cloning of DNA fragments > 2 kb (subsequently mentioned as TOPO-XL)

Vectors from Promega, Mannheim:

pGEM-TEasy™: 3,000bp, ampicillin resistance, lacZα-Gen

Vectors from Qiagen, Hilden:

pQE 30: 3,400 bp, ampicillin resistance, N-terminal 6xHis-tag

4.3 Bacterial cultures

- Plate cultures

The bacteria for a cloning reaction were plated on an Agar plate with the appropriate antibiotic and then cultured overnight at 37 °C in an incubator. As required, this plate could then be stored at 4°C.

- Overnight cultures

To prepare an overnight culture a bacterial colony was picked via a sterile pipette tip and transferred into LB-medium. The addition of the appropriate antibiotic depended on the vector used and its resistance, and is mentioned in the appropriate section, below. These samples were shaken at 37°C at 200 rpm overnight.

- Glycerine cultures

For the long-term storage of bacterial cultures, 500 µl of an overnight culture were mixed with an equal volume of 10 % v/v glycerol and frozen at - 80 °C.

- Chemically competent bacterial cells

Chemically competent cells were used for cloning (*E. coli* DH5α) and for propagating the vectors (*E. coli* DB3.1). Competent cells were self-made and prepared out using the $CaCl_2$-method (Cohen *et al*., 1972). Depending on the use, bacterial cells were propagated by plating them on Agar-plates without antibiotics with overnight growth at 37 °C in an incubator. Then, different colonies were picked from the plate and cultured overnight in LB-medium. The next day, a 1:100 dilution was made and shaken at 37 °C until the optical density reached ~ 0,2 at a wavelength of 600 nm. At this time point the bacteria are in a logarithmic growth phase and in the best condition for becoming competent. The culture was centrifuged at 3,500 rpm for 5 minutes (Sorvall® RC-5B; Du Pont Instruments, Bad Homburg) and the pellet resuspended in 20 ml of ice-cooled 0.1 M $MgCl_2$. After a second centrifugation step the bacterial cells were taken up in 2 ml of ice-cooled 0.1 M $CaCl_2$ and allowed to swell on ice, at least overnight, so that they became competent by the following morning. This swelling process destabilizes the bacterial membrane, so that a certain percentage of the cells are able to take up the desired plasmid. The cells are now chemically competent and ready for the transformation step. At 4 °C these bacterial cells are stable for at least one week.

4.4 Propagating the vectors

Because of the lethal effect of the CcdB protein, all Gateway™ vectors containing the *ccdB* gene must be propagated in an *E. coli* strain that is resistant to *ccdB* effects. Therefore the *E. coli* strain DB 3.1 was used, which contains a gyrase mutation (*gyr*A462) that renders it resistant to the *ccdB* effects (Bernard and Couturier, 1992; Miki *et al.*, 1992; Bernard *et al.,* 1993).

To propagate a vector, 75 ng of the commercial available Gateway™ vectors from Invitrogen, Karlsruhe, were transformed into 50 µl of competent *E. coli* DB3.1 (refer to: *5.14 Cloning*, Transformation). Colonies containing the vector were picked and amplified in overnight cultures. After plasmid isolation (refer to: *5.15 Plasmid isolation*), the propagated vectors were kept frozen at - 20°C.

5. Common molecular biological methods

5.1 Total RNA extraction

For the total RNA extraction the E.Z.N.A.® Total RNA Kit from peqlab, Erlangen, was used. A 1.2 g piece of frozen *Megathura crenulata* mantel tissue, stored at - 80 °C, was was purified according to the manufacturer's instructions. Without forgetting to dry the RNA on the column by centrifugation, which is a crucial step that should not be skipped, the RNA was eluted into a fresh tube by adding 50 - 100 µl of DEPC-treated water, depending on the desired final concentration.

For the quantification of the RNA, one can determine the absorption of an appropriate dilution aliquot at 260 nm and then at 280 nm. One A_{260}-unit indicates about 40 µg RNA/ml. The RNA concentration is calculated as follow:

$$\text{RNA conc. } (\mu g/ml) = \text{Absorption}_{260} \times 40 \times \text{dilution factor}$$

The ratio of $A_{260/280}$ is an indication of nucleic acid purity. A value higher than 1.8 indicates a purity of > 90 % nucleic acid.

RNA samples were stored at - 80 °C in sterile DEPC-dH_2O. Under such conditions RNA prepared with the E.Z.N.A. system is stable for at least one year.

5.2 Primer design

The primers used were purchased by the companies Roth (Karlsruhe) and Sigma-Genosys (Deisenhofen) and are listed in the appendix.

Gene specific primers should be chosen conform to the following rules:

- varying length of 18 - 23 nucleotides
- G+C content of about 40 - 60 %
- adapted annealing temperature of both, upstream and downstream, primers
- no self-compatibility
- no palindrome sequence

- no possibility to build primer-dimers or any secondary structures

These features were tested using the computer program Oligo Calculator, version 3.07.

Gateway™ attachment primers were selected according to the gene specific primers and the corresponding cDNA fragments, with appropriate 5' or 3' elongations, depending on forward or reverse application and expression of an N- or C-terminal tag. Gateway™ attachment primers are also listed in the appendix. On all occasions the applied primer concentration was 10 pmol, in dH_2O.

5.3 Polymerase chain reaction (PCR) and its variations

- Standard PCR

PCR is the enzymatic amplification of a specific DNA sequence *in vitro*. This process uses multiple cycles of template denaturation, primer annealing and primer elongation to amplify DNA sequences. It is an exponential process since the amplified products from each previous cycle serve as templates for the next cycle of amplification.

There are three major steps in a PCR, which are repeated for 25 to 40 cycles. One cycle consists of:

- **Denaturation:**
  During the denaturation (at 94 °C), the double stranded DNA melts and opens to give single stranded DNA, and all enzymatic reactions stop (for example: the extension from a previous cycle).
- **Annealing:**
  The primers are mixed by Brownian motion. Ionic bonds are constantly formed and broken between the single stranded primer and the single stranded template. The more stable bonds last slightly longer (primers that fit exactly) and on that piece of double stranded DNA (template and primer) the polymerase can attach and starts copying the template. Once there are a few bases built in, the ionic bond is so strong between the template and the primer, that it does not break anymore. The annealing temperature depends on the melting point of both primers and should be approximately 5 °C beneath.
- **Extension:**
  Depending on the enzyme, 68 °C - 72 °C is the ideal working temperature for the polymerase. The growing complementary DNA strand is completed with dNTPs,

starting from the free 3'OH group of the primer, so that a double stranded DNA-molecule is generated.

The template can include purified genomic or plasmid DNA, RNA converted by reverse transcriptase to complementary DNA (cDNA), or unpurified, crude biological samples such as bacterial colonies or phage plaques. The forward and reverse primers determine the sequence and the length of the amplified product.

The single cycle steps are displayed in Figure 6.

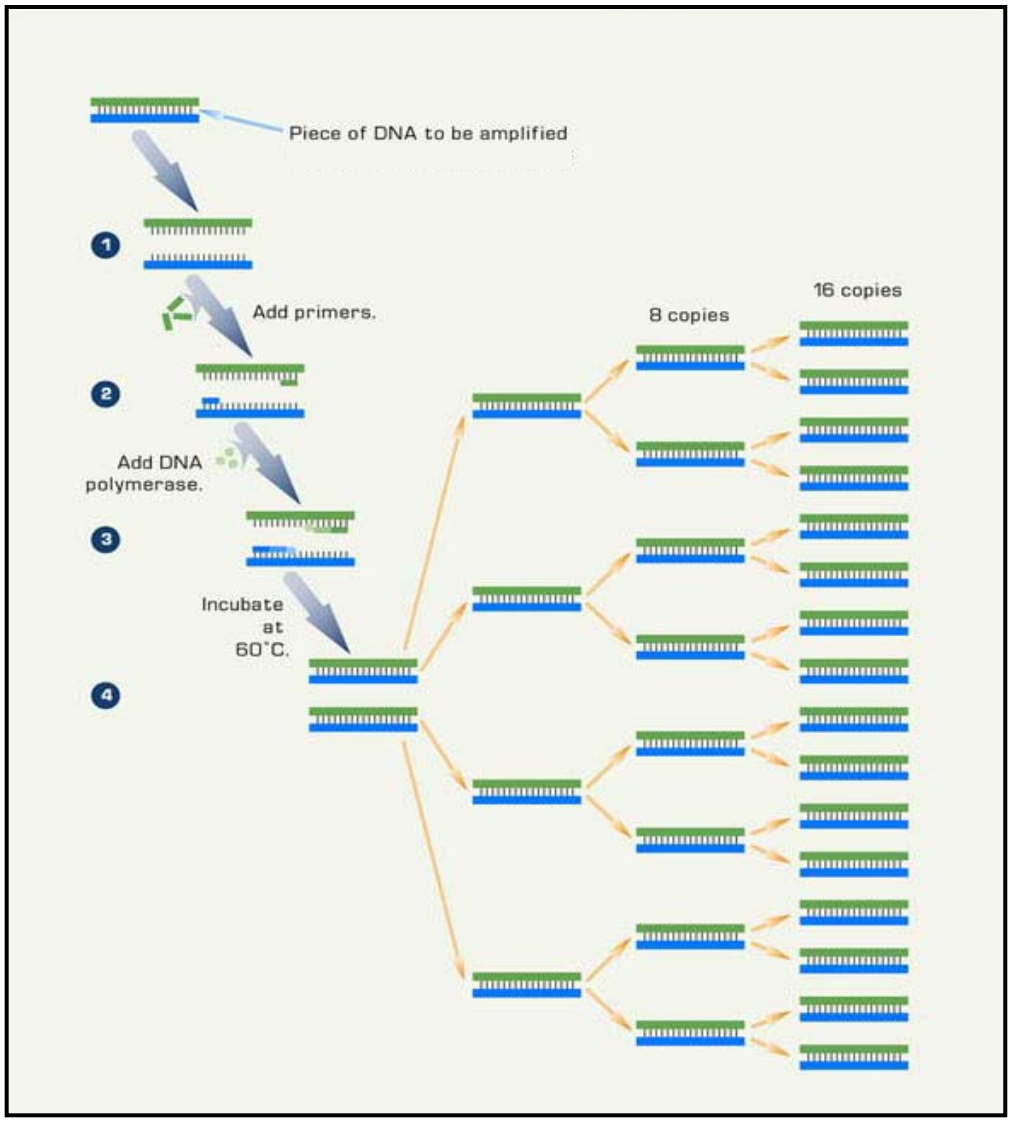


Fig. 6: Scheme of polymerase chain reaction (PCR), (From: http://www.learner.org/ channel/ courses/biology/images/archive/detail/1867_d.jpg).

As enzymes, the following thermostable polymerases were used for the respective PCR variations, following the general manufacturer's instructions:

- **Taq**™ polymerase (Invitrogen, Karlsruhe): for the everyday amplification of simple fragments, adding a single free adenosine-triphosphate at the end of the double stranded DNA molecule, out of ***T****hermophilus* ***aq****uaticus*,
- **Pwo** (Roche, Mannheim): highly processive polymerase with a 3' - 5' exonuclease activity for highest possible fidelity applications, from ***P****yrococcus* ***wo****esei*
- **Expand** DNA-polymerase mix (Roche, Ingelheim): a mixture of Taq and Pwo polymerases, especially for the amplification of long fragments
- **AccuTaq™** (Sigma-Aldrich, Deisenhofen): a polymerase with a 3' - 5' proofreading exonuclease activity for an increased yield and greater fidelity
- **Pfu** (Promega, Mannheim): a polymerase that exhibits 3' - 5' proofreading exonuclease activity for high fidelity, from ***P****yrococcus* ***fu****riosus*
- **Phusion** (Finnzymes, Espoo, Finnland): an accurate and fast polymerase with a processivity-enhancing domain
- **Red Taq ready Mix** (Sigma-Aldrich, Deisenhofen): a ready-to-use mix containing buffer, $MgCl_2$, dNTPs, Taq-polymerase and red loading buffer, for checking clones by PCR

General cycling parameter:

96 °C	2 - 5 minutes	
96 °C	10 seconds	
50 - 70 °C	30 seconds	25 - 40 x
68 - 72 °C	1 minute/ 1 kb	
68 - 72 °C	10 minutes	
4 °C	∞	

All PCRs were run in the automated cycler "T-Gradient" from Biometra, which can heat and cool the tubes with the reaction mixture in a very short time.

- *att*-site PCR

This is a specific PCR technique to add the Gateway™-specific attachment (*att*) sites by extended primers to an already existing PCR fragment, to enable recombination into one of the Gateway™ vectors (refer to: *7.1 Gateway™ Technology*). These *att*-PCRs were

exclusively performed with proofreading polymerases, following the manufacturer's instructions.

- Clone PCR

The clone PCR is a fast procedure to check the integration and the length of an insert after a cloning reaction. After picking the desired clone from a plate, the pipette tip is melted into a clone PCR preparation and then ejected into 5 ml LB-medium containing the appropriate antibiotics. For clone PCRs the Red Taq ready mix from Sigma-Aldrich (Deisenhofen) was used mainly in combination with vector specific primers. 1/4th of the recommended amount of PCR reaction was found to be sufficient to provide a visible result.

- Sequencing PCR

The sequencing PCR was done in our laboratory following the *cycle-sequencing*-method. A specific mix of buffer, polymerase, dNTPs and fluorescence-labelled ddNTPs (dideoxynucleotide-triphosphates) was used. During cycling, the polymerase chain reaction is stopped whenever a ddNTP is randomly incorporated into the growing strand. The result is a mixture of different fragments of varying length, each ending with a fluorescence labelled nucleotide. The detection of the sequence was performed by GENterprise, Mainz.

General pipetting scheme:

Big Dye 3.1	1 µl
Big Dye buffer	3 µl
primer 10 pmol	1 µl
DNA	5 µl
	10 µl

General cycling parameters for sequencing:

96 °C	10 seconds	30 x
55 °C	55 minutes	
4 °C	∞	

Special cycling parameters for the sequencing of Gateway™ vectors:

94 °C	5 minutes	
96 °C	10 seconds	30 x
50 °C	5 seconds	
60 °C	4 minutes	
4 °C	∞	

- SOE-PCR (Splicing by Overlap Extension)

When using SOE-PCR two DNA fragments are fused together by PCR without the use of restriction digestion (splicing). Either both fragments already have complementary overlapping sequences, or they are supplemented with primers by PCR (Fig. 7).

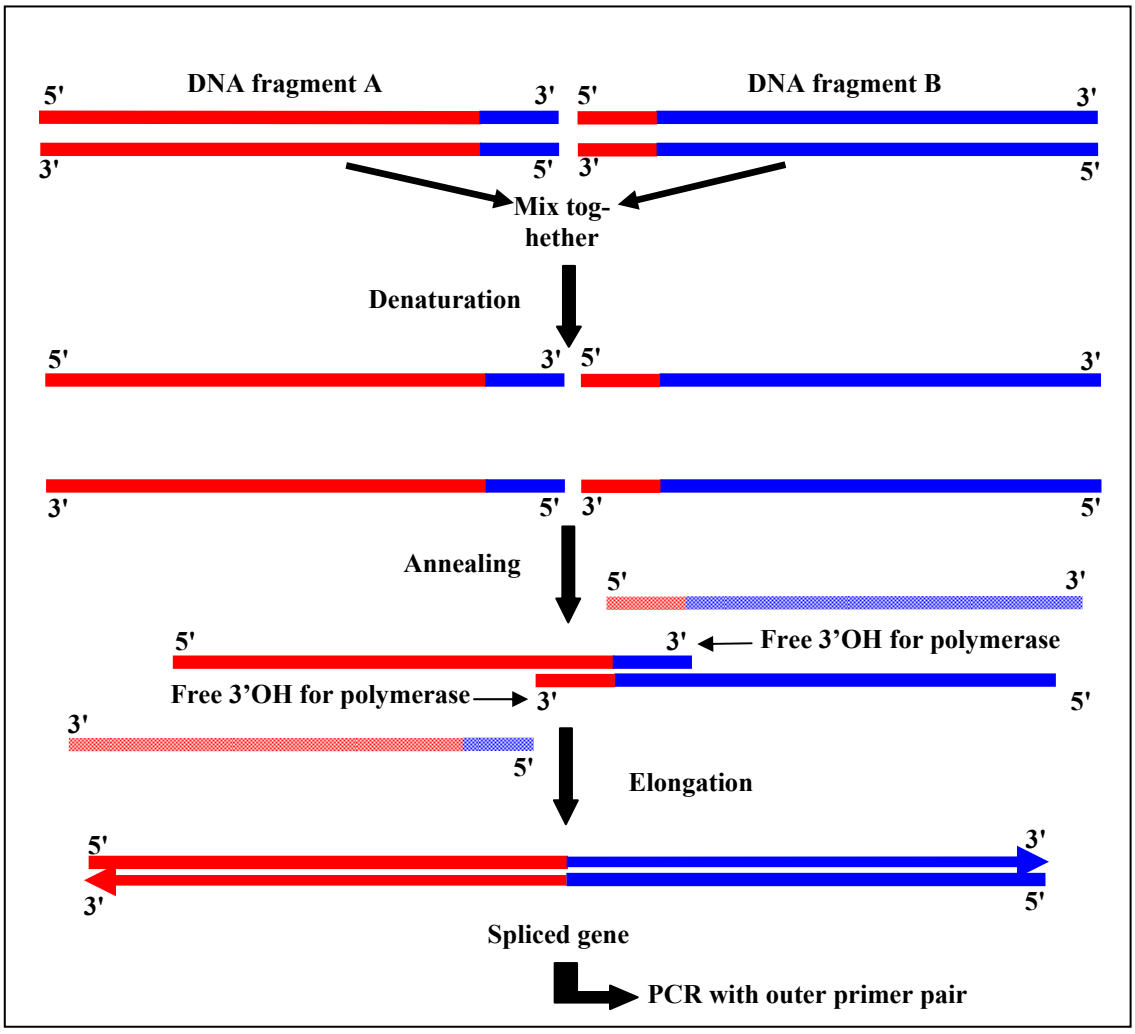


Fig. 7: Scheme of SOE-PCR (**S**plicing by **O**verlap **E**xtension).

After mixing and denaturation of the two fragments, the single strands will anneal to each other on the complementary, overlapping sequence. The resulting free 3‘OH groups so produced serve the polymerase as primers to elongate the strands. The added outer primers help to amplify the spliced fragment during the PCR reaction (Fig. 7).

- Asymmetric PCR

Before starting an SOE-PCR an additional step for asymmetric PCR synthesis can be performed to generate an excess of single stranded DNA-fragments. This is done by applying only one of both primers flanking the required DNA fragment (Fig. 8).

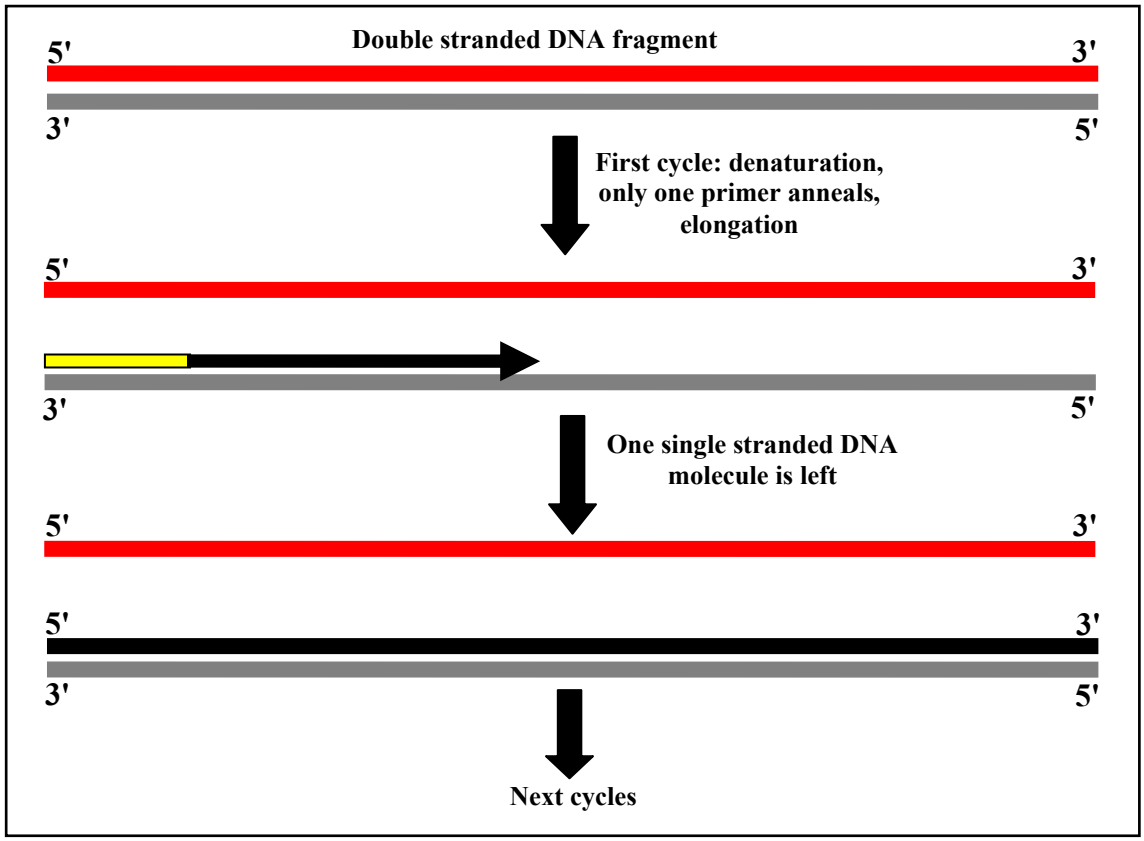


Fig. 8: Scheme of an asymmetric PCR.

- Touchdown PCR

Touchdown PCR uses a cycling program with varying annealing temperatures. It is a useful method to increase the specificity of RT-PCR and PCR. The annealing temperature in the initial PCR cycle should be 5 - 10 °C above the melting point of the primers. In subsequent cycles, the annealing temperature is decreased in steps of 1 - 2 °C per cycle until a temperature is reached that is equal to or 2 - 5 °C below the melting point of the primers. Touchdown PCR enhances the specificity of the initial primer-template duplex formation and hence the specificity of the final RT-PCR product.

5.4 Reverse transcription

Reverse transcriptase, extracted from retroviruses, is an enzyme able to transcribe RNA into a complementary strand of DNA, by incorporating deoxyribonucleotide-triphosphates (dNTPs). Reverse transcriptase is able to copy genetic information from RNA to DNA, which is the reverse of the usual direction (DNA to RNA). To start the reverse transcription, the transcriptase needs a primer with a free 3'OH group. Adding RNase Inhibitor (Rnasin) hinders the degradation of RNA by contaminating RNases. The scheme in Figure 9 depicts the reverse transcription:

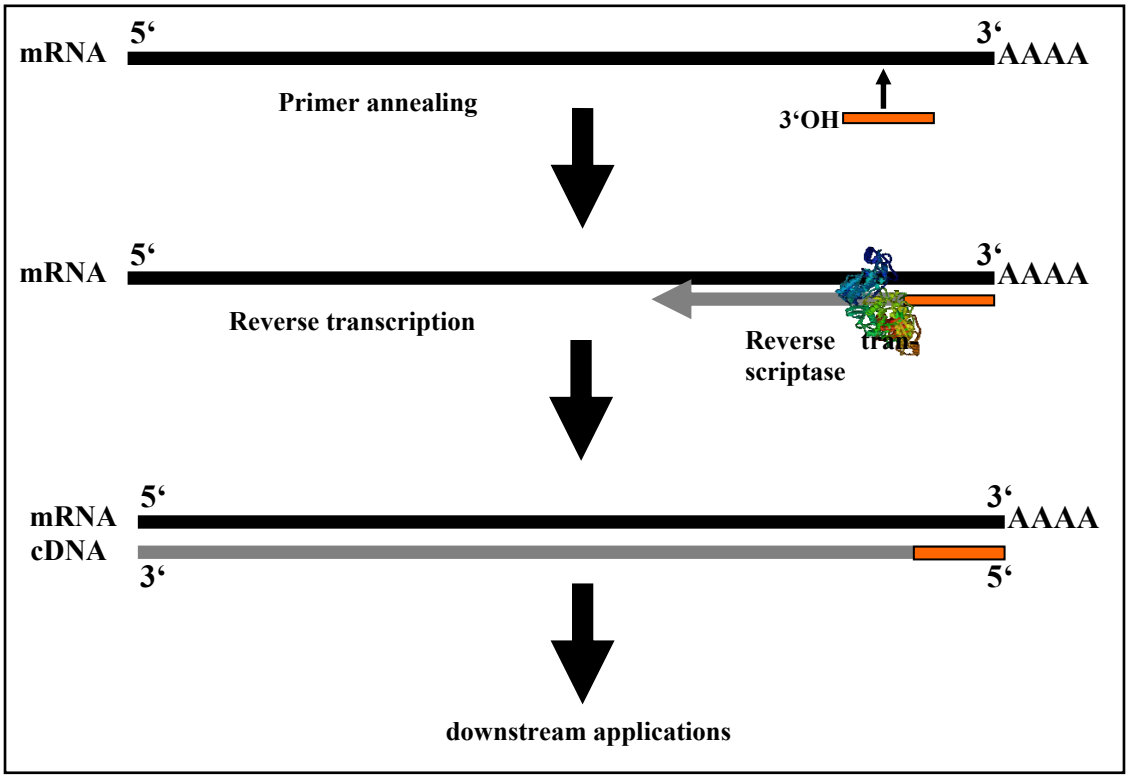


Fig. 9: Scheme of reverse transcription.

An oligo(dT)-primer or a gene-specific downstream primer with a free 3'-OH group binds to its complementary target sequence into the mRNA, where the reverse transcriptase binds (Fig. 9). The single stranded RNA serves as template and by introducing the four corresponding dNTPs the cDNA (complementary DNA) is synthesized (reverse transcription). This cDNA strand can now be applied in further downstream applications.

As enzymes the following reverse transcriptases were used, following the manufacturer's instructions:

- **Superscript III** (Invitrogen, Karlsruhe)
- **Improm II** (Promega, Mannheim)

5.5 RT-PCR

RT-PCR combines cDNA synthesis from RNA templates with PCR. The template for RT-PCR can be total RNA or poly $(A)^+$ selected RNA. Reverse transcription reactions can be primed with a random primer, an oligo (dT) or a gene-specific primer (GSP) using a reverse transcriptase. RT-PCR can be carried out either in two-step or one-step format. In two-step RT-PCR, each step is performed under optimal conditions. cDNA synthesis is performed first in RT buffer and a portion of the reaction is removed for PCR. In one-step RT-PCR, reverse transcription and PCR take place sequentially in a single tube under conditions optimized for both RT and PCR.

The applied reverse transcriptases and DNA-polymerases for two-step RT-PCR are listed under *5.4 Polymerase chain reaction (PCR) and its variations* and *5.5 Reverse transcription*. For one-step RT-PCR the following kits were used:

- **Qiagen® One step RT-PCR** (Qiagen, Hilden), a combination of Omniscript and Sensiscript Reverse Transcriptases for RT and HotStarTaq DNA Polymerase for PCR reaction
- **RobustT™ I RT-PCR Kit** (Finnzymes, Espoo, Finnland), first strand synthesis with AMV reverse transcriptase, second strand synthesis with DyNAzyme™ EXT DNA Polymerase

The reactions were set up following the manufacturer's instruction.

5.6 PCR clean up

For cleaning PCR products three different clean up kits were used:

- Charge switch® PCR clean up kit (Invitrogen, Karlsruhe)
- E.Z.N.A cycle pure Kit (peqlab, Erlangen)
- Nucleospin Extract II (Machery-Nagel, Düren)

The Charge switch® PCR clean up kit is based on a magnetic bead technology providing a switchable surface that is charged, dependent on the surrounding buffer pH to facilitate nucleic acid purification. In low pH conditions, the ChargeSwitch® Magnetic Beads have a positive charge and bind the negatively charged nucleic acid backbone. Proteins and other contaminants are not bound; they are washed away using the wash buffer. To elute nucleic acids, the charge on the surface is neutralized by raising the pH to 8.5 using a low salt elution buffer.

The E.Z.N.A cycle pure kit as well as the Nucleospin Extract II Kit uses HiBind® technology to recover DNA bands of 50 bp to 40 kb free of oligonucleotides, nucleotides and

polymerase. Binding conditions are adjusted by addition of a specially formulated buffer, and the sample was applied to a HiBind® DNA spin-column. Following a wash step, and after drying the column, the DNA was eluted with deionized water (or low salt buffer) and ready for downstream applications.

5.7 DNA gel electrophoresis

DNA fragments can be separated according to their size by using gel electrophoresis. Negatively charged DNA molecules migrate to the anode according to their molecular mass. To visualize the DNA within the gel, ethidium bromide is added to the agarose. This molecule is incorporated between the nucleotides of the two DNA strands and fluoresces purple when illuminated with UV-light at a wavelength of 254 nm. Because ethidium bromide is a strong carcinogen and toxic, alternatively, SYBRsafe™ (Invitrogen, Karlsruhe) was used to stain the agarose gels, which has a green fluorescence at the same wavelength. The result of DNA separation by gel electrophoresis is visible as distinct fluorescent bands within the gel.

- Solutions

10 x TBE buffer:	890 mM Tris/HCl pH: 7.5 890 mM boric acid 20 mM Na_2EDTA
Electrophoresis buffer:	1 x TBE
6 x loading dye:	50 % glycerine 0.4 % bromophenol blue
Gel matrix:	0.8 % agarose in 1 x TBE
Ethidium bromide stock solution:	10 mg/ml
SYBRsafe stock solution:	10,000 x conc.

- Documentation and evaluation

The analysis of the DNA containing gels was performed using UV light (UV-Transilluminator, Bachhofer, Reutlingen). The gels were photographed with a CCD-camera and documented with a “Video Copy Processor” (Mitsubishi). This method enables visualization of bands containing only 20 - 50 ng of DNA. Gels with weakly detectable bands were stained again for 20 minutes in 100 ml 1 x TBE containing 0,5 µg/ml ethidium bromide or 5 µl 10000 x SYBRsafe, respectively.

To analyse the size of the separated DNA bands, a marker was also applied on the gel. This marker contains distinct DNA fragments of known size (Fig. 10). The comparison of the migration pattern of the marker bands and the samples enables conclusions to be drawn as to the molecular size of the fragments.

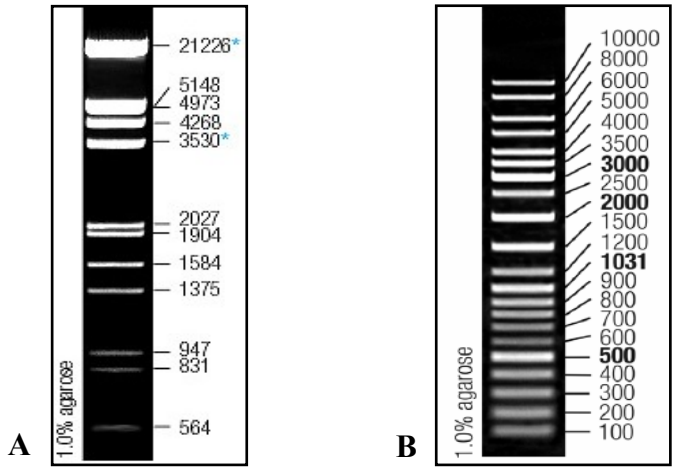


Fig. 10: DNA markers used in this work. **A**: Lambda DNA/ EcoRI+HindIII; **B**: GeneRuler™ DNA Ladder Mix, both from Fermentas, St. Leon-Roth.

5.8 DNA extraction from agarose gels

- Gel extraction Kits

To purify the DNA fragments out of the gel matrix the method of gel extraction is used. The following kits for gel extraction were applied and handled following the manufacturer's instructions:

- **PureLink™ Gel Extraction Kit**, Invitrogen (Karlsruhe)
- **Perfectprep Gel Cleanup Kit**, Eppendorf (Hamburg)
- **PCR clean-up Gel extraction**, Macherey-Nagel (Düren)
- **Qiaquick Gel Extraction Kit**, Qiagen (Hilden)

Each of these kits is based on the reversible binding of DNA to a silica gel matrix. The DNA bands of interest were cut from the gel matrix and dissolved in the appropriate buffer by heating and melting the agarose slice. This mixture was applied on a centrifugation column, where the DNA binds and in an additional centrifugation step, the contaminations were washed away. After drying the column, the DNA was eluted in adequate buffer or deionized water.

The concentration of the eluted DNA was measured at 260 nm using a Biophotometer from Eppendorf (Karlsruhe). The conversion of the absorption into the corresponding concentration was done automatically (1 OD_{600} corresponds to 50 µg/ml double stranded DNA).

- Freeze and squeeze

An alternative way to extract DNA from agarose gels, other than with the above mentioned kits, is to use the so called “freeze and squeeze” method, available on the web site http://www.biotech-europe.de/rubric/tricks/index.html. The directions given here were applied. The excised piece of agarose containing the desired DNA molecule was placed in the middle of a piece of parafilm, which was than folded to include the agarose. After placing this package for 30 min at - 20 °C, the piece was slightly squeezed between thumb and forefinger to thaw the frozen agarose. The emerging fluid drop contained the desired DNA and was collected in a reaction tube.

5.9 Digestion of nucleic acids with S1 nuclease

To remove overhanging single-stranded nucleic acids from a double-stranded DNA, digestion with S1 nuclease (Fermentas, St. Leon-Roth) was applied according to the manufacturer’s instructions. S1 Nuclease, a single-strand specific nuclease, degrades single-stranded nucleic acids, releasing 5'-phosphoryl mono- or oligonucleotides and creating blunt end double-stranded DNA.

5.10 Phenol/chloroform extraction of nucleic acids

The phenol/chloroform extraction is used to remove proteins and single nucleotides out of a nucleic acid solution.

For this purpose the same volume of a 1:1 phenol/chloroform mixture was added to the contaminated sample and mixed thoroughly by vortexing and then centrifuged at 14,000 x g for 2 minutes. Two different phases appear, the milky interphase containing the proteins and the upper aqueous phase the nucleic acids, which was transferred into a fresh tube. The remaining phenol was removed using alcohol precipitation. Therefore 1 volume of chloroform/isoamyl-alcohol (1:24) was added, the mixture was vortexed and again centrifugated for 2 minutes at 14,000 x g. To make sure that all contaminating proteins are removed, this step was repeated.

5.11 Cloning

This method is used to integrate a DNA fragment into a cloning vector (ligation), which is amplified in competent host bacterial cells (transformation). The advantage is to have a large number of copies of this insert due to the high reproductivity of bacteria. After the transformation the clones can be checked either by clone PCR or, after plasmid isola-

tion, by digestion with the appropriate restriction enzymes, if the insert of interest has been incorporated.

Five different cloning methods were used in this work:

- **pGEM-T Easy** (Promega, Mannheim)
- **QIAExpressionist** (Qiagen, Hilden)
- **TOPO-TA** and **TOPO-XL Cloning Kit** (Invitrogen, Karlsruhe)
- **Gateway™ Technology** (Invitrogen, Karlsruhe)
- **pENTR Directional TOPO™ Cloning Kit** (Invitrogen, Karlsruhe)

- Ligation

Ligation was used with two different cloning systems: the pGEM-T Easy (Promega, Mannheim) and the QIAExpressionist (Qiagen, Hilden).

With the first system the isolated DNA band was ligated into the pGEM-T Easy vector using T4 DNA Ligase. To counteract a self-ligation of the vector, the insert must be added in a much higher molar ratio (minimum 1:3).

With the second system, the vector and the DNA fragment were digested previously with the same restriction enzymes, creating either blunt ends or corresponding sticky ends (refer to: *5.16 DNA digestion using restriction endonucleases*).

Irrespective of the applied system the ligation step requires more of the enzyme ligase (Roche, Ingelheim), which catalyses the linkage of free 5' phosphate groups with free 3' hydroxyl groups under ATP-consuming conditions.

The ligation assay was left for 1 hour at room temperature or overnight at 4°C.

- TOPO cloning

This technique is a fast and efficient method of cloning and there is no requirement for ligase. Instead, a different mechanism is used: the topoisomerase I binds to the double-stranded DNA of the vector and cleaves it after the specific sequence 5'-CCCTT-3'. Binding between the 3' phosphate group of the cleaved strand and the hydroxyl group of a tyrosine of the topoisomerase is generated, which can be affected by the 5' OH group of the PCR fragment. In this way, the desired insert can be ligated into the vector (Fig. 11).

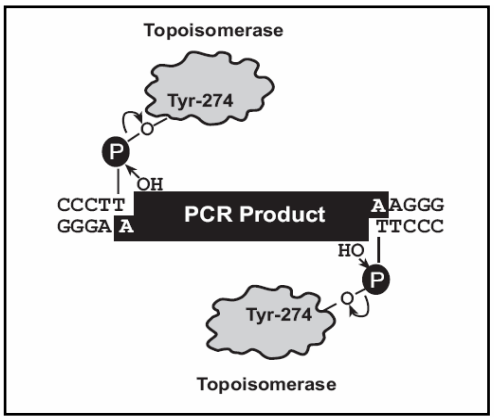


Fig. 11: Scheme of a TOPO-cloning reaction. (From: http://www.invitrogen.com/content/sfs/manuals/topota_man.pdf).

For this cloning two different kits were used: firstly the TOPO-TA Cloning Kit, and secondly the TOPO-XL Cloning Kit, which was particularly adapted for larger fragments (> 2 kb). Both vectors contain a 3' overhanging thymidine, which provides the advantages on the one hand ligation with the 3' overhanging adenosine of a *Taq*-amplified PCR product, and on the other hand prevents the circularisation of the vector. The cloning reaction was done following the manufacturer's instruction, however, only the half of the specified amount of vector and bacterial cells were used.

In addition to these two cloning techniques, a third technique was used: the Gateway™ Technology from Invitrogen (Karlsruhe), which is based on a recombination reaction between vector and insert. Because most cloning reactions were done using this method, it is explained in a separate section (refer to: *7.1 Gateway™ Technology*).

- Cloning without UV damage

As it is well known, UV light can damage DNA. The consequences are single strand breaks within the double stranded DNA molecule and the generation of smaller degradation products. This, in turn, could lead to the fact that these smaller degradation products will be cloned more efficiently into the vector than the desired molecule. According to a protocol found on the web site http://www.biotecheurope.de/rubric/ tricks/index.html, this problem can be by-passed.

This protocol was applied following the instructions given. After completing the PCR reaction, the amplified DNA was applied into the first and the second pocket of the agarose gel, whereas approximately 70 % of the sample was loaded into the first pocket and only 20 % into the second one. The next lane contained the marker. After running the gel, the first lane was excised and set aside. The remaining gel was viewed under UV light and a photo was taken. The location of the desired band appearing within

the second lane was assigned by incision above and under the band and the UV light was switched off. After putting back the first lane accurately next to the remaining gel, the incision was extended across the first lane. The piece of gel containing the desired band was excised from the first lane, extracted using the common agarose gel extraction methods and cloned.

- Transformation

During the transformation, DNA molecules occurring as plasmids are introduced by heat shock into competent bacterial cells (refer to *4.3 Bacterial cultures*, - Chemically competent bacterial cells). The ligation or cloning reaction is added to 25 µl of competent cells and left on ice for 30 minutes. Subsequently, heat shock at 42 °C (30 seconds) is the best way to induce up-take of the plasmid. For recovery and expression of the selection marker the bacteria were incubated shaking for 1 - 2 hours at 37 °C in 200 µl of SOC-medium. After this incubation the sample was plated on Agar plates containing the appropriate antibiotic and these were incubated overnight at 37 °C.

- Lethal gene selection

The morning after the transformation, the bacterial cells can be selected following the lethal gene expression.

By producing a toxic substance due to the *ccdB* gene, this selection is lethal for the bacterial cells when not substituted by the ligated insert. This produces positive clones, because the insert-less ones die due to the *ccdB* gene.

For the selection of bacteria without a plasmid, the transformation is plated on Agar plates containing either kanamycin or ampicillin, depending on the vector and its resulting resistance to antibiotics. Positive clones were picked using a pipette tip and first separately transferred into a clone-PCR sample (refer to *5.4 Polymerase chain reaction (PCR) and its variations*) by pipetting up and down and afterwards the whole pipette tip was ejected into 5 ml LB-medium containing the adequate antibiotics for generating an overnight culture (refer to: *4.3 Bacterial cultures*).

5.12 Plasmid isolation

For the isolation of plasmid DNA from bacterial cells the E.Z.N.A plasmid Miniprep Kit I from Peqlab (Erlangen) was used, following the manufacturer's instructions. This kit combines a modified, alkaline lysis of the cells and precipitation of bacterial proteins

and genomic DNA with the selective and reversible DNA binding abilities of HiBind® spin columns.

2 ml of an overnight culture were spun down at 13,000 x g for 2 minutes and the pellet was resuspended in an alkaline lysis buffer. A second solution helps neutralizing the sample, and the addition of a third solution causes the precipitation of bacterial proteins and cell debris. The following centrifugation step pellets this precipitated material, and the clear supernatant contains the plasmid DNA, which is applied to the column. The DNA binds reversibly to the silica matrix and after washing and drying the column the DNA was eluted in deionized water.

5.13 DNA digestion using restriction endonucleases

This technique was used in the present work for two approaches: the insertion of a DNA fragment into a vector as well as the excision of a DNA fragment out of a vector.

- Insertion of a DNA fragment into a vector: the digestion of the DNA fragment and the vector, prior to ligation, creates blunt ends or corresponding sticky ends, depending on the used restriction endonuclease.

A piece of target DNA can be inserted into a plasmid if both the circular plasmid and the target DNA have been cleaved by the same restriction nuclease in such a way as to create sticky ends. The newly created recombinant molecule is stabilized with the DNA ligase enzyme, which repairs nicks in the backbone of the DNA molecule. The same ligation can be done if both DNA strands were digested resulting in blunt ends.

The pQE 30 vector from Qiagen (Hilden) was designed to have different restriction sites incorporated only at the multiple cloning site. The desired insert, after adding the restriction sites by PCR at the 5' and 3' end of the fragment, was also digested with the chosen endonucleases to have corresponding sticky ends to the digested vector and then both were ligated using T4 DNA ligase. For this approach, depending on the existing restriction sites, only one enzyme needs to be applied, that cuts two times. Also, two endonucleases can be combined, to enable uni-directional cloning (Fig. 12).

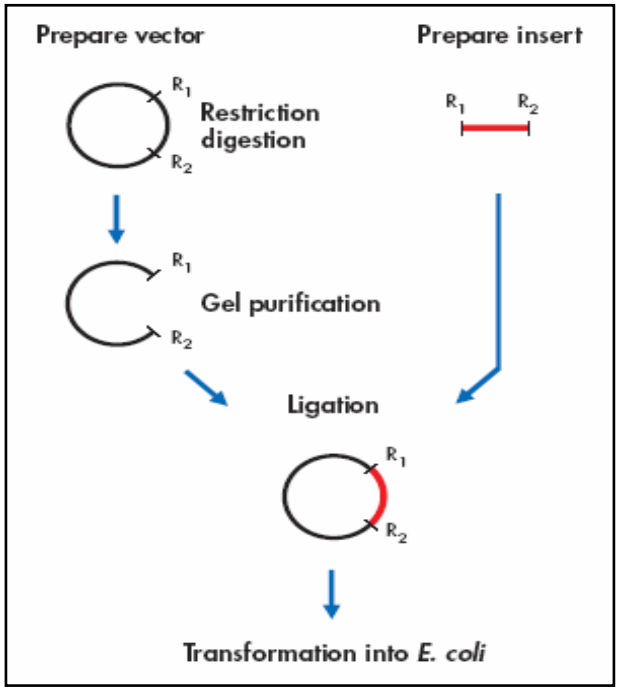


Fig. 12: Scheme of the insertion of a DNA fragment into a vector after digestion of both by restriction enzymes. R_1, R_2: restriction sites (From: QIAexpressionist 06/2003).

However it is clear that the DNA fragment should not contain a restriction site for the chosen enzyme. This ligation reaction is then transformed into competent bacterial cells. The applied restriction enzymes were HindIII and KpnI from Fermentas, St Leon-Roth.

- Excision of a DNA fragment from a vector: the digestion of an insert containing plasmid to analyse the size.

The excision of an insert from the plasmid using restriction enzymes is an alternative method to clone-PCR for the analysis of the correct size. The isolated plasmid is digested using endonucleases that cut the vector within the multiple cloning site. The result can be visualized by agarose gel electrophoresis, where two distinct bands should appear if the enzyme cuts only at these two restrictions sites: one for the linearized vector and one for excised insert. If the sequence of the insert is known, one can check if the proposed endonuclease also cleaves within the insert and then calculate the size of the resulting bands. If not, the addition of the apparent different sized bands, except the one for the vector, should result in the original size of the expected complete insert size.

As a control, the vector per se and a sample of digested vector alone should always be applied simultaneously on the gel. EcoRI (Fermentas, St Leon-Roth) was the chosen endonuclease for this approach.

6. Routine protein biochemical methods

6.1 Determination of hemocyanin concentration and absorption spectrum

To determine the concentration of native hemocyanin an absorption spectrum was produced from 220 nm to 600 nm in an Ultraspec 3100 Pro UV/Visible Spectrophotometer. For the native hemocyanin, two characteristic absorption maxima occur, at 280 nm and 340 nm, respectively. The first maximum is typical for proteins, due to the absorption of aromatic amino acids such as tryptophan and tyrosine. An OD_{600} of 1 corresponds to a protein concentration of ~ 1mg/ml. The oxygenated active centre of hemocyanin (Cu-O_2-Cu-complex) has an absorption maximum at 340 nm and therefore provides evidence for the quantitative presence of hemocyanin, when measured at this wavelength.

6.2 Dialysis

The principle of dialysis is based on the semi-permeable properties of membranes or dialysis tubes to exclude specific substances from a solution. Due to the pore size of the membrane or the tube, only small molecules below the molecular mass cut off value can pass through the pores of the membrane. Larger proteins are retained within the dialysis tube. This technique is generally used to remove undesirable small molecules from a protein solution or to change the buffer.

In this work dialysis was used to remove 8 M urea from protein samples by dialysis against 50 mM Tris buffer, pH 8.0. Samples were placed into dialysis tubing, molecular mass cut off of < 10,000 (Visking, Art. 0653, Typ 20/32; Roth, Karlsruhe), which had been boiled for 10 minutes in bidistilled water prior to use. The dialysis tube was then positioned in 2 litres of 50 mM Tris buffer. After two or three renewals of the outer buffer and after 24 hours of constant stirring, the urea will be completely removed.

6.3 Polyacrylamide gel electrophoresis (PAGE)

Proteins can be separated within an electrical field due to their varying mass and net charge. The most commonly used support matrices - agarose and polyacrylamide - provide a means for separating molecules by size, in that they are porous gels. A porous gel may act as a molecular sieve by retarding, or in some cases completely obstructing the movement of large macromolecules, while allowing smaller molecules to migrate freely. Two types of buffer systems are used in electrophoresis, continuous and discontinuous. A continuous system has a single separating gel and uses the same buffer in the tanks and throughout the gel. In a discontinuous system, a non-restrictive large pore gel,

termed the stacking gel, is layered on top of a separating gel, termed the resolving gel. Each gel is made with a different buffer, and the tank buffers are different from the gel buffers. Due to the stronger concentration of the samples by using a stacking gel, the resolution obtained in a discontinuous system is much greater than that obtained with a continuous system, so only the former system was used. In the presence of the radical former ammonium persulfate (APS) and the catalyst tetramethylendiamine (TEMED), acrylamide can polymerize into long chains, which are linked to form a net-like porous gel structure by N,N'-methylenbisacrylamide. The size of the pores of the matrix is determined by the concentration of acrylamide and N,N'-methylenbisacrylamide, so that the range of separation can be selected, as required. Table 1 shows the correlation between the concentration of the gel and the molecular mass separation range of proteins.

Polyacrylamide conc. in %	Separation range in kDa
3	100 - 1000
5	80 - 500
8	60 - 400
12	40 - 200
20	10 - 100

Tab. 1: Separation patterns of different polyacrylamide gel concentrations.

For performing PAGE, vertical electrophoresis tanks were used for minigels (7 x 10 cm) and for large gels (14.5 x 18 cm). Separation of the samples was done in 10 % resolving gels, thickness of 1 mm, overlaid with a 3.9 % stacking gel. In the stacking gel, a plastic comb was inserted before polymerisation, to create the sample pockets.
PAGE was also carried out under denaturing conditions, in a similar manner to non-denaturing (native) conditions, following the procedure of Laemmli (1970).

- SDS-PAGE

SDS-PAGE has a number of uses, which include the establishment of protein size, protein identification, determination of sample purity, identification of disulfide bonds, quantification of proteins and blotting applications.
SDS, sodium dodecyl (lauryl) sulphate, is an anionic detergent that binds quantitatively to proteins, giving them linearity and uniform negative charge, so that they can be separated solely on the basis of their size. The SDS has a high negative charge that overwhelms any charge of the proteins, giving them an approximately equal negative charge. The number of SDS molecules that bind to a protein is proportional to the number of amino acids in the protein. SDS also disrupts the forces that contribute to protein folding

(tertiary structure), ensuring that the protein is not only uniformly negatively charged, but then linear. SDS is present in the loading dye, the gel and the running buffer. All proteins migrate to the anode. Furthermore, β-mercaptoethanol is also present in the loading dye, so that the tertiary structure of the proteins is destroyed by cleavage of disulfide bonds, thus breaking the only known covalent bond between the polypeptide chains. The following denaturation at 95 °C for 5 minutes is responsible for an unfolded state and an elongated conformation of the proteins. The migration distance during the electrophoresis is then linearly dependent on the logarithm of the molecular size (Weber & Osborn, 1969).

To be able to determine the size of the proteins in relation to the migration distance, a calibration marker containing several known proteins was also applied to the gel. For minigels 5 µl and for large gels 15 µl of the marker SDS-6H (Sigma, Deisenhofen; Table 2) is used.

Protein	Molecular weight (kDa)
Myosin	205
β-Galactosidase	116
Phosphorylase b	97.4
Bovine serum albumin (BSA)	66
Ovalbumin	45
Carboanhydrase	29

Tab. 2: Composition of the protein marker SDS-6H and the respective molecular masses of the calibrating proteins.

- Native PAGE

While the separation of proteins during an SDS-PAGE exclusively occurs due to their molecular size, during a native PAGE, untreated proteins are separated electrophoretically due to their molecular size combined with their net charge and conformation (tertiary structure). The system of Laemmli (1970) was used for gel production. Note that the buffers contained neither SDS nor β-mercaptoethanol, and that the sample denaturation step at 95 °C was omitted.

- Set up of the gels

Solutions used for both systems, SDS and native PAGE:

Acrylamide (30 % (w/v) Acrylamide, 0.8 % (w/v) Bisacrylamide (Protogel-ready-to-use-solution, Helmut Schröder, Stuttgart)

Polymerization starter: 10 % (w/v) APS in dist. H_2O

	TEMED
Staining solution:	0.2 % (w/v) Coomassie Brilliant Blue R 205
	40 % methanol
	20 % acetic acid (60 %)
Decolouring solution:	20 % isopropanol
	7 % acetic acid (60 %)

Solutions used depending on the system:

	SDS-PAGE	**Native PAGE**
Resolution gel buffer	1.5 mol/l Tris/HCl 0.4 % SDS pH 8.8	1.5 mol/l Tris/HCl pH 8.8
Stacking gel buffer	0.5 mol/l Tris/HCl 0.4 % SDS pH 8.8	0.5 mol/l Tris/HCl pH 6.8
Running buffer	0.023 mol/l Tris/HCl 0.19 mol/l Glycine 0.2 % SDS pH 8.3	0.023 mol/l Tris/HCl 0.19 mol/l Glycine pH 8.3
Loading dye (4 x)	20 % Glycerine 8 % SDS 36 % Stabilizing buffer 4 % β-Mercaptoethanol 1 Spatula tip of Bromophenol blue	90 % Glycerine 10 % Stabilizing buffer 1 Spatula tip of Bromophenol blue

Tab. 3: System specific solutions for SDS and native PAGE.

Before pouring the gels, two glass plates with three side-spacers inserted inbetween, were held together by two or four clamps. All sides except the upper one were sealed with 1 % agarose. The resolving gel was mixed and flowed between the two glass plates, until 1cm below the upper edge and than overlaid immediately with distilled water. After polymerization of the resolving gel, the stacking gel was mixed and applied, and a plastic comb was promptly inserted, without air bubble formation. After this second gel had polymerized, the clamps were removed and the gel was inserted into the electrophoresis tank (PHERO-minivert, Biotec Fischer, Reiskirchen). The upper and lower tanks were filled with running buffer and the comb was removed. The protein samples were mixed with a loading dye (bromophenol blue) containing glycerol, to increase the sample density, and injected into the pockets created by the comb. The gels were electrophoresed at 75 - 100 mA for minigels and at 175 - 200 mA for large gels, until the bromophenol blue

tracker dye reached the bottom of the gel. After completion of the electrophoresis the gel was either used for protein transfer onto a nitrocellulose membrane (refer to: *6.5 Western blot*) or incubated with shaking for one hour in staining solution and then destained, with shaking for an additional hour. For long-term storage the gels were kept in 7.5 % v/v acetic acid.

Gel	**Resolving gel**	**Stacking gel**
percentage	**10 %**	**3.9 %**
Rothiphorese® [ml]	1.6	0.2
Resolving gel puffer [ml]	1.25	-
Stacking gel puffer [ml]	-	0.4
H_2O [ml]	2.06	1
TEMED [µl]	1.6	1.6
APS 10 % [µl]	50	50

Tab. 4: Pipetting scheme for 1 minigel, for SDS or native PAGE. For a large gel, a six-fold increase of the minigel volume was used.

6.4 Western blot

This blotting technique is used to detect electrophoretically separated proteins by transferring them onto a nitrocellulose membrane. Protein detection is done by polyclonal or monoclonal antibodies, directed against a known protein.

- Transfer of electrophoresed proteins in an SDS-PAGE

Sufficiently separated proteins in an SDS-PAGE can be transferred and immobilized on a solid membrane for Western blot analysis. To make the proteins accessible to antibody detection, they are electrophoretically moved from within the gel onto a membrane made of nitrocellulose (blotting). The membrane is placed face-to-face with the gel, and current is applied to large plates on either side. The charged proteins move from within the gel onto the membrane while maintaining the same position they had within the gel. As a result of this "blotting" process, the proteins are attached as a thin surface layer for detection (see below: - Detection and analysis). This nitrocellulose membrane is chosen for its non-specific protein binding properties (i.e. it binds all proteins equally well). Protein binding is based upon hydrophobic interactions, as well as charged interactions between the membrane and protein. After blotting, the proteins on the membrane can be stained

reversibly with a 1:10 dilution of Ponceau S solution (Sigma, Deisenhofen) and the stained protein bands can be scanned for documentation. The red colour can be removed by washing with bidistilled water.

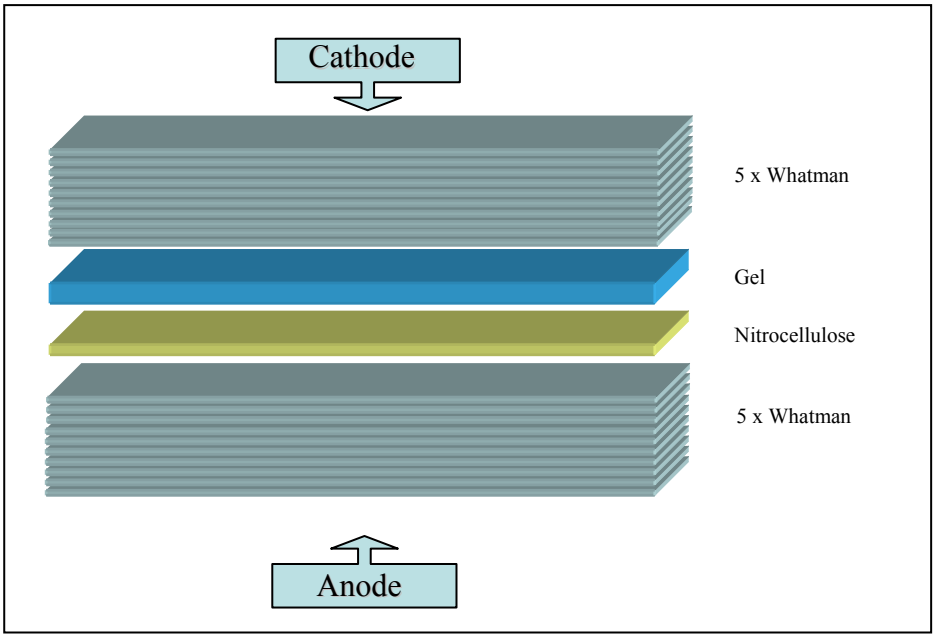


Fig. 13: Assembly of a western blot (From: PhD thesis, Sandra Bergmann, 2004).

In this work, the *semi-dry* electroblotting method (Khyse-Andersen, 1984) was used, where sheets of Whatman papers were soaked in transfer buffer to facilitate the current flow. These sheets were placed above and under the gel-membrane assembly, which were also placed in this buffer for a short time before blotting (Fig. 13).

Transfer buffer: 250 mM Tris pH 8.3
1.5 M Glycerol
10 % Methanol

The electrotransfer of the proteins was performed at 0.8 mA/cm^2 gel area for 2 hours.

- Transfer of electrophoresed proteins in a native PAGE

Untreated proteins applied on a native PAGE are separated electrophoretically due to their molecular size combined with their net charge and conformation (tertiary structure). Therefore it is not possible to transfer them onto a nitrocellulose membrane by applying current. In this work the wet press-blot method was used, where the protein transfer occured only due to capillary forces. The driving force is a fluid stream which flows from the buffer reservoir through the gel and thereby moves the proteins onto the nitrocellulose membrane. Sheets of Whatman papers and sponges above and under the

native acrylamide gel were soaked with transfer buffer, wherefore nitrocellulose membranes can be placed on both sides of the gel. To exert pressure, this assembly was then placed between two heavy stone plates, additionally to a compression by four vices.

Transfer buffer: 130 mM Glycin, pH 9.6

The transfer of the proteins was performed for at least two hours or overnight.

- Blocking

Once the proteins had been transferred onto a nitrocellulose membrane, independently if theywere derived from a native gel or an SDS gel, steps had to be taken to prevent non-specific protein interactions between the membrane and the antibody used for detection of the target protein. Blocking of non-specific binding is achieved by placing the membrane in a diluted solution of proteins - typically 3 % w/v bovine serum albumin (BSA) or 5 % w/v non-fat milk powder, in TBS or TBST buffer, respectively. The choice of the blocking reagent depends on the compatibility with the antibodies used.

TBS (per litre) 10 mM Tris/Cl
140 mM NaCl

TBST (per litre) 10 mM Tris/Cl
140 mM NaCl
0.1 % Tween 20

- Detection and analysis

The aim of the detection process is to probe the membrane for the protein of interest with antibodies, and to link them to a reporter enzyme, which produces a colorimetric or photometric signal. After blocking, a dilute solution of primary antibody (generally between 0.5 and 5 μg/ml) is incubated with the membrane under gentle agitation. Typically, the solution comprises buffered saline solution with a small percentage of detergent, sometimes including powdered milk or BSA. After rinsing the membrane to remove unbound primary antibody and in order to detect the antibodies which have bound specifically, anti-immunoglobulin antibodies are added. This anti-Ig-enzyme is commonly termed the secondary antibody or conjugate. The secondary antibody is usually linked to biotin or a reporter enzyme such as alkaline phosphatase or horseradish peroxidase. This step confers an advantage in that several secondary antibodies will bind one primary antibody, providing an enhanced signal. Finally, after excess secondary antibody is washed off the blot, a substrate is added which will yield a precipitate upon

reaction with the conjugate, resulting in a visible band where the primary antibody has bound to the protein.

The antibodies used for this work and the explicit procedures will be given at the appropriate place in Chapter C, Results.

- Dot blot

This blotting technique is carried out by the same procedure than in Western blot, despite the fact that the protein samples are not transferred from a gel onto a membrane by applying electric current; they are spotted directly onto the membrane and then detected with a primary and a secondary, labelled antibody. This very rapid method is used to test different antibody concentrations for a specific antigen analyzed by Western blot without the lengthy procedure of running a gel and transferring the proteins onto a membrane.

6.5 Two dimensional immunoelectrophoresis

With this method, which was performed following the procedures of Laurell (1965) and Weeke (1973), soluble proteins can be characterized immunologically. By determining the different immuno-precipitation patterns of the proteins, one can draw conclusions about their structural relationship, one to another. Furthermore, by comparing the different height of the precipitation peaks, information is given on the protein concentration. Indeed, in the past the procedure has been used for the quantitative determination of proteins (the Rocket technique).

In a 1 % w/v agarose gel, the protein mixture is separated electrophoretically within the first dimension, where the diverse native proteins migrate in different rates from the cathode to the anode, due to their different net charge. Due to the large pores of the gel (1 %), protein size and the tertiary structure of the proteins do not significantly influence the migration pattern. In the second dimension, the separated proteins are electrophoresed through an antibody-containing 1 % w/v agarose gel, also from the cathode to the anode. If there is an antigen present in the protein sample, an antigen-antibody complex formation occurs. These complexes migrate within the gel until their concentration is so high that a three dimensional immunoprecipitate is formed. At this point a characteristic peak of precipitated protein is formed, which can be visualized by staining.

In this work, crossed immunoelectrophoresis as well as crossed-line immunoelectrophoresis and rockets were kindly done by [redacted] wherefore the method is not explained in details.

6.6 Carbohydrate digestion via N-glycosidase

N-glycosidase F$^-$ or PNGase F is a glycopeptidase which cleaves asparagine-linked high mannose as well as hybrid and complex oligosaccharides from glycoproteins. It deaminates the asparagine to aspartic acid, but leaves the oligosaccharide intact (see Figure 14). A tripeptide with the oligosaccharide-linked asparagine as the central residue is the minimal substrate for PNGase F.

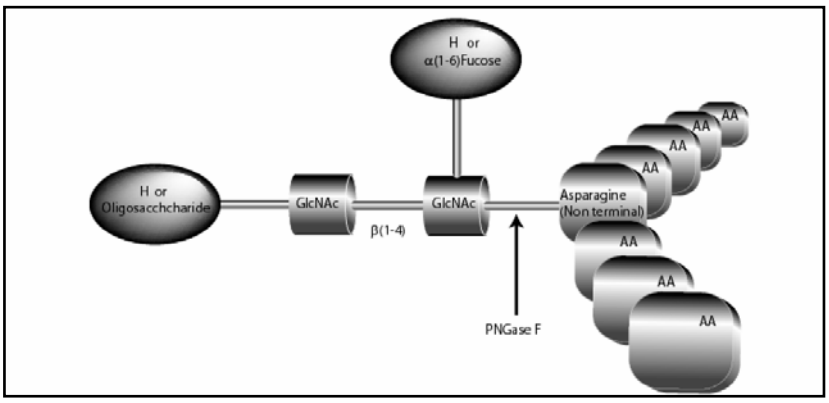


Fig. 14: Specificity of PNGase F (From: www.sigmaaldrich.com/sigma/datasheet/g5166dat.pdf).

In the present work the PNGase F from Sigma-Aldrich (Deisenhofen) was used exclusively. PNGase F is isolated from *Chryseobacterium (Flavobacterium) meningosepticum* and has a molecular weight of 36 kDa. It is active in the pH range of 6 to 10, but the optimal pH is 8.6. Detergent and heat denaturation are known to increase the rate of cleavage up to 100 times. Most native proteins can still be completely N-deglycosylated, but incubation time must be increased. Following the protocol of PNGase F (Sigma, Deisenhofen), three consecutive steps can be applied preceding the deglycosylation, to maximize the recovery. The first one is the addition of 10 µl phosphate buffer, pH 7.5, to up to 200 µg glycoprotein, followed by the incubation with PNGase F overnight at 37 °C. The second step includes, after the first one, the addition of 2.5 µl of 2 % SDS with 1 M β-mercaptoethanol. The third step includes the two previously mentioned ones followed by an addition of 2.5 µl 15 % Triton-X 100 and by a denaturation for 5 minutes at 95 °C, which should lead to the highest recovery of deglycosylated protein. PNGase F is known to remain active under incubation conditions for at least 72 hours.

For experiments where the protein has to be kept in its native conformation, the protocol was shorten to the following steps: up to 200 µg of the glycoprotein were employed in a maximum of 35 µl of water, and the addition of 10 µl of 250 mM phosphate buffer, pH 7.5, was followed by the addition of 2 µl of the PNGase F solution (7,690 Units/ml). This mixture was incubated at 37 °C for the maximum incubation time (72 hours), to

ensure the complete deglycosylation of the glycoprotein. The result was visualized by testing the protein with the DIG glycan detection Kit (Roche, Mannheim), which is described in the following section.

6.7 Glycan detection

To determine the glycosylation status of a protein or to assess whether or not carbohydrate side chains have been removed successfully from glycoproteins by PNGase F, the DIG glycan detection kit from Roche, Mannheim, was used. This kit is based on an enzyme immunoassay. Adjacent hydroxyl groups in sugars of glycoconjugates are oxidized to aldehyde groups by mild periodate treatment. The spacer-linked steroid hapten digoxigenin (DIG) is then covalently attached to these aldehydes via a hydrazide group. Digoxigenin labelled glycoconjugates are subsequently detected in an enzyme immunoassay, using a digoxigenin specific antibody conjugated with alkaline phosphatase (Fig. 15).

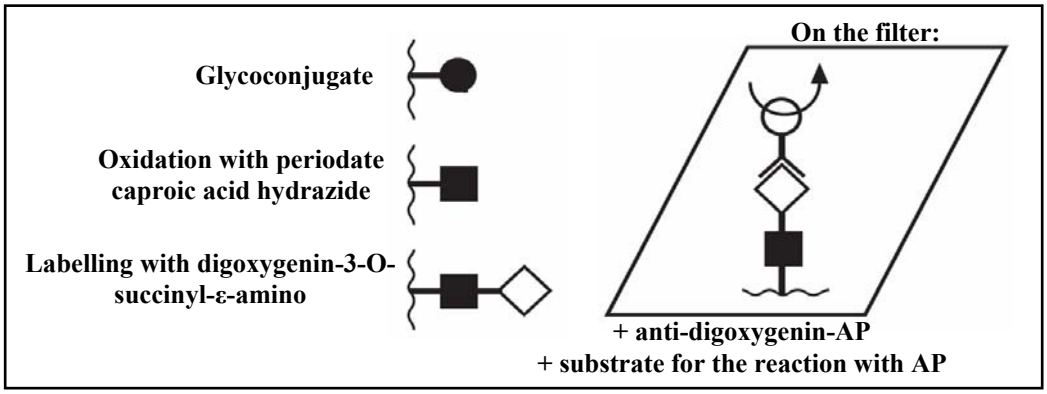


Fig. 15: Detection of glycoproteins and other glycoconjugates using the DIG glycan detection kit from Roche, Mannheim (Modified from: http://www.roche-applied-science.com/pack-insert/1142372a.pdf).

All procedures were performed following the manufacturer's instructions.

7 Recombinant protein expression in *E. coli*

7.1 Gateway™ Technology

The Gateway™ Technology from Invitrogen, Karlsruhe, is a universal cloning method based on the site-specific recombination properties of bacteriophage lambda (Landy, 1989). The Gateway™ Technology provides a rapid and highly efficient way to move DNA sequences into multiple vector systems for functional analysis and protein expression (Hartley *et al.*, 2000) (Fig. 16).

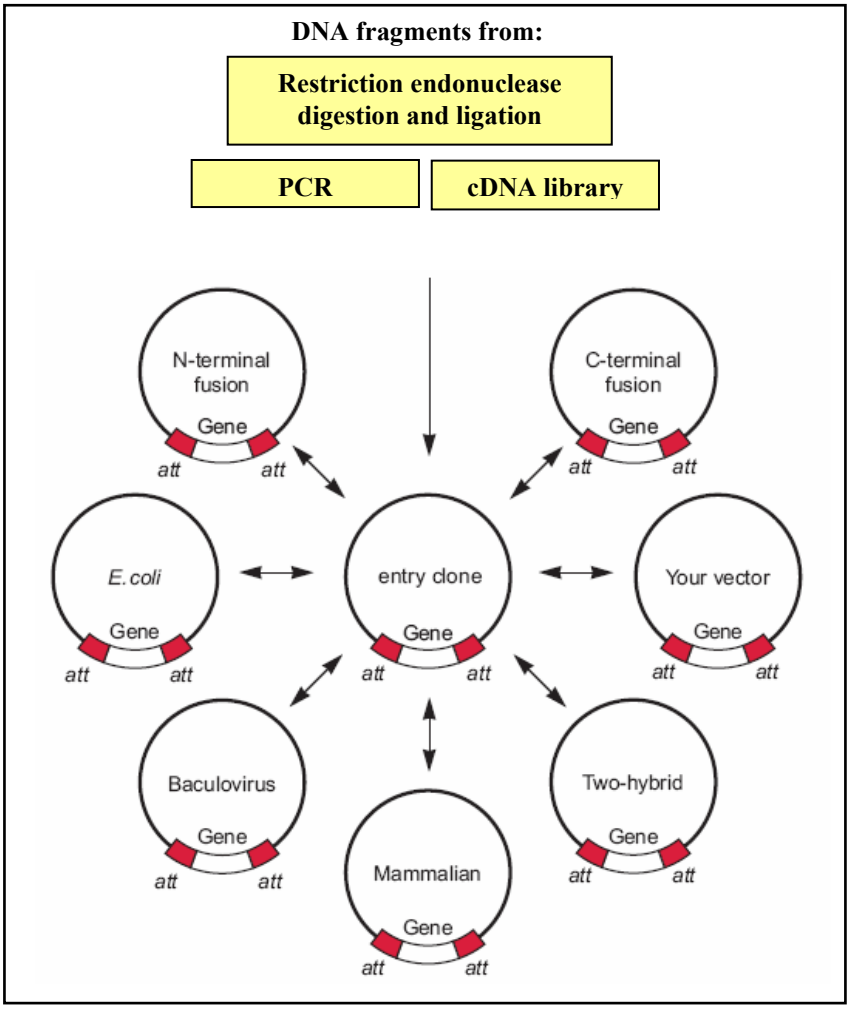


Fig. 16: Diagram of the Gateway™ Technology from Invitrogen, Karlsruhe (From: https://www.invitrogen.com/content/sfs/manuals/gatewayman.pdf).

The Gateway™ Technology provides the following theoretical advantages:

- Rapid and highly efficient transfer of DNA sequences into multiple vector systems for protein expression and functional analysis while maintaining orientation and reading frame.
- Permits the use and expression from multiple types of DNA sequences (*e.g.* PCR products, cDNA clones, restriction fragments).
- Easily accommodates the transfer of a large number of DNA sequences into multiple destination vectors.
- No need of restriction enzymes or ligase

- The basis of Gateway™ Technology

The Gateway™ Technology is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome

and the switch between the lytic and lysogenic pathways (Ptashne, 1992). In the Gateway™ Technology, the components of the lambda recombination system are modified to improve the specificity and efficiency of the system (Bushman *et al.*, 1985). Lambda integration into the *E. coli* chromosome occurs via intermolecular DNA recombination that is mediated by a mixture of lambda and *E. coli*-encoded recombination proteins (*i.e.* Clonase™ enzyme mix). The hallmarks of lambda recombination are listed below:

- Recombination occurs between specific attachment (*att*) sites on the interacting DNA molecules.
- Recombination is conservative (*i.e.* there is no net gain or loss of nucleotides) and requires no DNA synthesis. The DNA segments flanking the recombination sites are switched, such that after recombination, so that the *att* sites are hybrid sequences containing sequences donated by each parental vector. For example, *att*L sites contain sequences from *att*B and *att*P sites.
- Strand exchange occurs within a core region that is common to all *att* sites.
- The recombination can occur between DNAs of any topology (*i.e.* supercoiled, linear, or relaxed), although efficiency varies. For more detailed information about lambda recombination, see published references and reviews (Landy, 1989; Ptashne, 1992).

Lambda recombination occurs between site-specific attachment (*att*) sites: *att*B on the *E. coli* chromosome and *att*P on the lambda chromosome. The *att* sites serve as the binding site for recombination proteins and have been well-characterized (Weisberg and Landy, 1983). Upon lambda integration, recombination occurs between *att*B and *att*P sites to give rise to *att*L and *att*R sites. The actual crossover occurs between homologous 15 bp core regions on the two sites, but surrounding sequences are required as they contain the binding sites for the recombination proteins (Landy, 1989).

Lambda recombination is catalyzed by a mixture of enzymes that bind to specific sequences (*att* sites), bring together the target sites, cleave them, and covalently attach the DNA. Recombination occurs following two pairs of strand exchanges and ligation of the DNAs in a novel form. The recombination proteins involved in the reaction differ depending upon whether lambda utilizes the lytic or lysogenic pathway (Table 5). The lysogenic pathway is catalyzed by the bacteriophage λ Integrase (Int) and *E. coli* Integration Host Factor (IHF) proteins (BP Clonase™ enzyme mix) while the lytic pathway is catalyzed by the bacteriophage λ Int and Excisionase (Xis) proteins and the *E. coli* Inte-

gration Host Factor (IHF) protein (LR Clonase™ enzyme mix). For more information about recombination enzymes, see Landy (1989) and Ptashne (1992).

Pathway	Reaction	Catalyzed by
Lysogenic	*att*B x *att*P → *att*L x *att*R	BP Clonase™ (Int, IHF)
Lytic	*att*L x *att*R → *att*B x *att*P	LR Clonase™ (Int, Xis, IHF)

Tab. 5: Lambda pathways and the corresponding recombination proteins.

- Gateway recombination reactions

The Gateway™ Technology uses the lambda recombination system to facilitate transfer of heterologous DNA sequences (flanked by modified *att* sites) between vectors (Hartley *et al.*, 2000). Two recombination reactions constitute the basis of the Gateway™ Technology:

- **BP Reaction:** Facilitates recombination of an *att*B substrate (*att*B-PCR product or a linearized *att*B expression clone) with an *att*P substrate (donor vector) to create an *att*L-containing entry clone (see Fig. 17). This reaction is catalyzed by the BP Clonase™ enzyme mix.

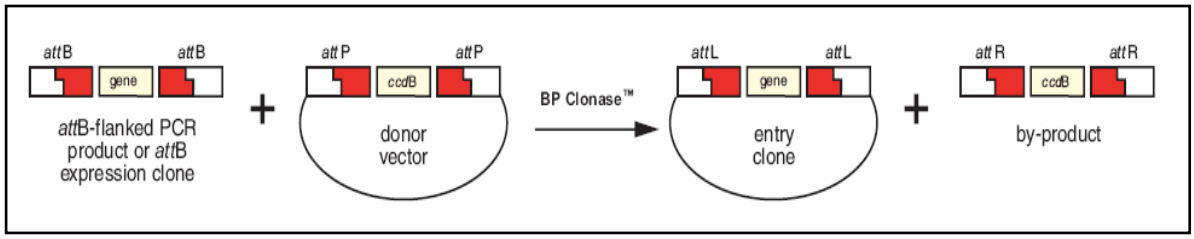


Fig. 17: Diagram of the BP reaction (from: https://www.invitrogen.com/content/sfs/ manuals/gatewayman.pdf).

- **LR Reaction:** Facilitates recombination of an *att*L substrate (entry clone) with an *att*R vector (destination vector) to create an *att*B-containing expression clone (Fig. 18). This reaction is catalyzed by the LR Clonase™ enzyme mix.

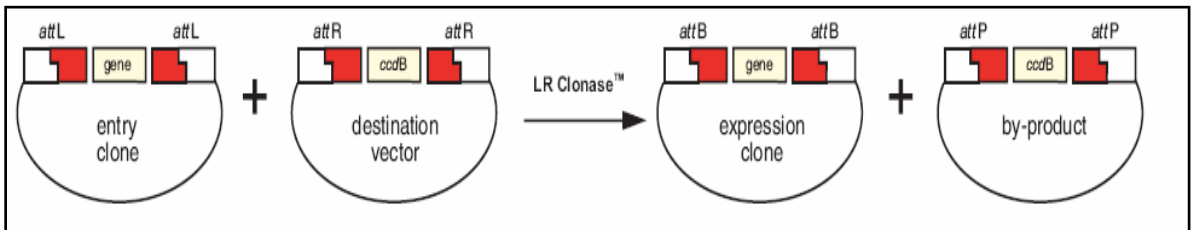


Fig. 18: Scheme of the LR reaction (from: https://www.invitrogen.com/content/sfs/ manuals/gatewayman.pdf).

- Common features of the Gateway™ vectors

To enable recombinational cloning and efficient selection of entry or expression clones, most Gateway™ vectors contain two *att* sites flanking a cassette containing:

- The *ccd*B gene for negative selection.
- Chloramphenicol resistance gene (CmR) for counter-selection.

After a BP or LR recombination reaction, this cassette is replaced by the gene of interest to generate the entry clone or expression clone, respectively.

The presence of the *ccd*B gene allows negative selection of the donor and destination vectors in *E. coli* following recombination and transformation. The CcdB protein interferes with *E. coli* DNA gyrase (Bernard and Couturier, 1992), thereby inhibiting growth of most *E. coli* strains (*e.g.* DH5α™, TOP10). When recombination occurs, the *ccd*B gene is replaced by the gene of interest. Cells that take up unreacted vector carrying the *ccd*B gene or by-product molecules retaining the *ccd*B gene will fail to grow. This allows high-efficiency recovery of the desired clones.

Because of the lethal effects of the CcdB protein, all Gateway™ vectors containing the *ccd*B gene must be propagated in an *E. coli* strain that is resistant to CcdB effects. Therefore, the DB3.1™ *E. coli* strain is recommended, which contains a gyrase mutation (*gyr*A462) that renders it resistant to the CcdB effects (Bernard & Couturier, 1992; Bernard *et al.*, 1993; Miki *et al.*, 1992).

- Designing *att*B-PCR primers

To generate PCR products suitable for use as substrates in a Gateway™ BP recombination reaction with a donor vector, *att*B sites need to be incorporated into the PCR products. The design of the PCR primers to amplify the gene of interest is critical for recombinational cloning using Gateway™.

Guidelines for designing the forward primer:

- To enable efficient Gateway™ cloning, the forward primer must contain the following structures: 1) Four guanine (G) residues at the 5′ end followed by 2) the

25 bp *att*B1 site followed by 3) at least 18-25 bp of template- or gene-specific sequence.

- The *att*B1 site ends with a thymidine (T). If the PCR products should be fused in frame with an N-terminal tag, the primer must include two additional nucleotides to maintain the proper reading frame with the *att*B1 region. These two nucleotides **cannot** be AA, AG or GA, because these additions will create a translation termination codon. Therefore, all primers were designed with the two additional nucleotides TC to maintain the proper reading frame.

Example 1 for designing a forward primer for native expression (in **bold** the *att*B1 site, underlined is the ATG initiation codon):

Shine-Dalgarno **Kozak**

5'-GGGG**ACAAGTTTGTACAAAAAAGCAGGCT**TCGAAGGAGATAGAACCATGG (18-25 gene-specific nucleotides)-3'

The inclusion of the Shine-Dalgarno and Kozak sequence allows protein expression in both *E. coli* and mammalian cells.

Example 2 for designing a forward primer for N-terminal fusions (in **bold** the *attB*1 site):

5'-GGGG**ACAAGTTTGTACAAAAAAGCAGGCT**TC (18-25 gene-specific nucleotides)-3'

The gene-specific nucleotides should be in frame with the *attB*1 sequence and no stop codons should be introduced.

Guidelines for designing the reverse primer:

- To enable efficient Gateway™ cloning, the reverse primer must contain the following structure: + four guanine (G) residues at the 5′ end + the 25 bp *att*B2 site + 18-25 bp of template- or gene-specific sequence.
- If the PCR product should be fused in frame with a C-terminal tag, the primer must include one additional nucleotide to maintain the proper reading frame with the *att*B2 region and any in-frame stop codons between the *att*B2 site and the gene of interest must be removed
- If the PCR product should not be fused in frame with a C-terminal tag the primer must include a stop codon.

Example 1 for designing a reverse primer with no C-terminal fusion tag (underlined is the stop codon, in **bold** the *att*B2 site):

5'GGGG**ACCACTTTGTACAAGAAGCTGGGT**CCTA (18-25 gene-specific nucleotides)-3'

The gene-specific nucleotides should be in frame with the stop codon.

Example 2 for reverse primer design, for C-terminal fusions (in **bold** the *att*B2 site):

5'GGGG**ACCACTTTGTACAAGAAAGCTGGGT**C (18-25 gene-specific nucleotides)-3'

In this second example the reverse *att*B PCR primer is designed to allow the expression of a C-terminal fusion protein of interest. The gene-specific nucleotides should be in frame with the *att*B2 sequence and stop codons should be removed.

- Producing an *attB*-PCR product

As DNA templates genomic DNA, mRNA, cDNA libraries or plasmids containing cloned DNA sequences can be used for amplification with *att*B-containing PCR primers. To ensure the maintenance of the sequence it is recommended to use proofreading polymerases for amplification (refer to *5.4 Polymerase chain reaction (PCR) and its variations*). Standard PCR conditions were used to prepare *att*B-PCR products following the manufacturer's instructions for the polymerase and suitable cycling parameters were chosen depending on the primers and the template. The PCR reaction was checked for correct size by agarose gel electrophoresis (refer to *5.9 DNA gel electrophoresis*).

After the generation of the correct *att*B-PCR product it is recommended to purify the PCR product to remove *att*B-primers and any *att*B primer-dimers. Primers and primer-dimers can recombine efficiently with the donor vector in the BP reaction and may increase background after transformation into *E. coli*. For purification of the PCR product different approaches were used: PCR clean up, PEG purification and DNA extraction from agarose gels (refer to *5.8, 5.7, 5.10*, respectively).

If the PCR template is a plasmid that contains the kanamycin resistance gene, it is recommended to treat the PCR reaction mixture with *Dpn* I before purifying the *att*B-PCR product. This treatment degrades the plasmid (*i.e. Dpn* I recognizes methylated GATC sites) and helps to reduce background in the BP recombination reaction associated with template contamination.

- Performing the BP recombination reaction

The BP recombination reaction facilitates the transfer of a gene of interest in an *att*B expression clone or *att*B-PCR product to an *att*P-containing donor vector to create an entry clone. Once the entry clone is created, the gene of interest may then be easily shuttled into a destination vector using the LR recombination reaction.

To perform a BP recombination reaction following substrates are required:

- Linear *att*B-flanked PCR products **or** *att*B-containing expression clones
- Supercoiled *att*P-containing donor (pDONR™) vector

The vectors are supplied as supercoiled plasmid, lyophilized in TE buffer, pH 8.0, and should be resuspended in 40 µl of sterile water to a final concentration of 150 ng/µl.

pEXP7-tet is provided as a positive control for the BP reaction. pEXP7-tet is an approximately 1.4 kb linear fragment and contains *att*B sites flanking the tetracycline resistance gene and its promoter (Tcr). Using the pEXP7-tet fragment in a BP reaction with a donor vector results in entry clones that express the tetracycline resistance gene.

For optimal efficiency, it is recommended to use the following amounts of *att*B-PCR product (or linearized *att*B expression clone) and donor vector in a 10 µl BP recombination reaction:

- An equimolar amount of *att*B-PCR product (or linearized *att*B expression clone) and the donor vector.
- 50 (fmol) each of *att*B-PCR product (or linearized *att*B expression clone) and donor vector is preferred, but the amount of *att*B-PCR product used may range from 20-50 fmol (50 fmol of donor vector is approximately 150 ng).
- For large PCR products (>4 kb) use at least 50 fmol of *att*B-PCR product, but no more than 250 ng.

The following formula is used to convert fmol of DNA to ng of DNA:

$$ng = (fmol)(N)\left(\frac{600fg}{fmol}\right)\left(\frac{1\ ng}{10^6\ fg}\right)$$

where N is the size of the DNA in bp.

Example of fmol to ng conversion for a 2,500 bp DNA *att*B-PCR product:

$$(50\ fmol)(2{,}500\ bp)\left(\frac{600fg}{fmol}\right)\left(\frac{1\ ng}{10^6\ fg}\right) = 82{,}5\ ng\ of\ PCR\ product\ required.$$

Gateway™ BP Clonase™ II enzyme mix catalyzes *in vitro* recombination between an *att*B-PCR product (or *att*B-containing expression clone) and an *att*P-containing donor vector to generate an *att*L-containing entry clone.

For setting up the BP recombination reaction, the following components were added to a reaction tube:

Components	Sample	Positive Control	Negative Control
*att*B-PCR product or linearized *att*B expression clone (20 - 50 fmol)	1 - 7 µl	--	1 - 7 µl
pDONR™ vector (150 ng/µl)	1 µl	1 µl	1 µl
pEXP7-tet positive control (50 ng/µl)	--	2 µl	--
TE Buffer, pH 8.0	To 8 µl	To 5 µl	To 10 µl

Tab. 6: Required components for setting up the BP recombination reaction.

After removing the BP clonase II from - 20 °C and vortexing briefly twice, 2 µl were added to the sample and to the positive control, but not to the negative control. This reaction was incubated for 1 hour at 25 °C. For most applications, one hour of incubation will yield a sufficient number of entry clones. Depending on requirement, the length of the recombination reaction can be extended for up to 18 hours. For large PCR products (≥ 5 kb), longer incubations (*i.e.* overnight incubation) will increase the yield of colonies and are recommended. 1 µl of the proteinase K solution was added to each reaction and incubated at 37 °C for 10 minutes. The BP reaction may now be stored for one week at - 20 °C or used for transformation.

In this work, the PCR reaction itself, as well as the PEG purification, the PCR clean up and the gel extraction were used as *att*B-containing PCR product for the BP recombination reaction. The vectors used were pDONR™201 and pDONR™221, both containing the kanamycin resistance.

- Transforming chemically competent cells

Once the BP recombination reaction has been performed, transformed competent *E. coli* are required and selected for entry clones. Any *rec*A, *end*A *E. coli* strain (like OmniMAX™ 2-T1R, TOP10, DH5α™, DH10B™ or equivalent) can be used for transformation. pUC19 is supplied with the PCR Cloning System and was used as a positive control for transformation.

The transformation procedure was applied to introduce the BP recombination reaction into One Shot® DH5α chemically competent *E. coli* following the protocol mentioned in *5.14 Cloning:* Transformation.

- The LR recombination reaction

After the generation of the entry clone, the LR reaction is performed to transfer the gene of interest into an *att*R-containing destination vector to create an *att*B-containing expression clone. To perform an LR recombination reaction, the following substrates are needed:

- Supercoiled *att*L-containing entry clone.
- Supercoiled *att*R-containing destination vector.

The LR recombination reaction is mediated by the LR Clonase™ II Enzyme Mix, a mixture of the bacteriophage λ Integrase (Int) and Excisionase (Xis) proteins, and the *E. coli* Integration Host Factor (IHF) protein.

The pENTR™-gus plasmid is provided with the LR Clonase™ II Enzyme Mix for use as a positive control for recombination and expression. Using the pENTR™-gus entry clone in an LR recombination reaction with a destination vector allows generation of an expression clone containing the gene encoding for β-glucuronidase (*gus*) (Kertbundit *et al.*, 1991).

For setting up the LR recombination reaction, the following components were added to a reaction tube:

Components	Sample	Positive Control	Negative Control
Entry clone (50 - 150 ng/reaction)	1 - 7 µl	--	1 - 7 µl
Destination vector (150 ng/µl)	1 µl	1 µl	1 µl
pENTR-gus (50 ng/µl)	--	2 µl	--
TE Buffer, pH 8.0	To 8 µl	To 5 µl	To 10 µl

Tab. 7: Required components for setting up the LR recombination reaction.

After removing the LR clonase II from - 80 °C and vortexing briefly twice, 2 µl were added to the sample and to the positive control, but not to the negative control. This reaction was incubated for 1 hour at 25 °C. Depending on the needs, the length of the recombination reaction can be extended for up to 18 hours. For large PCR products (≥ 5 kb), longer incubations (*i.e.* overnight incubation) will increase the yield of colonies. 1 µl of the proteinase K solution was then added to each reaction and incubated at

37 °C for 10 minutes. The LR reaction may now be stored for one week at - 80 °C or used for transformation into a suitable *E. coli* host. Any *rec*A, *end*A *E. coli* strain (like OmniMAX™ 2-T1R, TOP10, DH5α™, DH10B™or equivalent) can be used for transformation. This was done as for the BP reaction (except the selection marker: ampicillin for destination vectors), as well as the analysis of the transformants by PCR or digestion by restriction enzymes. Positive clones were amplified in overnight cultures, the DNA isolated by plasmid isolation and the clones sequenced entirely to ensure that no stop codon has been introduced by incorrect amplification during PCR.

7.2 Recombinant protein expression in E. coli with Gateway™ technology

The *E. coli* Expression System with Gateway® Technology contains a series of Gateway®-adapted destination vectors designed to facilitate high-level, inducible expression of recombinant proteins in *E. coli* using the pET system. Depending on the vector chosen, the pDEST™ vectors allow production of native, N- or C-terminal-tagged recombinant proteins (see Table below).

Vector	Fusion Peptide	Fusion Tag	Size increase
pDEST™14	---	---	---
pDEST™15	N-terminal	Glutathione *S*-transferase (GST) (Smith *et al.*, 1986)	27.7
pDEST™17	N-terminal	6xHis	2.6
pDEST™24	C-terminal	Glutathione *S*-transferase (GST) (Smith *et al.*, 1986)	27.9

Tab. 8: Overview of the Gateway™ destination vectors.

pDEST™14, pDEST™15, pDEST™17 and pDEST™24 contain the following elements:

- T7 promoter for high-level, T7 RNA polymerase regulated expression of the gene of interest in *E. coli* (Studier and Moffatt, 1986; Studier *et al.*, 1990).
- N- or C-terminal fusion tags for detection and purification of recombinant fusion proteins (choice of tag depends on the particular vector; see above).
- Two recombination sites, *att*R1 and *att*R2, downstream of the T7 promoter for recombinational cloning of the gene of interest from an entry clone.
- Chloramphenicol resistance gene (CmR) located between the two *att*R sites for counter-selection.
- The *ccd*B gene located between the *att*R sites for negative selection.
- Ampicillin resistance gene for selection in *E. coli.*

- pBR322 origin for low-copy replication and maintenance of the plasmid in *E. coli.*

The pET expression system uses elements from bacteriophage T7 to control expression of heterologous genes in *E. coli*. In the pDEST™14, pDEST™15, pDEST™17, and pDEST™24 vectors, expression of the gene of interest is controlled by a strong bacteriophage T7 promoter. In bacteriophage T7, the T7 promoter drives expression of gene 10 (φ10). T7 RNA polymerase specifically recognizes this promoter. To express the gene of interest, it is necessary to deliver T7 RNA polymerase to the cells by inducing expression of the polymerase or infecting the cell with phage expressing the polymerase. In the *E. coli* Expression System with Gateway™ Technology, T7 RNA polymerase is supplied by the BL21 host *E. coli* strain in a regulated manner.

Two different *E. coli* strains were used in this work for expression: (1) BL21-AI™ and (2) BL21 Star™(DE3)pLysS One Shot® chemically competent cells.

The BL21-AI™ *E. coli* strain is included in the kit and is intended for use as a host for expression of T7 RNA polymerase-regulated genes. This strain is derived from the BL21 strain (Grodberg and Dunn, 1988; Studier and Moffatt, 1986) and contains a chromosomal insertion of the gene encoding T7 RNA polymerase (T7 RNAP) into the *ara*B locus of the *araBAD* operon, placing regulation of the T7 RNAP gene under the control of the *araBAD* promoter. The *ara*B gene is deleted in this strain. Because the T7 RNAP gene is inserted into the *ara*B locus of the *araBAD* operon, expression of T7 RNA polymerase can be regulated by the sugars, L-arabinose and glucose. To induce expression from the *araBAD* promoter, L-arabinose should be used (Lee, 1980; Lee *et al.*, 1987). To modulate expression, the concentration of L-arabinose added can simply be varied. The repression of basal expression from the *araBAD* promoter is done by addition of glucose. In the absence of glucose, basal expression from the *araBAD* promoter is generally low (Lee, *et al*., 1980 & 1986). Adding glucose further represses expression from the *araBAD* promoter by reducing the levels of 3´, 5´-cyclic AMP (Miyada *et al.*, 1984).

The BL21 Star™(DE3)pLysS strain is derived from the BL21(DE3)pLysS strain. The DE3 designation means the strain contains the λ DE3 lysogen that carries the gene for T7 RNA polymerase under control of the *lacUV*5 promoter. IPTG (Isopropyl-ß-D-thiogalactopyranoside) is required to induce expression of the T7 RNA polymerase. This strain is an *E. coli* B/r strain and does not contain the *lon* protease. It is also deficient in the outer membrane protease, OmpT. The lack of two key proteases reduces degradation of heterologous proteins expressed in this strain. The pLysS plasmid carried by the

BL21(DE3)pLysS strain produces T7 lysozyme to reduce basal level expression of the gene of interest. pLysS confers resistance to chloramphenicol (CamR) and contains the p15A origin. This origin allows pLysS to be compatible with plasmids containing the ColE1 or pMB1 origin (i.e. pUC- or pBR322- derived plasmids). This strain is recommended for use in expressing toxic genes. The IPTG-inducible *lacUV5* promoter controls expression of the T7 polymerase gene in the BL21(DE3)pLysS strain. Because of the extremely high activity of T7 RNA polymerase, some basal level expression of the gene of interest may occur in uninduced cells. This creates problems in cases where the gene of interest is toxic to bacterial cells. In these cases, expression of the toxic gene under uninduced conditions leads to selection of cells that express the lowest levels of the toxic gene. These cells are often unable to express high levels of the gene of interest upon IPTG induction of the T7 polymerase. For expression of toxic genes, it is especially recommended to use BL21(DE3)pLysS. This strain produces T7 lysozyme, which helps to reduce basal levels of T7 RNA polymerase. Although levels are reduced, the cells may still contain a small amount of T7 RNA polymerase.

- Basic transformation procedure

Once the expression clone was generated via the LR recombination reaction and amplified using transformation, this plasmid was transformed into one of the expression *E. coli* strain using the same transformation protocol as for the DH5α cells (refer to: *7.1 Gateway™ technology*, transforming competent cells). One difference, however, is that one shot of cells was always used for one transformation reaction, instead of the half quantity described for subcloning.

- Expression guidelines for BL21-AI and BL21(DE3)pLysS

Following transformation, 3 or 4 transformants were picked and cultured at 37 °C in 5 ml LB medium containing the appropriate antibiotic to select for the expression plasmid (ampicillin). When using BL21(DE3)pLysS, 34 µg/ml chloramphenicol needs to be added to select for pLysS. These cultures were then grown overnight to saturation (OD_{600} = >2). With these overnight cultures fresh LB medium containing the appropriate antibiotic was inoculated to an OD_{600} of 0.05 - 0.1 (~ 1:50 dilution of the overnight culture). This dilution allows the cells to quickly return to logarithmic growth and to reach the appropriate cell density. If using BL21(DE3)pLysS it is not necessary to include again chloramphenicol in these cultures. Generally, the cells will not lose the pLysS plasmid during the limited number of cell doublings that occur in the growth and induc-

tion stages. The remainder of the overnight culture was taken to create a glycerol stock of the clones. Once the clone that best expresses the protein has been identified, the glycerol stock can be used to perform additional expression experiments. The cultures were grown until they reach mid-log (OD_{600} = ~ 0.4, 2 to 3 hours), induced by adding either L-arabinose (BL21-AI) or IPTG (BL21(DE3)pLysS) to a final concentration of 0.2 % and 0.5 mM, respectively, and cultured for additional 4 hours. Time points can be taken to analyze for optimal expression of the protein. The clones can now be analyzed by SDS-PAGE, Western blot or enzymatic assay to determine which clone expresses the protein of interest the best. The glycerol stock created from this clone can now be used for expression experiments. If the expression levels in subsequent inductions were found to decrease or the plasmid is lost, the protein may be toxic to *E. coli.*

To shorten this protocol, a variation was found to work well. After the overnight culture of the expression clones, the cultures were diluted only to an OD_{600} = ~ 0.4 and directly induced. The expression was done at 37 °C overnight, ensuring sufficient recombinant protein. Additional expression variations for difficult proteins were applied with respect to temperature, expression time, bacterial strain and medium. When a parameter was changed, it is mentioned at the appropriate position in the results (chapter C).

The expressing cells were pelleted by centrifugation at maximum speed and frozen at - 20 °C or - 80 °C.

7.3 Determination of target protein solubility

To set up the best purification strategy, it is important to determine whether the protein is soluble in the cytoplasm or located in cytoplasmic inclusion bodies. Many proteins form inclusion bodies when they are expressed at high levels in bacteria, while others are tolerated well by the cell and remain in the cytoplasm in their native configuration. If the protein forms inclusion bodies it must be solubilised with strong denaturants, such as guanidine hydrochloride (GuHCl) or urea, prior to purification. Purification under denaturing conditions ensures that all tagged proteins in the cell are solubilised and can be purified. In addition, the tag is fully exposed under denaturing conditions, which leads to more efficient purification. If purification under native conditions is preferred or necessary, the tagged protein must be soluble. Purification under non-denaturing conditions can result in reduced yields if the tagged protein is only partially exposed due to native protein folding. However, even when most of the recombinant protein is present in inclusion bodies, there may be some soluble protein, which can be purified in its native form.

So, the target protein solubility has to be determined and therefore a protocol from the handbook "The Qiaexpressionist™", from Qiagen (Hilden), was used. The cells were resuspended in 5 ml of lysis buffer for native purification. The sample was then frozen in liquid nitrogen and thawed in cold water. Alternatively, lysozyme can be added to 1 mg/ml and the sample incubated on ice for 30 minutes. The lysate was then sonicated 6 x 10 seconds at 200 - 300 W, with 10 seconds pauses, and should be kept on ice all the time. A centrifugation step at 10,000 x g at 4 °C for 20 - 30 minutes enables separation of the supernatant (soluble protein) from the pellet (insoluble protein), and the pellet was resuspended in 5 ml of lysis buffer. Both fractions were analyzed by SDS-PAGE to settle on the location of the recombinant protein.

7.4 Purification of 6xHis-tagged fusion proteins using Ni-NTA spin columns

The presence of the N-terminal 6xHis-tag in pDEST™17 allows affinity purification of recombinant fusion protein using a nickel-chelating resin such as Ni-NTA. The procedure in the handbook "The Qiaexpressionist™", from Qiagen (Hilden), was used.

Immobilized-metal affinity chromatography (IMAC) was first used to purify proteins by Porath *et al*. (1975), using the chelating ligand iminodiacetic acid (IDA). IDA was charged with metal ions such as Zn^{2+}, Cu^{2+}, or Ni^{2+}, and was then used to purify a variety of different proteins and peptides (Sulkowski, 1985). IDA has only 3 metal-chelating sites, while Nitrilotriacetic acid (NTA) is a tetradentate chelator. NTA occupies four of the six ligand binding sites in the coordination sphere of the nickel ion, leaving two sites free to interact with the 6xHis- tag (Figure 19). NTA binds metal ions far more stably than other available chelating resins (Hochuli, 1989) and retains the ions under a wide variety of conditions, especially under stringent wash conditions.

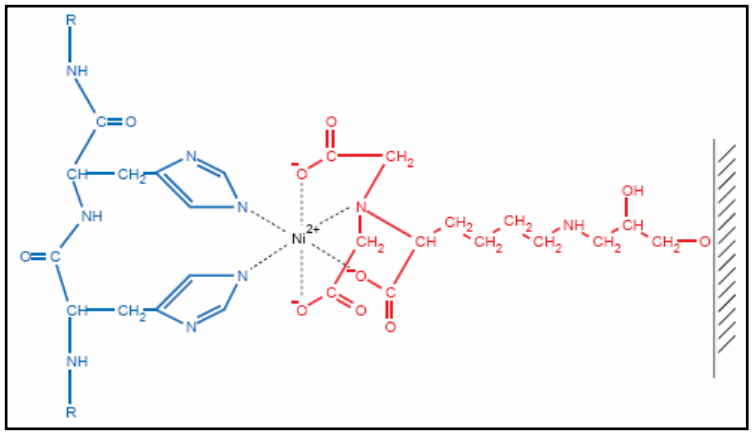


Fig. 19: Interaction between neighbouring residues in the 6xHis- tag and the Ni-NTA-matrix (From:http://www1.qiagen.com/literature/handbooks/PDF/Protein/Expression/QXP_QIAexpressionist/ 1024473 _QXPHB_0603.pdf).

The QIA*express*® System is based on the selectivity and affinity of nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography matrices for biomolecules which have been tagged with 6 consecutive histidine residues (6xHis-tag). This purification method can be performed under native and denaturing conditions. The decision whether to purify 6xHis-tagged proteins under native or denaturing conditions depends on protein location and solubility, the accessibility of the 6xHis-tag, the downstream application, and whether biological activity needs to be retained. Ni-NTA, from Qiagen, Hilden, is composed of Ni-NTA coupled to Sepharose® CL-6B.

Ni-NTA spin columns were used, where the Ni-NTA silica combines Ni-NTA with a macroporous silica support material, optimized to suppress non-specific hydrophobic interactions. Each spin column can purify up to 150 µg 6xHis-tagged protein from cellular lysates.

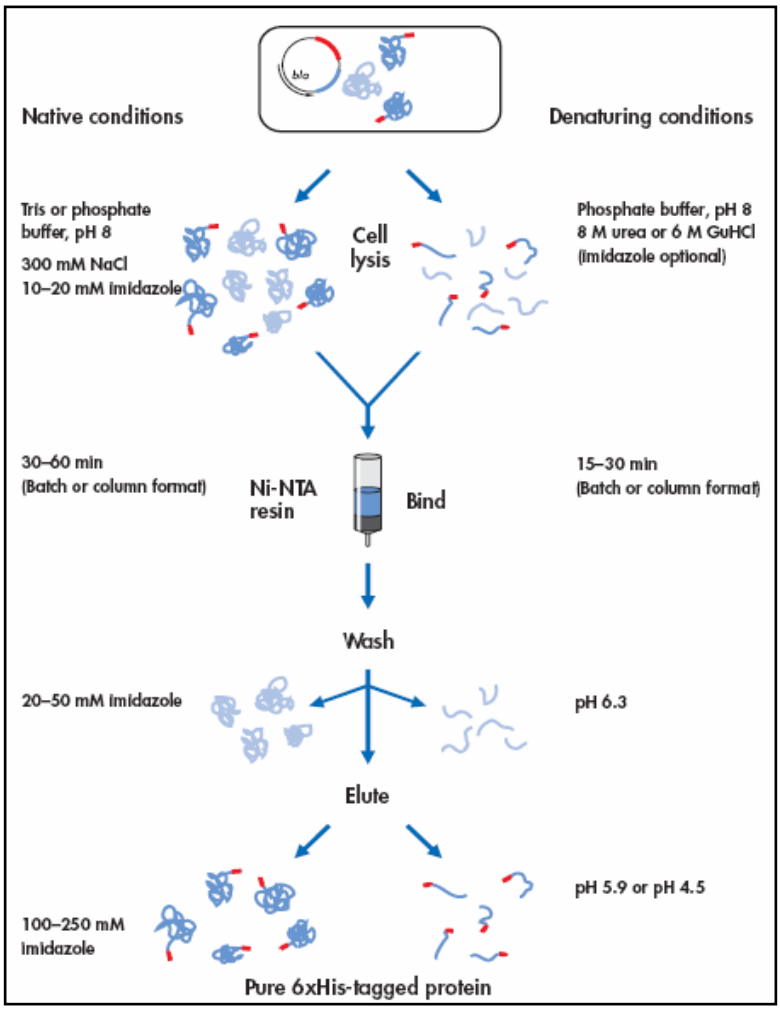


Fig. 20: Overview of the purification of 6xHis-tagged proteins under native and denaturing conditions using the Ni-NTA spin columns from Qiagen (Hilden).

Before purifying proteins under native conditions, it is important to check that the protein is soluble. However, even if most of the protein is insoluble, it is often possible to purify traces of soluble material on Ni-NTA resin. In the absence of strong denaturing agents such as urea, unstable proteins may be subject to degradation during cell harvest and lysis. It is best to work quickly and to keep the cells at 0 - 4 °C at all times. Addition of PMSF or other protease inhibitors may be helpful on a case-by-case basis, but their potential effect on the recombinant protein must be taken into consideration.

Under native conditions the frozen cell pellet was thawed on ice for 15 minutes and re-suspended in lysis buffer at 2 - 5 ml per gram wet weight. Lysozyme was added to 1 mg/ml and the solution incubated on ice for 30 minutes. A sonication step was inserted on ice using a sonicator equipped with a microtip to clear the lysate and to cleave the DNA. If the lysate is very viscous, RNase A (10 µg/ml) and DNase I (5 µg/ml) can be added and incubated on ice for 10 - 15 minutes. The lysate was centrifuged at 10,000 x *g* for 20 - 30 minutes at 4 °C to pellet the cellular debris and the supernatant was trans-

ferred to a fresh tube. A Ni-NTA spin column was equilibrated with 600 µl lysis buffer. The cleared lysate supernatant containing the 6xHis-tagged protein was loaded onto the equilibrated Ni-NTA spin column and then washed twice with 600 µl wash buffer. The elution of the protein was done with 2 x 200 µl elution buffer; the eluates were collected in fresh centrifugation tubes. Each of the centrifugation steps with the Ni-NTA columns was done for 2 minutes at 2000 rpm (approximately 700 x *g*) and from each step an aliquot was taken for SDS-PAGE analysis.

Lysis buffer:	50 mM NaH2PO4
	300 mM NaCl
	10 mM Imidazole
	Adjust pH to 8.0 using NaOH
Wash buffer:	50 mM NaH2PO4
	300 mM NaCl
	20 mM Imidazole
	Adjust pH to 8.0 using NaOH
Elution buffer:	50 mM NaH2PO4
	300 mM NaCl
	250 mM Imidazole
	Adjust pH to 8.0 using NaOH
Alternative elution buffer:	50 mM NaH2PO4
	300 mM NaCl
	500 mM Imidazole
	Adjust pH to 8.0 using NaOH

For protein purification under denaturing conditions, the cells were thawed for 15 minutes, resuspended in 1 ml buffer B and incubated with agitation for 1 hour at room temperature. The solution should become translucent when lysis is complete. If the cells do not solubilise in buffer B, then buffer A must be used. Since fractions which contain GuHCl will precipitate with SDS upon loading onto an SDS-PAGE, they must either be diluted (1:6), dialyzed before analysis or separated from GuHCl by TCA precipitation. The lysate was centrifuged at 10,000 x *g* for 20 - 30 minutes at 4 °C to pellet the cellular debris and the supernatant was collected. A Ni-NTA spin column was equilibrated with 600 µl buffer B and the cleared lysate supernatant containing the 6xHis-tagged protein ws loaded onto the pre-equilibrated Ni-NTA spin column. The

column was then washed twice with 600 µl buffer C and the elution of the protein with 2 x 200 µl buffer E. The eluates were collected in fresh centrifugation tubes. Each of the centrifugation steps with the Ni-NTA columns were done for 2 minutes at 2,000 rpm (approximately 700 x *g*) and from each step an aliquot was taken for SDS-PAGE analysis.

Buffer A:	100 mM NaH2PO4
	10 mM Tris/Cl
	6 M GuHCl
	Adjust pH to 8.0 using NaOH
Buffer B:	100 mM NaH2PO4
	10 mM Tris/Cl
	8 M Urea
	Adjust pH to 8.0 using NaOH
Buffer C:	100 mM NaH2PO4
	10 mM Tris/Cl
	8 M Urea
	Adjust pH to 6.3 using HCl
Buffer D:	100 mM NaH2PO4
	10 mM Tris/Cl
	8 M Urea
	Adjust pH to 5.9 using HCl
Buffer E:	100 mM NaH2PO4
	10 mM Tris/Cl
	8 M Urea
	Adjust pH to 4.5 using HCl

This cell lysis and protein purification under denaturing conditions was also done using Ni-NTA magnetic agarose beads, following the manufacturer's instructions in the Qiaexpressionist™ handbook, from Qiagen (Hilden).

Cell lysis was done in the same way as for the purification by spin columns under denaturing conditions. The Ni-NTA magnetic agarose beads were resuspended by vortexing for 2 seconds and then immediately 200 µl of this 5 % suspension were added to 1 ml of the lysate containing the 6xHis-tagged protein. The suspension was mixed gently on an end-over-end shaker for 30 minutes to 1 hour at room temperature. The tube was placed

on a magnetic separator for 1 minute and the supernatant was removed with a pipette. After removing the tube from the magnet, 500 µl of buffer C were added, mixed and the tube was replaced on a magnetic separator for 1 minute, to remove the buffer. Buffer E was added, 100 µl, the suspension was mixed and incubated for 1 minute. By repeated placing into the magnetic separator, the eluate can be collected and all fractions analyzed by SDS-PAGE.

7.5 Purification of GST-tagged fusion proteins using glutathione agarose

The presence of the N- or C-terminal GST tag in pDEST™15 and pDEST™24, respectively, allows purification of recombinant fusion protein using glutathione agarose. Prepacked columns from Sigma, Deisenhofen, were used. The resin in these columns consists of glutathione attached through the sulfur to epoxy-activated 4 % cross-linked beaded agarose, resulting in a 12-atom spacer. Affinity chromatography using glutathione-agarose permits mild, non-denaturing and highly selective purification of glutathione binding enzymes such as glutathione-S-transferase, glutathione peroxidase and glyoxalase I.

The resin of the columns was equilibrated with several column volumes of equilibration buffer. The column should not run dry at any time. Cell lysate was prepared in appropriate buffer. Tris or phosphate buffers, pH 6.5 to 9.5, are typical lysis buffers compatible with glutathione affinity chromatography and salt concentrations of up to 1 M do not interfere with binding. Protease inhibitors such as EDTA or PMSF are often included in the lysis buffer. The binding of GST to glutathione-agarose is unaffected by 1 % Triton X-100, 1 % Tween-20, 1 % CTAB, 10 mM DTT or 0.03 % SDS. Triton X-100 was added to a final concentration of 1 % (v/v) and the mix was centrifuged for 10 minutes at 10,000 x *g* at 4 °C to clear the cell lysate. To prevent clogging of the column, only clarified supernatant should be applied and highly viscous samples containing chromosomal DNA or RNA should be sonicated or treated with nuclease to reduce the viscosity. Cellular debris and particulate matter must be removed by centrifugation or filtration. Then the clarified supernatant was loaded onto the column under gravity flow. Depending on the sample and the flow rate, not all the protein may bind. Multiple passes over the column or closing the loaded column and incubating it on a rotator may improve the binding efficiency. The resin was then washed four times with PBS-T at 4 °C and the elution of the GST-tagged protein from the resin was done with Elution Buffer (3 times, 1 ml each). Each fraction should be analyzed by SDS-PAGE. Free glutathione can be

removed from sample by dialysis against buffer of choice. Regeneration of the column was performed following the manufacturer's instructions.

Equilibration Buffer:	Phosphate buffered saline (PBS): 10 mM phosphate buffer pH 7.4, 150 mM NaCl Add appropriate protease inhibitors when preparing cell lysate.
PBS-T:	PBS containing 1 % Triton X-100
Elution Buffer:	5 mM to 10 mM reduced glutathione 50 mM Tris-HCl, pH 8.0, freshly prepared, final pH 7.5.

7.6 Isolation of proteins from inclusion bodies

The accumulation of high amounts of recombinant protein in a host cell can lead to the formation of intracellular inclusion bodies.

To be able to purify the recombinant protein from these inclusion bodies, different approaches were used:

- Different lysozyme concentrations at different incubation temperatures
- Native lysis buffer under repeated addition of Tween and EGTA
- Different sonication durations and repetitions
- Denaturing conditions like 6 M GuHCl and 8 M Urea with varying incubation times
- Different concentrations of Triton X-100, SDS, DTT, alone and in combination
- Alkaline lysis using the E.Z.N.A plasmid Miniprep Kit I from peqlab, Erlangen

In addition to these procedures, two different protocols were also used:

- **Protocol 1:** Purifying proteins from inclusion bodies (from: http://structbio.vanderbilt.edu/chazin/wisdom/labpro/inclusion.html)
- **Protocol 2:** Purification of active eukaryotic proteins from the inclusion bodies in *E. coli* (from: http://www.dwalab.com/labman/op5.html),

Both protocols are based on the denaturing forces of 8 M urea and were applied following the instructions mentioned at the respective web sites.

Protocol 1:

Buffer A: 50 mM Tris-HCl, pH 8.0
1 mM EDTA
10 mM NaCl

Buffer B: 20 mM Na_2HPO_4, pH 7.2
20 mMNaCl
1 mM EDTA
25 % sucrose

The centrifugation bottle had to be weighed before and after the addition of the culture media to measure the weight of cells. After the cells have been spun down at 6,000 rpm for 20 minutes, 3 ml of buffer A were added for each gram of cells and the cell pellet was fully resuspended. 10 µl of 100 mM PMSF as well as 16 µl of a 50 mg/ml solution of lysozyme were added per ml of solution to the resuspended cell pellet and mixed. This solution was placed in a 37 °C water bath until the solution became viscous. To cleave the DNA and reduce the viscosity the solution was sonicated and spun down at 18,000 rpm for 30 minutes. The supernatant was retained and the pellet fully resuspended in 3 ml of buffer B for each gram of cells. Again 10 µl of 100 mM PMSF and 10 µl of Triton X-100 were added per ml of solution to the resuspended cell pellet and mixed. The mixture was spun down at highest speed possible (~ 20,000 rpm) for 20 minutes. The supernatant should be saved for further analysis and the pellet contained the inclusion bodies. To dissolve this pellet, 20 ml of 8 M urea with 1 mM DTT were added. Heating in a bath at 37 - 50 °C may facilitate the dissolving. It is crucial to dissolve the inclusion bodies and to solubilise the recombinant protein. The next centrifugation step at maximum speed should separate soluble protein from any remaining inclusion bodies, and the addition of 8 M urea with 1 mM DTT followed by centrifugation should be repeated until no further pellet dissolves. The 8 M urea solution containing the recombinant protein was then dialyzed in 2 l of a 50 mM Tris-HCl buffer solution at pH 8.5 using a 10,000 Da molecular weight cut off dialysis tube for 2 days, with repeated exchange of the buffer (refer to: *6.3 Dialysis*). All fractions were collected and aliquots of the supernatants and the pellets were taken for SDS-PAGE analysis.

Protocol 2:

Buffer A:	20 mM Tris-HCl, pH 7.5
	20 % sucrose
	1 mM EDTA
Buffer B:	PBS w/o Ca, Mg
	1 mM EDTA
	1 x PMSF, added fresh
Buffer W:	PBS w/o Ca, Mg
	25 % sucrose
	1 mM EDTA
	1 % Triton X-100
Buffer R:	50 mM Tris-HCl, pH 8.0
	1 mM DTT
	20 % Glycerol
Buffer U1:	50 mM Tris-HCl, pH 8.0
	4 M urea
Buffer U2:	50 mM Tris-HCl, pH 8.0
	8 M urea

The essence of this method is to purify the inclusion bodies followed by selective solubilisation of the expressed protein. This protocol was designed for a 1 litre culture. However, one can scale the volumes up or down, depending on the type of experimentation performed.

The cells from 1 litre were pelleted by centrifugation at 5,000 rpm at 4 °C for 20 minutes. To prevent contamination of the inclusion bodies by proteins from the outer membrane, the outer membrane was removed by resuspending the cell pellet in 50 ml buffer A. This was incubated on ice for 10 minutes and then centrifuged for additional 5 minutes at 6,000 rpm and 4 °C. The pellet was now resuspended in 50 ml of ice cold water and left on ice for 10 minutes. Again the cells were centrifuged at 8,200 rpm for 5 minutes at 4 °C and resuspended now in 10 ml buffer P containing freshly added protease inhibitor. This mixture was sonicated 3x (50 W) each with 30 seconds pulse and 30 seconds pause in between each pulse. RNase T1 (1,3 x 103 U/10 ml) and DNase I (400 ug/10 ml) were added to the sonicated cell suspension and incubated at room temperature for 10 minutes. The suspension was further diluted by addition of 40 ml of washing buffer P and the crude inclusion bodies were pelleted by centrifugation at

11,000 rpm for 30 minutes at 4 °C. An aliquot of this supernatant should be kept for assay for proteins. The pellet containing the inclusion bodies was suspended in 40 ml of buffer W, incubated on ice for 10 minutes and centrifuged at 15,000 rpm for 5 minutes at 4 °C. After retaining an aliquot of the supernatant, this washing step was repeated and the pellet separated into two fractions. The pellets were now resuspended in 10 ml of buffer U1 and U2, respectively, with a brief sonication with a 5 seconds pulse (50 W), which should facilitate the solubilisation of the aggregated proteins. The resultant suspensions were incubated for an additional hour at room temperature and then centrifuged at 10,500 rpm for 30 minutes at 4 °C. The supernatant should contain the recombinant protein and was added to 100 ml of renaturation buffer R and stirred gently at 4 °C overnight to renature the proteins. The next day, the supernatant was clarified by centrifugation at 12,000 rpm for 30 minutes at 4 °C. The supernatant should contain highly expressed proteins.

C. Results

1 Course and strategy of the project

The main goal of this doctoral thesis was the recombinant expression of the hemocyanin (KLH) of the giant keyhole limpet *Megathura crenulata* in the *E. coli* prokaryotic system. This was possible because the complete genomic and cDNA sequence and therefore also the amino acid sequence have already been elaborated within our working group (unpublished). While KLH in its natively occurring form builds high molecular mass cylindrical didecamers, 8 MDa, one subunit consists of eight paralogous FUs (FUs), termed *a* to *h* (Fig. 21).

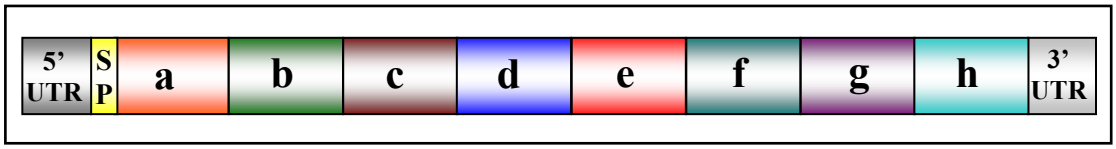


Fig. 21: Schematic depiction of the cDNA of KLH1, as well as KLH2. The eight FUs *a* to *h* are illustrated in different colours, SP represents the signal peptide and the 5' and 3' untranslated regions (UTR) are represented in grey.

The strategy was, starting from RNA, to produce KLH1 and KLH2 cDNAs of varying length (encompassing 1 and 2 FUs) and to clone these cDNAs using the Gateway™ Technology from Invitrogen, Karlsruhe. Therefore the required attachment (*att*) sites, which are a specific feature of Gateway™cloning, are added with extended primers by PCR to the cDNAs, flanking the respective 5' and 3' ends. These *att*-sites are required for the binding of the recombination enzymes and help to clone the cDNAs into one of the adapted vectors of this system. After a subcloning step, the gene of interest can be inserted into an expression vector and the coding sequences expressed recombinantly in the prokaryotic system *E.coli*. After the up-scaling of the expression cultures and purification from the bacterial cells, the recombinant proteins can be analyzed and compared with the native counterparts isolated from animals. In parallel to this work, recombinant expression of KLH was done by [redacted] in eukaryotic insect cells, within our working group, using the same clones. Both recombinant proteins were then compared, in addition to the comparison with the native form of KLH, isolated from animals.

Analysis of the prokaryotic proteins was done using biochemical methods, like native and SDS-polyacrylamide gel electrophoresis, immunological detection using different antibodies and also immunoelectrophoresis.

In parallel, the existing clones were taken as templates for additional PCRs: two consecutive cDNAs within the coding sequence of KLH are fused using the method of SOE-PCR (**S**plicing by **O**verlap **E**xtension). The overlapping sequences to the respective previous and following FU were already present in every cDNA. This is possible due to the selection of the gene specific primers for the RT-PCRs at the beginning of the work. In this way, longer cDNAs are created, adding one or two FUs at each step to the previous cDNA, until the whole subunit is built up. Between each fusion a cloning step is inserted for stabilization and amplification of the newly produced cDNA.

2 Generation of KLH cDNAs by RT-PCR and PCR

The one-step method for RT-PCR combines the generation of cDNAs from RNA templates and the subsequent amplification of these fragments in a single tube under conditions optimized for both RT and PCR.

Due to the fact that the complete hemocyanin sequences of both KLH1 and KLH2 were previously known, gene specific primers were already present and kindly made available by Klaus Streit. These primers were used for the generation of cDNA fragments of varying length, encompassing different FUs of KLH1 and KLH2. The applied mRNA and RNA were made available by the courtesy of Klaus Streit as well as of Alisa Hanisch, respectively. Each RT-PCR was set up twice to be able to apply two different annealing temperatures (50 °C and 60 °C, respectively).

2.1 KLH1-bc, KLH1-de, KLH1-fg and KLH1-h

The resulting cDNA fragments were separated electrophoretically in an agarose gel. Figure 22 shows a gel, where four different amplified cDNAs were generated using one-step RT-PCR. The cDNA is shown as distinct fluorescent bands in UV light. The comparison with the applied marker allows conclusions to be made about the mass of the cDNA bands.

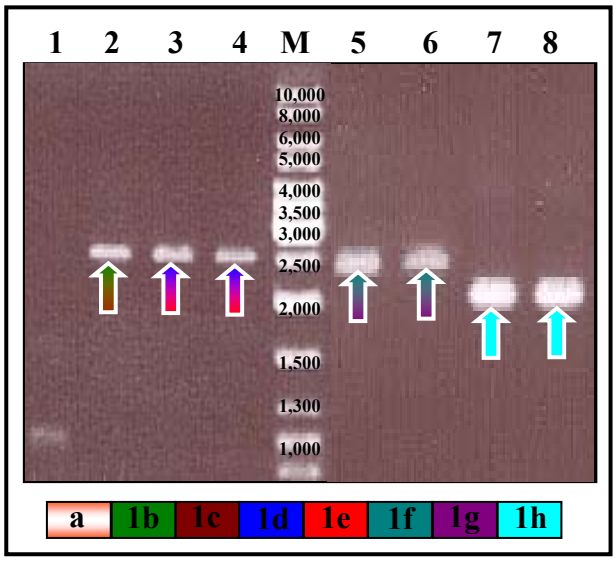


Fig. 22: Electrophoretic separation of different RT-PCR reactions. Apart from the first lane, all lanes show a distinct band at the expected mass, representing the cDNAs coding for the FUs KLH1-*bc* (green-brown arrow, lane 2), KLH1-*de* (blue-red arrows, lanes 3 and 4), KLH1-*fg* (seagreen-purple arrows, lanes 5 and 6) and KLH1-*h* (turquoise arrows, lanes 7 and 8), respectively. M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

In Figure 22, the lanes 1 and 2 represent the RT-PCRs resulting in the cDNA coding for KLH1-*bc*. Only in lane 2, where the reaction has taken place at an annealing temperature of 60 °C, does a distinct band appear at the expected mass of 2,574 bp (green-brown arrow). Lane 1 showed only a smaller amplification product, probably generated by unspecific primer hybridization. In each of the lanes 3 and 4 a distinct band of 2,576 bp is shown (blue-red arrows) which represents the cDNA of KLH1-*de*. Two other RT-PCR reactions were set up: the first results in a cDNA band of nearly 2,590 bp, representing the cDNA of KLH1-*fg* (Fig. 22, lanes 5 and 6, seagreen-purple arrows). The second RT-PCR reaction generates the cDNA coding for KLH1-*h*, including most of the 3′ UTR (2,167 bp, lanes 7 and 8, turquoise arrows).

2.2 KLH1-a, KLH1-cd, KLH1-ef and KLH1-gh

Due to newly designed gene specific primers other FU combinations were also generated by RT-PCR (Fig. 23).

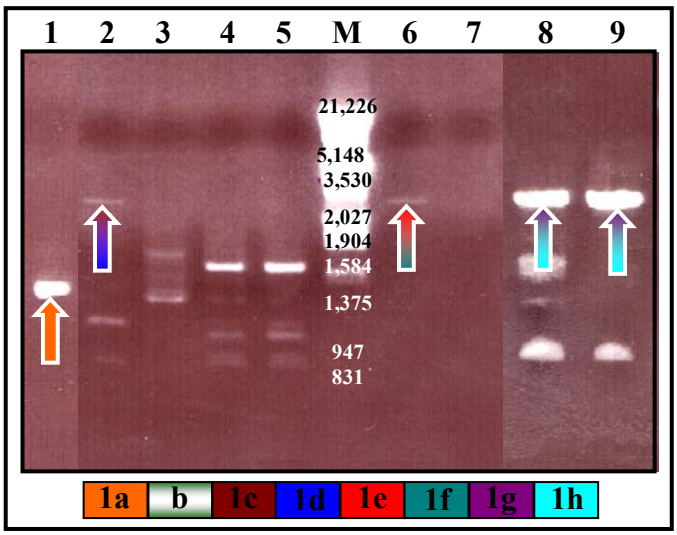


Fig. 23: RT-PCR reactions generating the cDNAs encoding KLH1-*a* (1,350 bp, lane 1, orange arrow), KLH1-*cd* with a mass of 2,582 bp (lane 2, brown-blue arrow) and KLH1-*ef* of 2,623 bp (lane 6, red-seagreen arrow). In the lanes 8 and 9 the RT-PCRs generating a cDNA encoding KLH1-*gh* are applied (purple-turquoise arrows). The lanes 3 - 5 and 7 lead to small cDNA fragments or to no bands at all. M represents the marker Lambda DNA/ EcoRI+HindIII, in bp.

Figure 23 shows that the cDNAs coding for KLH1-*a,* marked with an orange arrow in lane 1 (1,350 bp), and the cDNAs encoding KLH1-*cd* and KLH1-*ef* are only apparent at an annealing temperature of the respective primers of 50 °C (lanes 2 and 6, brown-blue and red-seagreen arrow, respectively). Other temperatures (lanes 3 and 7) and also different primer pair combinations (lanes 4 and 5) lead to small cDNA fragments or to no bands at all, probably due to unspecific primer hybridization. The RT-PCRs applied in lane 8 and 9 effectively produce a cDNA band at the expected mass of approximately 3,000 bp (Fig. 23, purple-turquoise arrows): the coding cDNA for KLH1-*gh*.

Thus, each FU of KLH1 was now available as cDNA: KLH1-*a* and KLH1-*h* as single cDNA and KLH1-*bc*, -1*cd*, -1*de*, -1*ef*, -1*fg* and -1*gh* as cDNAs encompassing two consecutive FUs.

2.3 Generation of KLH1 cDNAs encoding single FUs

Starting from the cDNAs generated by RT-PCR, each FU of KLH1 was amplified as a single cDNA. Each PCR generated a distinct band ranging from 1309 to 1366 bp, depending on the mass of the cDNA coding for the single FU and the associated overlapping sequences. The exception was the cDNA coding for KLH1-*h*, which is still larger than the other cDNAs, namely 1,641 bp, including the overlapping sequence to the previous FU-1*g* and a part of the 3' UTR.

2.4 KLH2: generation of cDNAs encoding each single FU

The same procedure was tried for KLH2: the RT-PCRs were performed using gene specific primers. This time, the forward and reverse primers were combined in the following way that each cDNA should be directly present as cDNAs coding for single the FUs of KLH2. These RT-PCR reactions were applied on an agarose gel to check the result (Fig. 24).

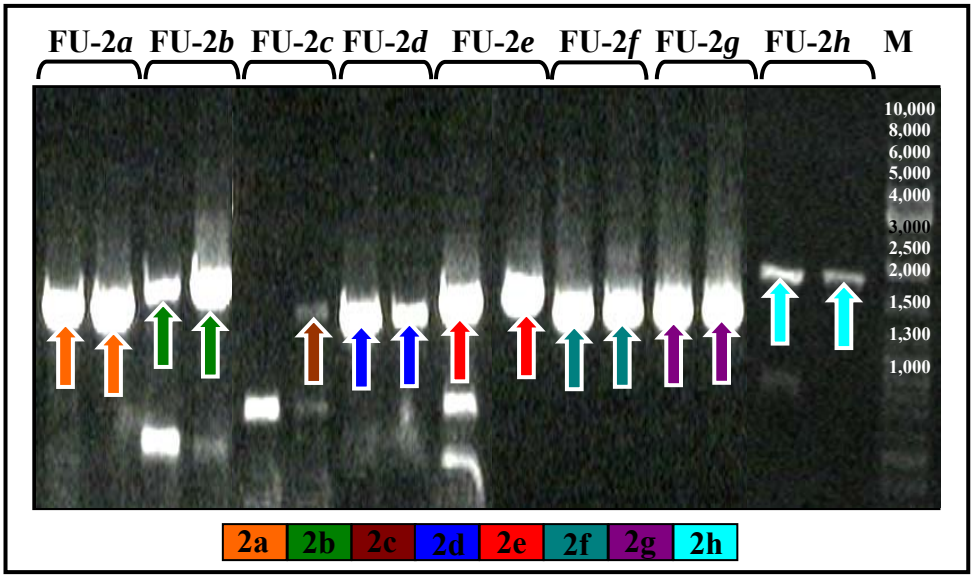


Fig. 24: Each RT-PCR generates a single cDNA of KLH2 at the expected mass, around 1,400 bp for the FUs -2*a* to -2*g* (orange, green, brown, blue, red, seagreen and purple arrows, respectively) and approximately 1,750 bp for KLH2-*h* (including a part of the 3' UTR, turquoise arrows). M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

As can be seen in Figure 24, each FU of KLH2 is represented by at least one distinct cDNA band, visible by electrophoretic separation, which corresponds to the expected mass. The cDNAs of KLH2-*a* to -2*g* are expected to be around 1,400 bp, depending on the chosen gene-specific primers: 1,394 bp for KLH2-*a* (orange arrows), 1,581 bp for KLH2-*b* (green arrows), 1,358 bp for KLH2-*c* (brown arrow), 1,300 bp for KLH2-*d* (blue arrows), 1,502 bp for KLH2-*e* (red arrows), 1,341 bp for KLH2-*f* (seagreen arrows) and 1,319 bp for KLH2-*g* (purple arrows), while the cDNA of KLH2-*h* has an expected mass of 1,755 bp (turquoise arrows). This is due to the 3′ extension of this cDNA of approximately 300 bp and to the specific selection of the reverse primer, which binds 134 bp within the 3' UTR. Apparent DNA bands, smaller than the expected ones, like observed in the RT-PCR reaction of KLH2-*b*, KLH2-*c* and KLH2-*e*, respectively, can occur by unspecific primer hybridization, especially at low annealing temperatures, like the chosen one (50 °C).

In addition, FU-*c* from KLH2 was already available at the start of the present work. This cDNA was generated by RT-PCR, cloned into the TOPO-TA vector and was kindly made available by Dr. Wolfgang Gebauer. The FU was then amplified from the vector by PCR to transfer it into one of the Gateway™ vectors.

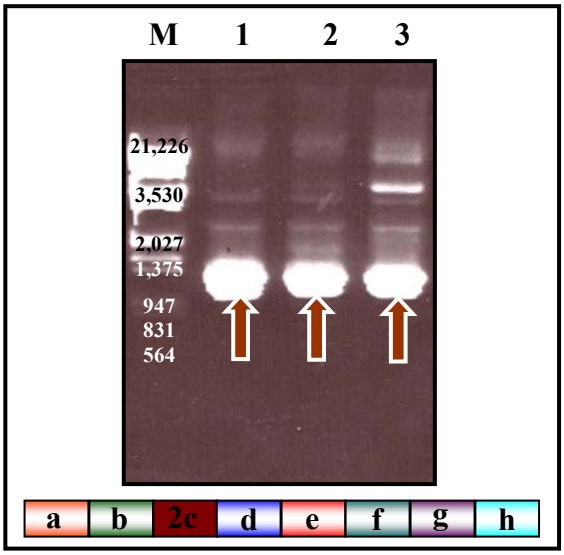


Fig. 26: The available cDNA KLH2-*c*. Very strong bands are visible at 1,249 bp in each lane (brown arrows). M represents the marker Lambda DNA/ EcoRI+HindIII, in bp.

Figure 26 shows the agarose gel of the amplification generating the cDNA for KLH2-*c* (brown arrows). These primers bind at the direct beginning and end of the FU, without generating overlapping sequences to the previous or following FU, as far the previously mentioned KLH1 and KLH2 gene-specific primers.

2.5 cDNAs encoding two consecutive FUs of KLH2: KLH2-bc and KLH2-de

Analogue to KLH1, RT-PCRs were done generating cDNAs encompassing two consecutive FUs of KLH2. This resulted in two different cDNAs: KLH2-*bc* and KLH2-*de*. The cDNA bands are clearly visible in the agarose gel represented in Figure 25.

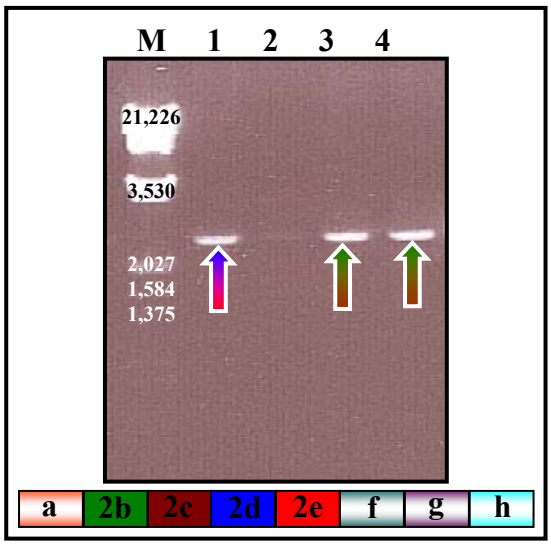


Fig. 25: Electrophoretic separation of the RT-PCRs generating the cDNAs encoding KLH2-*de* of 2,620 bp (lane 1, blue-red arrow) and KLH2-*bc* of 2,584 bp (green-brown arrows, lane 3 and 4, respectively). M represents the marker Lambda DNA/ EcoRI+HindIII.

From the second isoform of keyhole limpet hemocyanin, KLH2, two cDNAs encoding two consecutive FUs are generated by RT-PCR: KLH2-*bc* (Fig. 25, lanes 3 and 4, 50 °C and 60 °C, respectively, 2,584 bp, green-brown arrows) and KLH2-*de* (2,620 bp, lane 1, 50 °C, blue-red arrow). Further attempts to generate additional KLH2 cDNAs by RT-PCR failed.

3 Addition of Gateway™ specific *att*-sites by PCR

This PCR technique was applied when the Gateway™ Technology was used for cloning. As will be mentioned within the next chapter, the cloning occurs by recombination and therefore specific attachment sites (*att*B1 and *att*B2) are required for the binding of the enzymes. These additional sequences must flank the cDNA at the 5' and 3' end and were therefore added as extensions to the gene-specific primers used for amplification of the respective cDNAs. If the PCR primers become greater than 70 bp, the Gateway™ Technology handbook recommends the use of two sets of PCR primer pairs: a first one for the template specific amplification and a second one to install the complete *att*B sequences. Therefore, the first PCR was done using the first primer pair, and after purification of the PCR product, a second PCR installed the complete *att*B sequence.

This technique was tried for the cDNAs of KLH1-*bc*, -1*de*, -1*fg* and -1*h*, and while the first PCR generated fragments of the expected mass, the second PCR produced either no fragments at all or only smaller than expected fragments (data not shown). This result occurred independently from the cDNA employed. Therefore complete attachment primers of up to 75 bp length were generated, including the whole *att*B sites and

the template specific sequence. Because of the length, this was first tried only for the cDNAs encoding KLH1-*de* and KLH1-*fg*. The resulting agarose gel (not shown) revealed that the PCR with the complete attachment primers worked much better than the one with the two sets of primers. Thus, the required *att*B sites were successfully added by PCR to all fragments at this stage representing cDNAs of KLH1 and KLH2, using the appropriate complete attachment primer pair (data not shown). This was checked by agarose electrophoresis, and the DNA was purified using PCR clean up for subsequent cloning steps.

The following Table 9 shows an overview of all KLH1 and KLH2 cDNAs encompassing single or two consecutive FUs, generated by RT-PCR and PCR.

KLH1	1*a*	1*b*	1*c*	1*d*	1*e*	1*f*	1*g*	1*h*	1*ab*	1*bc*	1*cd*	1*de*	1*ef*	1*fg*	1*gh*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+

KLH2	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	2*g*	2*h*	2*ab*	2*bc*	2*cd*	2*de*	2*ef*	2*fg*	2*gh*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	-	-	-

Table 9: Overview of all the available cDNAs of KLH1 and KLH2 encompassing single FUs or two consecutive FUs, generated by RT-PCR and PCR ("+": available; "-": not available).

4 Cloning of the generated cDNAs using the Gateway™ technology

The Gateway™ Technology is based on the bacteriophage lambda site-specific recombination system which facilitates the integration of lambda into the *E. coli* chromosome and the switch between the lytic and lysogenic pathway (Ptashne, 1992). The recombination reaction occurs between specific attachment sites (*att*) on the interacting DNA molecules. They are added by PCR to the target cDNAs by extended primer pairs (*att*-site PCR) whereas the Gateway™ "donor" vectors already have the corresponding sites. The first reaction is called the BP-reaction because the *att*B1 and *att*B2 sites of the PCR fragment recombine with the vector specific *att*P1 and *att*P2 sites, respectively. The resulting clones are called "entry clones".

After the *att*-site PCR of all produced cDNAs of KLH1, the fragments (the PCR directly as well as the PCR clean up) were cloned into the vector pDONR™221. Clone PCR was done using the vector specific primers M13 forward and reverse. The agarose gels demonstrated that for every cDNA at least two clones were generated having the right insert mass of ~ 1,500 bp, depending on the FU and including the distance to the primer binding sites. FU-1*h* was visible at a mass of approximately 1,800 bp due to the

larger mass of the FU and including a part of the 3′ UTR. From the cDNAs encoding two FUs generated by RT-PCR, it was possible to clone KLH1-*bc*, -1*de*, -1*ef*, -1*fg* and -1*gh* into the donor vector.

From the second isoform, FU-2*c* was already available at the start of the present work and was cloned into the vector pDONR™221, generating the corresponding entry clone. The other cDNAs encoding the single FUs of KLH2, which were generated by RT-PCR, were also cloned into pDONR™221. The cDNAs encoding KLH2-*bc* and -2*de* failed to be cloned.

Once the cDNAs were present as entry clones, a second cloning step was performed (LR reaction), where the genes of interest are transferred into a Gateway™ specific destination vector, creating a so called expression clone. Four different standard destination vectors are available from Invitrogen for the expression in *E.coli*: pDEST™14, pDEST™15 (N-terminal GST-tag), pDEST™24 (C-terminal GST-tag) and pDEST™17 (N-terminal 6xHis-tag).

5 First recombinant expression of a KLH substructure in *E. coli*: establishing the method with KLH2-*c*

5.1 Expression experiment using the vector pDEST™14

Due to the fact, that the cDNA of the FU KLH2-*c* was already present at the beginning of this thesis, it was taken for the first expression experiments.

The entry clone was first subcloned into the destination vector pDEST™14 which contains neither N- nor C-terminal tag. The resulting expression clone was transformed into the expression bacterial strain BL21-AI™ and expressed following standard conditions. The SDS-PAGE with the applied samples showed a distinct band of approximately 50 kDa after staining with Coomassie Brilliant blue. However, these protein bands were seen in the uninduced cultures and not in the induced cultures (data not shown). To be sure that the samples were not mixed up, the expression was repeated but the results remained the same (data not shown).

5.2 Changing to pDEST™24

Thus, the KLH2-*c* cDNA was subcloned from the entry clone into the destination vector pDEST™24, which has a C-terminal GST tag. The expression was done as mentioned above and the result was visualized by SDS-PAGE. Now two additional protein bands appeared within the induced cultures, one at 44 kDa and one at approximately 48 kDa

(data not shown). The higher mass band can represent KLH2-*c*: the calculated theoretical molecular weight is of 48,031 Da. The origin of the smaller protein remains unknown.

5.3 First successful expression using pDEST™17

The cDNA of KLH2-*c* was again subcloned, but this time into the expression vector pDEST™17, having an N-terminal 6xHis-tag. To enable this step, specific measures had previously to be taken during forward primer design (refer to: B Materials and Methods, *7.1 Gateway™ technology*, - Designing *att*B-PCR primers). After the cloning into pDONR™221, a consecutive subcloning step transferred the cDNA into the destination vector pDEST™17. The expression was done under the same conditions as mentioned before. Six clones were picked and expressed and aliquots of the induced (I) and uninduced (U) cultures were applied on an SDS-PAGE (Fig. 27).

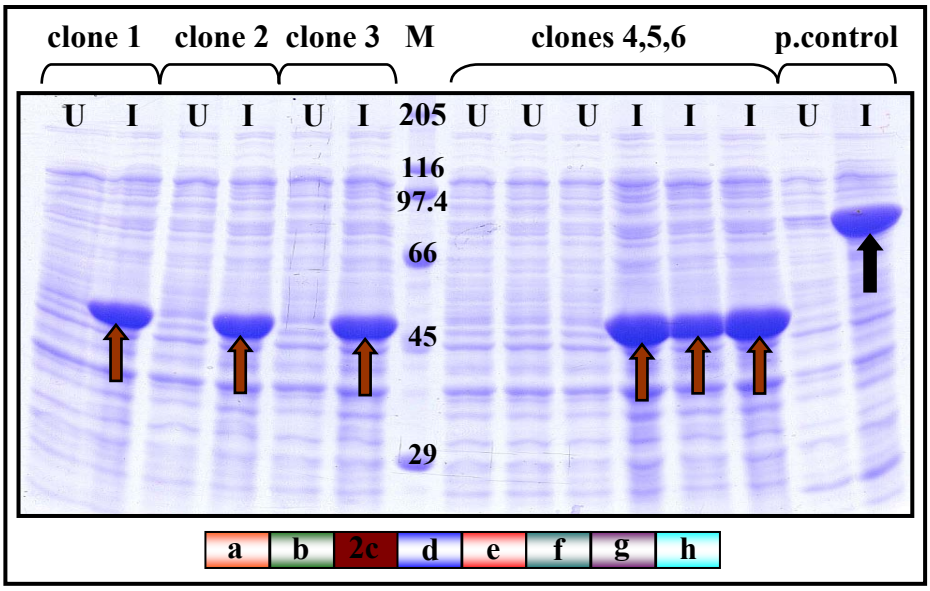


Fig. 27: SDS-PAGE of the aliquots of the uninduced (U) and induced (I) cultures of the six KLH2-*c* expressing clones in BL21-AI™ as swell as a clone expressing the positive control protein β-glucuronidase (p.control). Each induced KLH2-*c* culture shows a distinct band at approximately 50 kDa (brown arrows), whereas the positive control express a band at 80 kDa (black arrow). M represents the marker, in kDa.

The calculated theoretical molecular weight of KLH2-*c* including the 6xHis-tag is of 50,592 Da. As it can be seen in Figure 27, a strong, distinct, single band of approximately 50 kDa (brown arrows) appears in the induced cultures, which is not present in the uninduced cultures.

5.4 Immunological analyses

Two further SDS-PAGEs were performed containing the same KLH2-*c*-expressing aliquots and the respective proteins were transferred to nitrocellulose using the *semi-dry*

electroblotting method. A reversible staining of the proteins using a 1:10 dilution of Ponceau S (Fig. 28A) ensures that all protein had been transferred onto the membrane. The first blot was incubated using an α-His-antibody, which, in return, was detected using a secondary horseradish peroxidase labelled antibody (brown precipitation product, Fig. 28B).

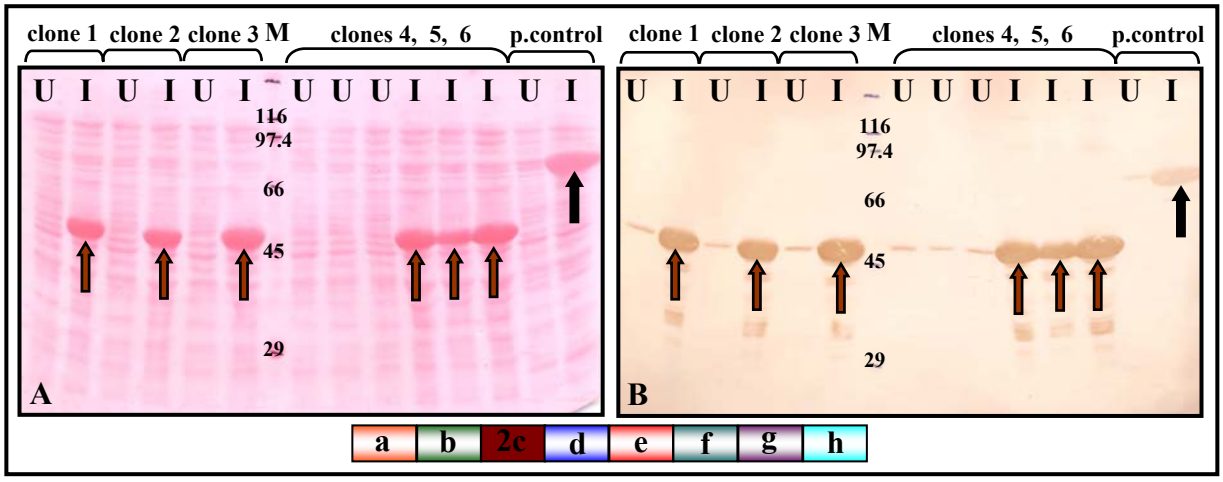


Fig. 28: Western blot of KLH2-*c*. **A**: Ponceau S staining of blotted proteins representing the uninduced (U) and induced (I) cultures of KLH2-*c* expressing clones 1 to 6 as well as the positive control protein β-glucuronidase (p.control). All proteins have been transferred onto the membrane (brown and black arrows). **B**: Specific detection using an α-His-antibody. Each lane of the induced KLH2-*c* expressing cultures (I) shows a thick band at 50 kDa (brown arrows) and the control protein at 80 kDa (black arrow). M represents the marker, in kDa.

Shown in Figure 28A, the Ponceau S staining of the blotted proteins shows that numerous proteins of the induced (I) as well as of the uninduced (U) cultures have been transferred onto the membrane. Part B of Figure 28 represents the specific detection of the proteins using an α-His-antibody. The above mentioned additional protein bands in the induced cultures of KLH2-*c* (brown arrows) as well as smaller degradation products, and also the β-glucuronidase (black arrow) are clearly detected by the antibodies recognizing the coexpressed N-terminal 6xHis-tag. All uninduced cultures exhibit weak bands of the same mass, also detected by the antibody. This can be due to basal expression levels in the bacteria cells having the plasmid, also without the addition of the inductor L-arabinose.

The second membrane containing the blotted proteins was incubated using a specific α-KLH2*c*-antibody. The detection was done due to the secondary antibody with the alkaline phosphatase as reporter enzyme (purple precipitation product). Figure 29 depicts the results.

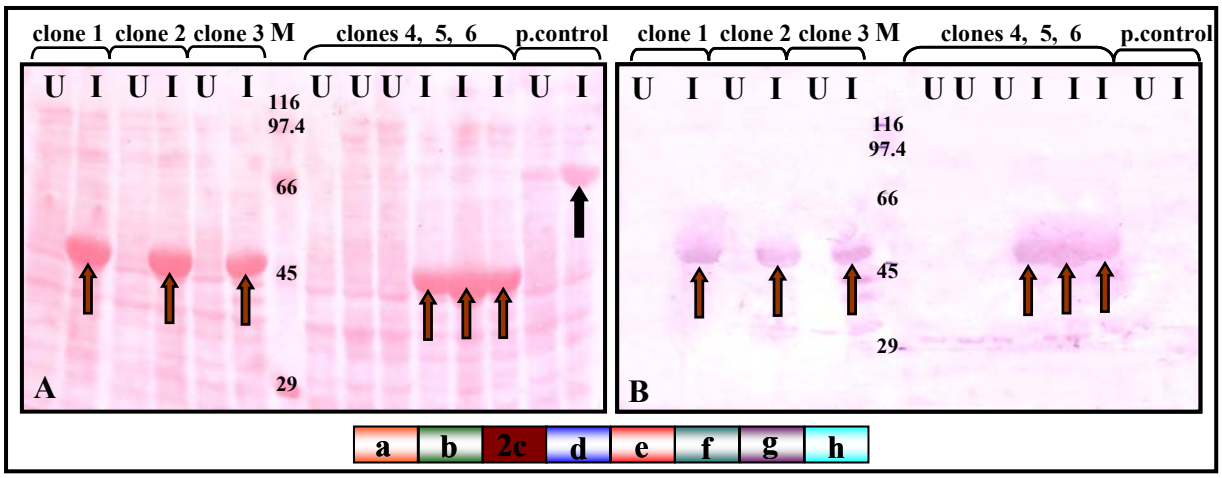


Fig. 29: Western blot of KLH2-*c*. **A**: Ponceau S staining of blotted proteins representing the uninduced (U) and induced (I) cultures of KLH2-*c* expressing clones 1 to 6 as well as the positive control protein β-glucuronidase (p.control). **B**: Specific detection using an α-KLH2-*c*-antibody. Each lane of the induced KLH2-*c* expressing cultures (I) shows a thick band at 50 kDa (brown arrows). M represents the marker, kDa, respectively.

The Western blot using the polyclonal α-KLH2-*c*-antibody displayed in Figure 29B, shows the specificity of the detection by antibodies: only the induced (I) cultures of the KLH2-*c* expressing clones show a band at approximately 50 kDa (brown arrows), which corresponds to the expected mass of the recombinant protein. The positive control protein β-glucuronidase (p.control) is, as expected, not recognized by the specific α-KLH2-*c*-antibody.

6 Cloning of all generated cDNAs of KLH1 and KLH2 into pDEST™17

Knowing now that the expression of KLH2-*c* worked well in combination with the destination vector pDEST™17, I tried to subclone all currently available cDNAs of KLH into this vector to create the respective expression clones. This was possible for each single FU of KLH1 and KLH2, as for the cDNAs KLH1-*bc*, -1*ef*, -1*gh*. Other cDNAs of KLH1 failed to be subcloned into any of the destination vectors. The sequencing showed that all expression clones were stop codon-free and maintained in the right reading frame. Table 10 shows an overview of all cloned KLH1 and KLH2 cDNAs into pDEST™17:

KLH1	1*a*	1*b*	1*c*	1*d*	1*e*	1*f*	1*g*	1*h*	1*ab*	1*bc*	1*cd*	1*de*	1*ef*	1*fg*	1*gh*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+
clone	+	+	+	+	+	+	+	+	-	+	-	-	-	+	+

KLH2	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	2*g*	2*h*	2*ab*	2*bc*	2*cd*	2*de*	2*ef*	2*fg*	2*gh*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	-	-	-
clone	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-

Table 10: Overview of the so far generated clones of KLH1 and KLH2 into the expression vector pDEST™17, compared to the already available cDNAs, generated by RT-PCR and PCR ("+": available; "-": not available).

7 First recombinant expression of a KLH1 substructure and its immunological analysis: KLH1-*h*

The next expression was performed with the FU KLH1-*h*. The transformation and expression procedure into the *E. coli* expression strain BL21-AI™ was done under the same conditions than for KLH2-*c*. Again six clones were picked and the aliquots of the uninduced and induced cultures were applied on an SDS-PAGE and stained with Coomassie Brilliant blue (Fig. 30).

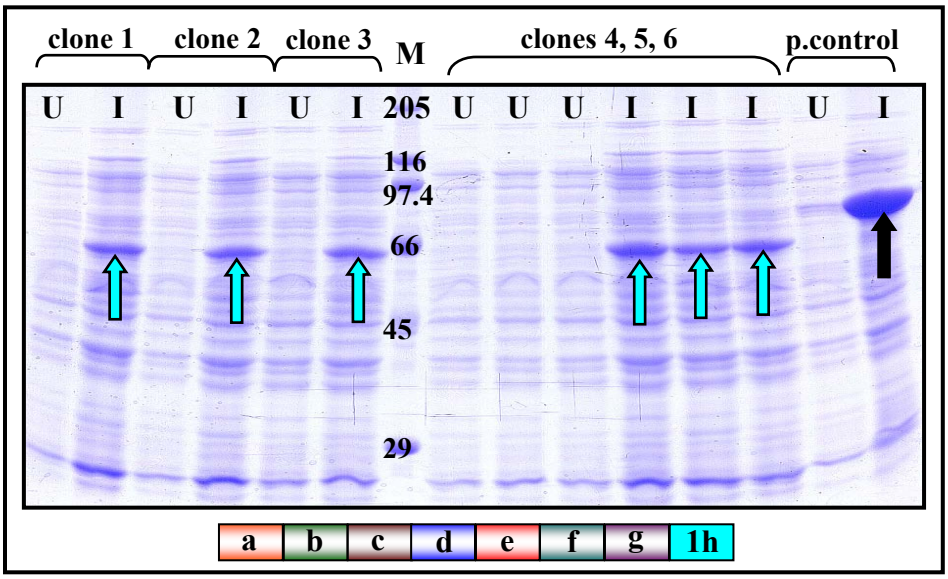


Fig. 30: SDS-PAGE of the expression of KLH1-*h*. In all induced (I) cultures of the KLH1-*h* clones a distinct band at 62.5 kDa appears (turquoise arrows), whereas the positive control protein β-glucuronidase (p.control) expresses a band of 80 kDa (black arrow). M represents the marker, in kDa.

Due to the 6xHis-tag and to the C-terminal extension of KLH1-*h* a band of approximately 62 kDa is expected. As it can be seen in Figure 30 the six aliquots of the induced

cultures expressing KLH1-*h* show a distinct band of the expected mass (turquoise arrows). The expression of the β-glucuronidase shows the expected 80 kDa (black arrow). The same samples of the KLH1-*h* expression were applied again on two additional SDS-PAGEs and the electrophoresed proteins were blotted onto two nitrocellulose membranes. The first blot was incubated with an α-His-antibody and the second one with a polyclonal KLH1-specific antibody (α-KLH1 no 4). The results are shown in Figures 31 and 32, respectively.

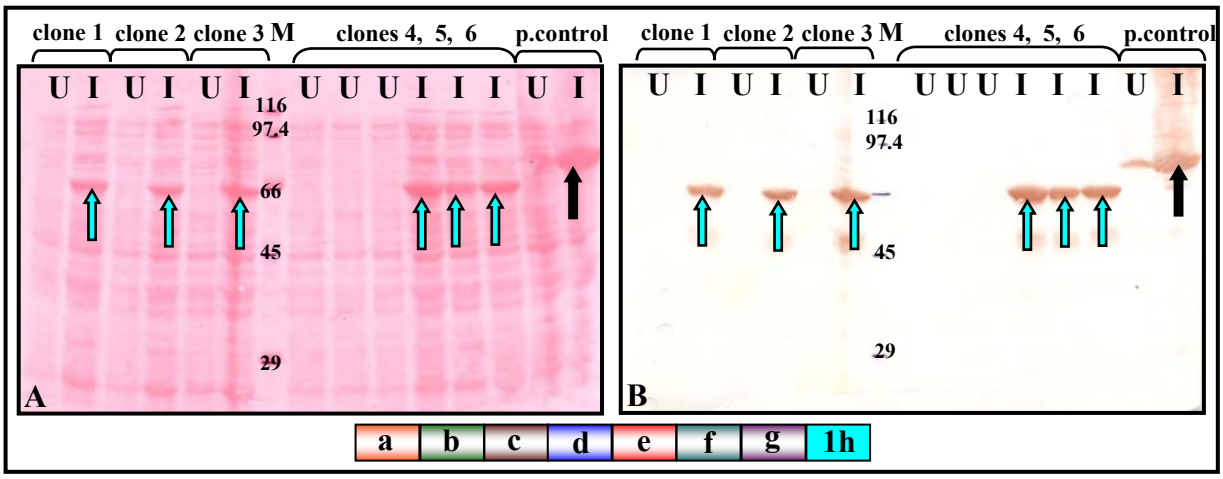


Fig. 31: Western blot of KLH1-*h*. **A:** Ponceau S staining of the expression of KLH1-*h*. **B:** Detection by an α-His-antibody. Protein bands of 62 kDa appear in the induced (I) cultures (turquoise arrows), and the 80 kDa band of the positive control (p.control, black arrow), which are not present in the uninduced (U) cultures. M represents the marker, in kDa.

Compared to the Ponceau S staining of the KLH1-*h* expression (Fig. 31A), where all the proteins reversibly have been stained, only the recombinantly expressed proteins are detected by the α-His-antibody due to the coexpressed N-terminal 6xHis-tag (Fig. 31B). KLH1-*h* is shown by a protein band of 62.5 kDa (turquoise arrows) and the control protein has a mass of 80 kDa (black arrow), as expected.

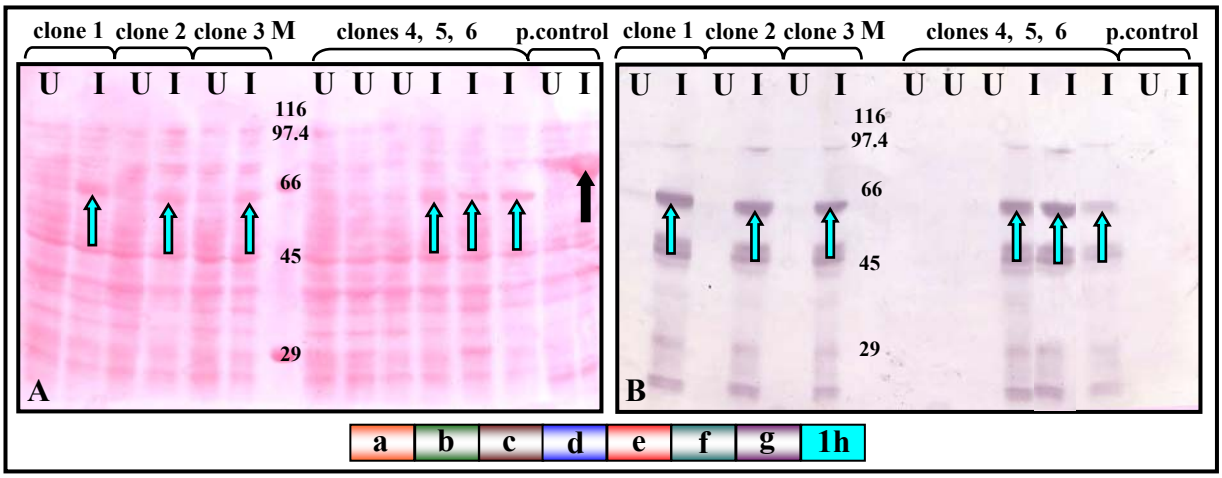


Fig. 32: Western blot of KLH1-*h*. **A:** Ponceau S staining of the expression of KLH1-*h*. **B:** Detection using a polyclonal α-KLH1-antibody. On both sides the distinct bands of 62 kDa appears in each induced (I) culture of the KLH1-*h* expression (turquoise arrows), which are not shown in the uninduced (U) cultures but clearly detected by the antibody. M represents the marker, in kDa.

The immunological detection by the polyclonal α-KLH1-antibody (Fig. 32B) confirms the results seen in Figure 30: the recombinantly expressed proteins representing KLH1-*h* are detected in a specific way by the antibodies within the induced (I) cultures (turquoise arrows), as well as smaller degradation products, but not the positive control protein (p. control).

8 Expressing KLH1-*a* to KLH1-*g*

8.1 Partial success and occurring problems

Since the expression of KLH2-*c* as well as of KLH1-*h* worked, I tried to also express the remaining single FUs of KLH1, which were already cloned into the destination vector pDEST™17.

Each expression clone of the single FUs KLH1-*a*, -*b*, -*c*, -*d*, -*e*, -*f* and -*g* was transformed independently into the expression bacteria BL21-AI™ and from each transformation one clone was picked and cultured. The expression conditions were maintained as previously used for KLH2-*c* and KLH1-*h* and aliquots of the induced (I) and uninduced (U) cultures were applied on a large SDS-PAGE. The Coomassie staining showed that most of the single FUs were represented by distinct additional protein bands around 50 kDa. KLH1-*b*, however, showed no protein band of the expected mass. Additionally, the induced culture of KLH1-*e* showed a weak band at the expected 50 kDa, but also a smaller band at approximately 26 kDa. KLH1-*g* showed only a band at approximately 27 kDa which do not match with the calculated theoretical molecular weight of 54 kDa, including the coexpressed 6xHis-tag (data not shown).

Because the sequencing of all clones showed that they are free of stop codons, the expression of the highly expressing clones (KLH1-*a*, -1*c*, -1*d* and -1*f*) as well as the expression of KLH1-*e* and -1*h* was repeated under the same conditions, except the expression time that was shortened to 4 hours, instead of overnight. Aliquots of the expression cultures were again applied on a large SDS-PAGE and stained with Coomassie Brilliant blue (Fig. 33).

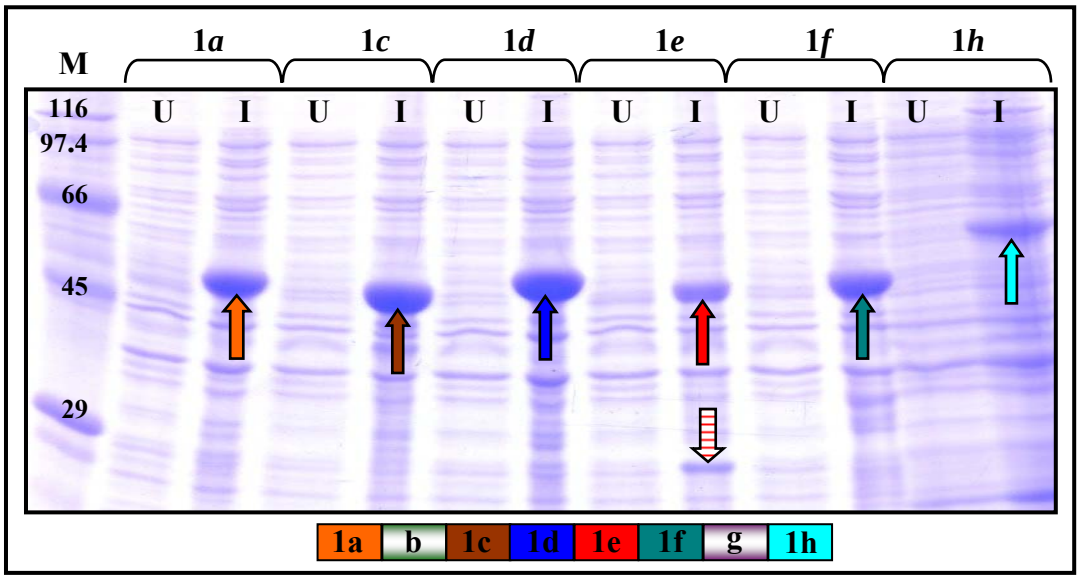


Fig. 33: SDS-PAGE of the expression experiment of the expressing single FUs KLH1-*a*, -1*c*, -1*d*, -1*e* and -1*f*. Each induced (I) culture shows a distinct band of approximately 50 kDa (orange, brown, blue, red and seagreen arrows, respectively), which is not present in the uninduced (U) culture. The expression of the expected 50 kDa band of KLH1-*e* is now stronger (red arrow). M represents the marker, in kDa.

Each induced culture showed a distinct protein band around 50 kDa. Due to the modified expression time (4 hours instead of overnight), the expression of KLH1-*e* works better (Fig. 33): the expected 50 kDa band is now stronger (red arrow) compared to the one observed in the above mentioned experiment (data not shown). The coexpressed smaller band of approximately 26 kDa is indeed still present, but weaker (striped red arrow).

8.2 Immunological analyses of the so far expressed FUs of KLH1

To confirm the origin of the expressed proteins, the electrophoresed samples were transferred onto two nitrocellulose membranes and incubated each by an α-His-antibody and by a polyclonal α-KLH1-antibody, respectively. The Figures 34 and 35 show the results of the Western blots.

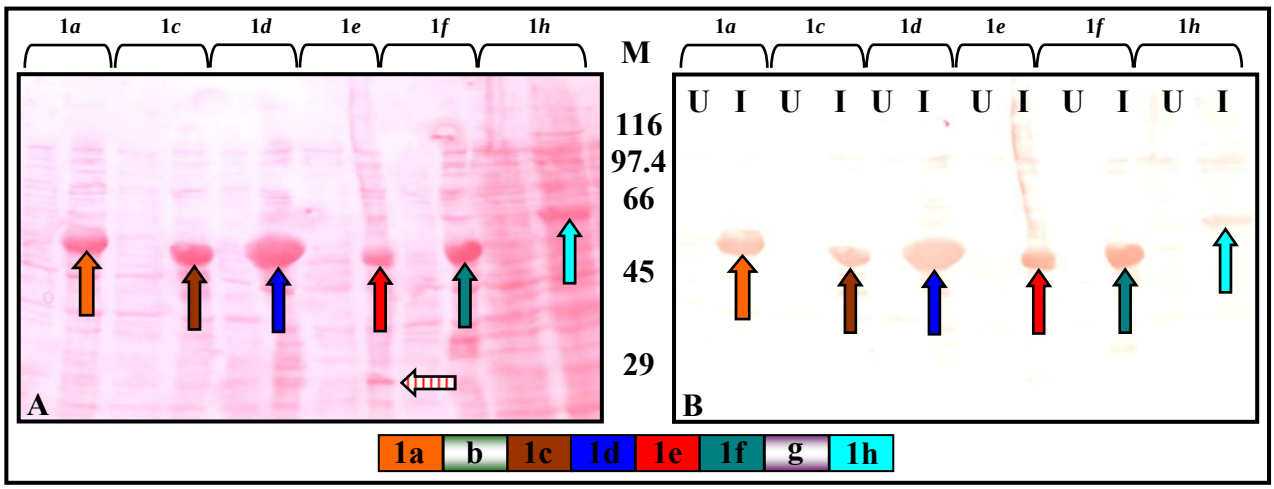


Fig. 34: Western blot. Ponceau S staining (A) and corresponding detection of the proteins of the uninduced (U) and induced (I) cultures of the expression of KLH1-*a* (orange arrow), -1*c* (brown arrow), -1*d* (blue arrow), -1*e* (red arrow), -1*f* (seagreen arrow) and KLH1-*h* (turquoise arrow) by an α-His-antibody (B). M represents the marker, in kDa.

The Western blot of the expression of the selected FUs of KLH1 (Fig. 34B) confirms the results shown by Coomassie staining (Fig. 33): the α-His-antibody recognizes in a specific way the coexpressed N-terminal 6xHis-tag of each recombinant FU. The bands correspond to the expected mass of each FU and masses range from 53 kDa to 55 kDa (theoretical calculated molecular masses: 52,980 Da for KLH1-*a*, orange arrow; 53,717 Da for KLH1-*c*, brown arrow; 53,149 Da for KLH1-*d*, blue arrow; 53,855 Da for KLH1-*e*, red arrow and 53,815 Da for KLH1-*f*, seagreen arrow) as well as 62.5 kDa for KLH1-*h* (exactly 62,528 kDa, turquoise arrow). Interestingly, the appearing mass of KLH1-*a* seems to be superior compared to the other FUs, although the theoretical calculated molecular mass of KLH1-*a* is the smallest one (52.9 kDa). The smaller coexpressed band of approximately 26 kDa in the induced culture of KLH1-*e* is not detected by the α-His-antibody (Fig. 34B).

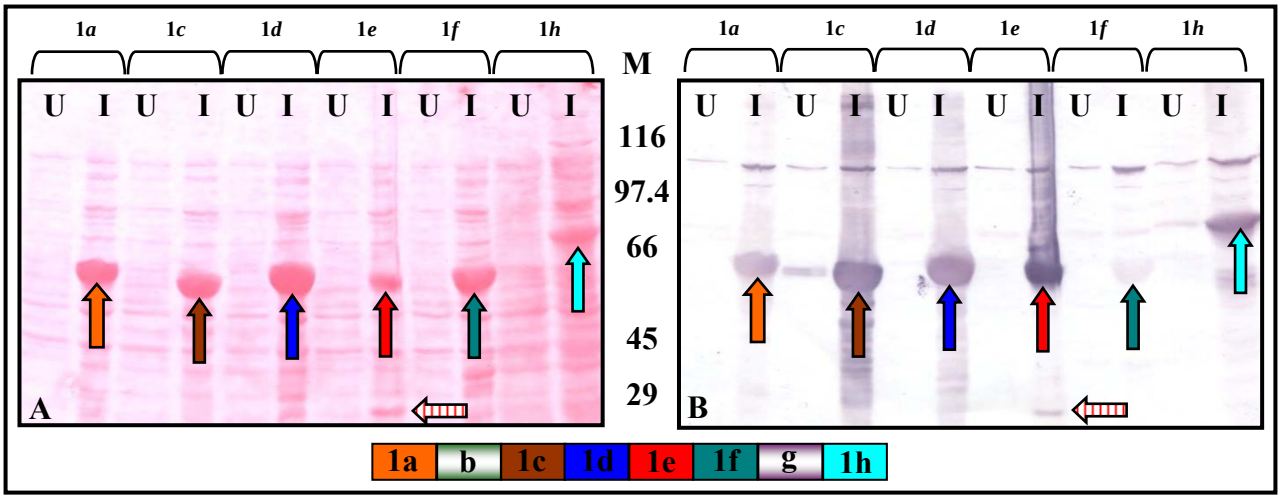


Fig. 35: A: Western blot. The strong protein bands of the recombinant KLH1 FUs as well all other prokaryotic proteins reversibly stained with Ponceau S. **B:** Specific detection of the FUs of KLH1 within the induced (I) cultures by the α-KLH1-antibody (KLH1-*a*, orange arrow; -1*c*, brown arrow, -1*d*, blue arrow; -1*e*, red arrow; -1*f*, seagreen arrow and -1*h*, turquoise arrow). M represents the marker, in kDa.

The second Western blot of the expression of KLH1-*a,* -*c*, -*d*, -*e*, -*f* and -*h* (Fig. 35B) also shows distinct protein bands, revealed by the polyclonal α-KLH1-antibody. In each induced (I) culture the strong band of recombinantly expressed protein is detected in a specific way as well as smaller degradation products. Interestingly, the additional band of 26 kDa in the expression culture of FU-1*e* is recognized by the antibody (striped red arrow), but also a lot of other smaller degradation products, also in the other induced cultures. At approximately 100 kDa, an additional band is also detected by the antibody, but this is also observed for the uninduced (U) culture. Interestingly, KLH1-*f* is more weakly stained by the α-KLH1-antibody than the other FUs (seagreen arrow). The Ponceau S staining shows, however, that a comparable amount of protein has been applied and transferred (Fig. 35A).

8.3 Changing the expression parameters for KLH1-b and KLH1-g

Because the expression of the FUs KLH1-*b* and KLH1-*g* failed completely, the expression experiment was repeated by changing the expression conditions: the culture medium (NZY instead of LB-medium), the expression duration (4 hours instead of overnight expression) and also the L-arabinose concentration was varied (ranging from 0.01 to 5 %) but the expression remained still unsuccessful (data not shown). Therefore, a new expression parameter was changed: the bacterial strain. The expression clones were transformed into the expression bacteria BL21Star™(DE3)pLysS. This bacterial strain carries both the DE3 lysogen and the plasmid pLysS. This latter constitutively expresses low levels of T7 lysozyme, which reduces basal expression of recombinant genes by

inhibiting basal levels of T7 RNA polymerase. This system is specially designed for an improved folding of recombinantly expressed proteins due to a tight regulation, so that the proteins can fold more slowly (Fig. 36).

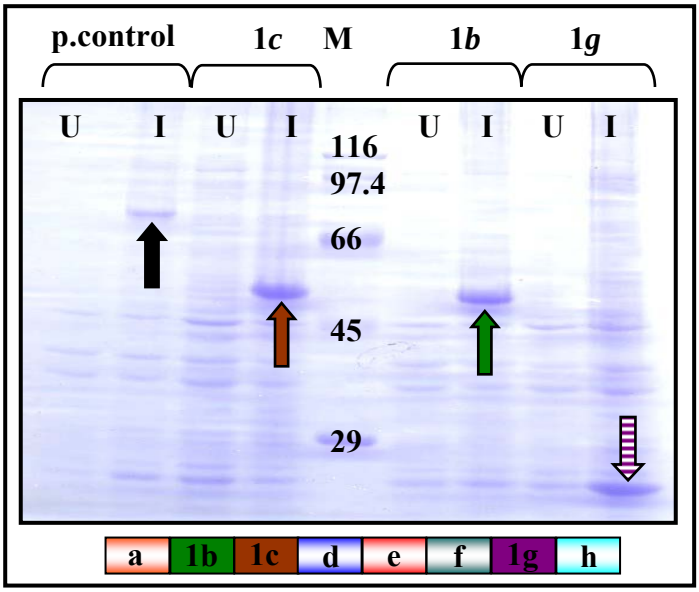


Fig. 36: SDS-PAGE of the expression of KLH1-*b* and -1*g* in a different bacterial strain: BL21Star™(DE3) pLysS. While the induced (I) cultures of the control protein β-glucuronidase and KLH-1*c* show the same result as before the expression of KLH1-*b* now reveals a distinct band of recombinant protein at the expected 52 kDa (green arrow). The expression of KLH1-*g* fails again (striped purple arrow). M represents the marker, in kDa.

The change of the bacterial strain from BL21-AI™ into BL21Star™(DE3)pLysS shows a difference regarding the expression behaviour of the clone encoding KLH1-*b* (Fig. 36). Now a distinct band of approximately 52 kDa appears in the induced (I) culture (green arrow) of this FU, which is not present in the uninduced (U) culture. As a control, the already in BL21™-AI expressing KLH1-*c* was transformed in BL21Star™(DE3)-pLysS, as well as the positive control protein β-glucuronidase, to be able to qualify the expression capacity of this new bacterial strain. As it can be seen in Figure 36, both the control protein and KLH1-*c*, show a strong band at the expected mass of 80 kDa (black arrow) and 54 kDa (brown arrow), respectively. The expression of KLH1-*g*, in contrast, still fails: again, no additional band is visible at the expected mass of 50 kDa, but only at 26 kDa (striped purple arrow).

8.4 Additional investigations concerning the expression of KLH1-g

Here, a further expression parameter was changed: the expression temperature. The expression procedure still took place in the new bacterial strain BL21Star™(DE3)pLysS, but the expression temperature after induction was lowered to 28 °C (instead of 37 °C), so that the bacteria cells grew more slowly. Aliquots of the induced (I) and uninduced

(U) cultures were taken after 2, 4, 6 and 24 hours after induction to monitor the expression course. The result is depicted in Figure 37.

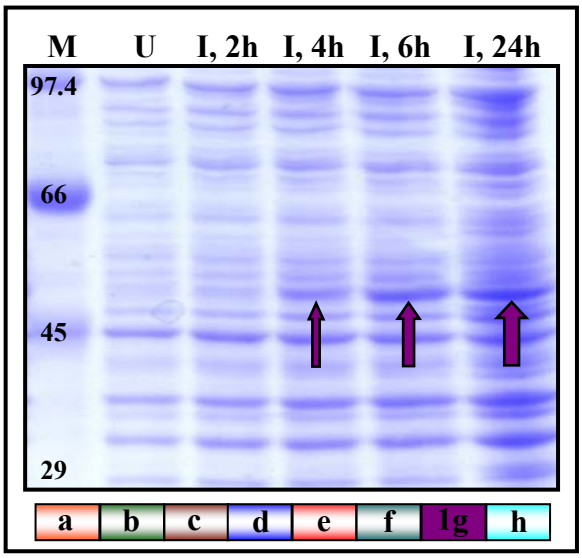


Fig. 37: SDS-PAGE of the time course of the recombinant expression of KLH-1*g* in BL21Star™(DE3)pLysS. While the quantity of all prokaryotic proteins seems to be proportionally more or less the same, a clear additional band appears at approximately 50 kDa the longer the expression experiments, starting 4 hours after induction (purple arrows).

Figure 37 shows the monitored time course of the expression of KLH1-*g*. After 4 hours induction, a weak band of recombinant protein appears in the induced culture (I, 4h, purple arrow), which is not present in the uninduced (U) culture. The mass of this protein band corresponds to the one expected: 50 kDa. The longer the expression experiment, the greater the additional band becomes (I, 6h; I 24 h, purple arrows).

8.5 Immunological detection of KLH1-b and KLH1-g

Now that the expression of both missing FUs (KLH1-*b* and KLH1-*g*) finally succeeded, the expression was repeated and the expressing clones of KLH1-*b* and -1*g*, as well as of the positive control and of KLH1-*c*, were applied on two additional SDS-PAGEs. The electrophoresed proteins were each transferred by Western blot onto nitrocellulose membranes: the first was incubated with an α-His-antibody, the second with a polyclonal α-KLH1-antibody (Fig. 38A and B, and Fig. 39A and B, respectively).

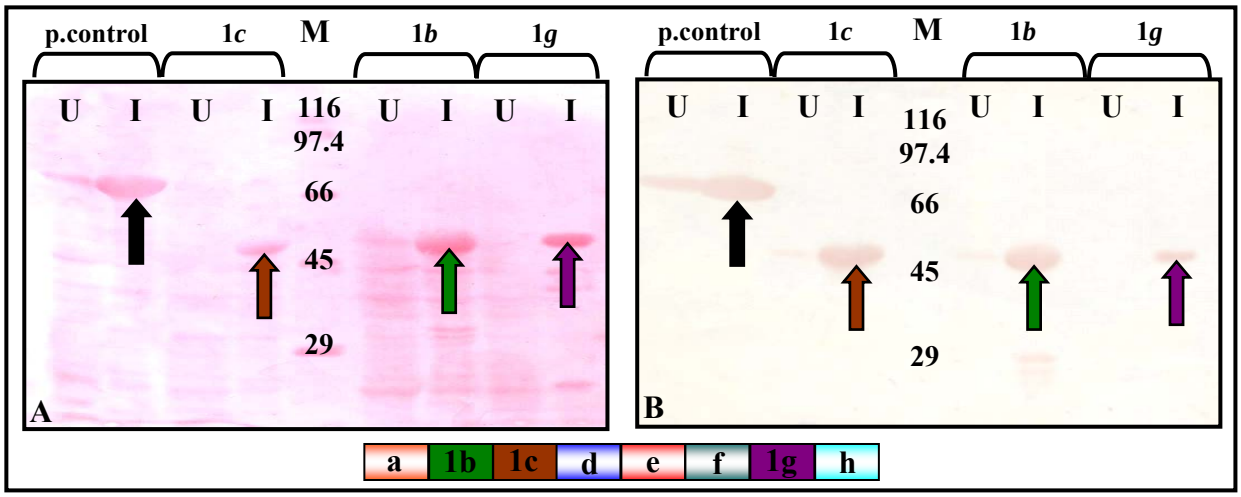


Fig. 38: Western blot. **A:** Ponceau S staining of the transferred proteins of the induced (I) and uninduced (U) cultures of the expression of the positive control protein and KLH1-*c*, -1*b* and -1*g*, where **B**, the recombinant proteins (black, brown, green and purple arrows, respectively) subsequently detected by the α-His-antibody. M represents the marker, in kDa.

The α-His-antibody recognizes all recombinantly expressed proteins in the respective induced cultures (Fig. 38B, black arrow: β-glucuronidase, 80 kDa, brown arrow: KLH1-*c*, 54 kDa, green arrow: KLH1-*b*, 52 kDa and purple arrow: KLH1-*g*, 50 kDa.

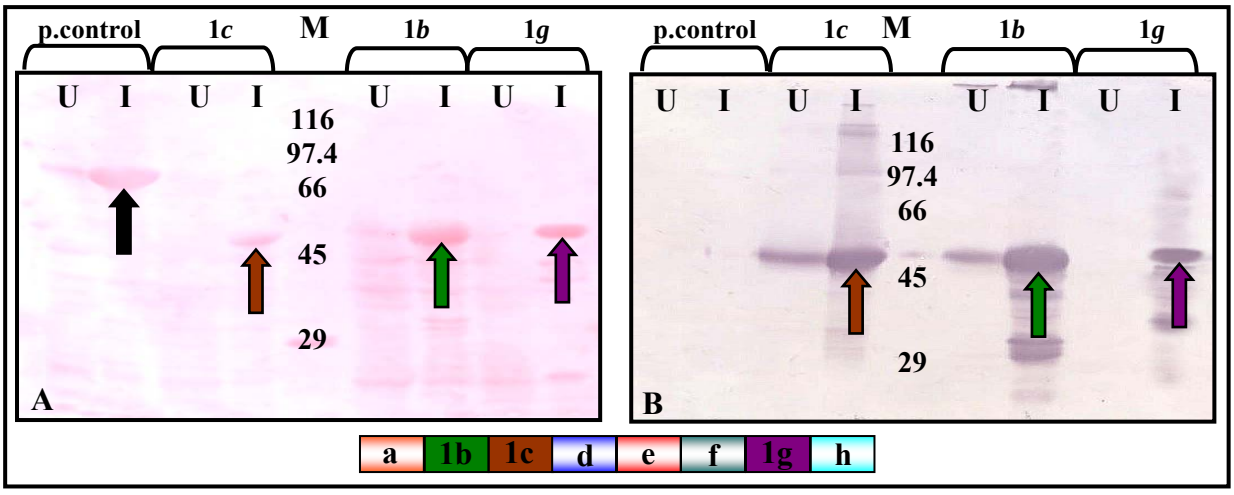


Fig. 39: Western blot. **A:** Ponceau S staining. **B:** Specific detection of the recombinantly expressed KLH1 FUs by the α-KLH1-antibody (brown, green and purple arrows, respectively). M represents the marker, in kDa.

The Western blot analysis of the expression of the FUs KLH1-*b* and -1*g*, as well as of the additionally applied positive controls KLH1-*c* and β-glucuronidase depicted in the Figures 39 and 40 shows clearly that the expressed proteins are well recognized by both antibodies (except the positive control protein β-glucuronidase). While the positive control protein β-glucuronidase has a distinct band at approximately 80 kDa (Fig. 38A and B and Fig. 39A, black arrows), the FU KLH1-*c* reveals the expected protein band, at

approximately 54 kDa (brown arrows). The two FUs KLH1-*b* (green arrows) and -*g* (purple arrows) show also positive recognition by both antibodies (Fig. 38A and B and Fig. 39A and B). In the induced culture of KLH1-*bc* also smaller degradation products are detected by the antibody.

9 Expression of single FUs of KLH2 as comparison to KLH1

9.1 Expression experiment

To check whether the expression of the single FUs of KLH2 behaves in a similar way, each expression clone encoding a single FU of KLH2 was also first transformed into the *E. coli* host BL21-AI™ and the recombinant proteins expressed following the above mentioned standard protocol. Aliquots of the induced (I) and uninduced (U) cultures were were applied to a large SDS-PAGE. In parallel, the positive control protein β-glucuronidase was also expressed.

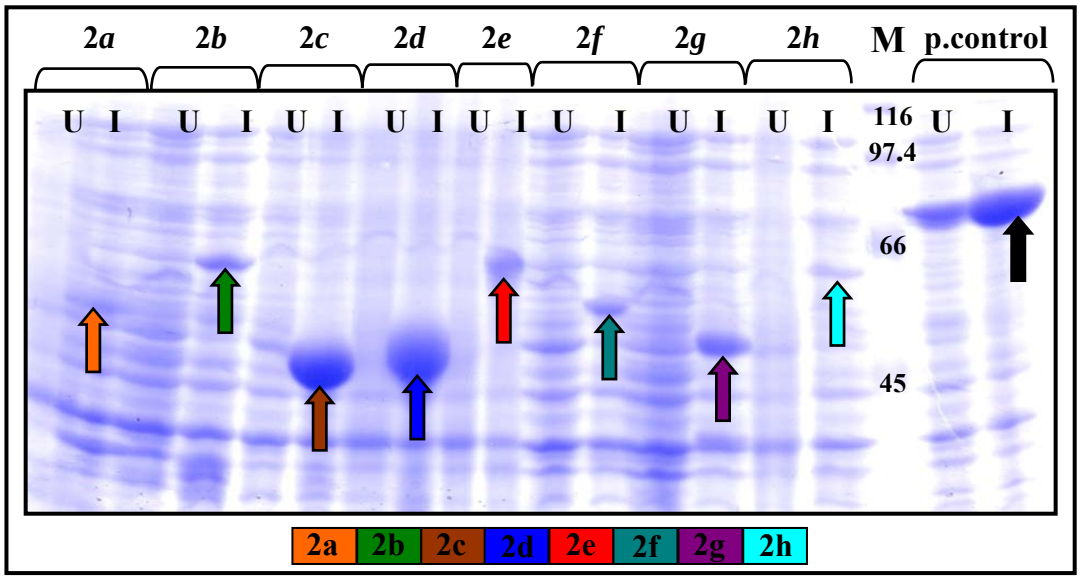


Fig. 40: SDS-PAGE of the expression of the single FUs of KLH2. All induced cultures of KLH2 shows a distinct band of recombinant protein at the expected mass. M represents the marker, in kDa.

The expression of the single FUs of KLH2 leads to the following result: the clones KLH2-*b*, -*c*, -*d*, -*e*, -*f* and -*g* express a strong protein band at the expected mass. The FU KLH2-*b* is represented by a protein band of about 62,682 Da, because the applied down-stream primer for the RT-PCR was located such that almost 300 basepairs of the con-secutive FU KLH2-*c* were coamplified. This shifts the expected average molecular mass of this FU from around 50 kDa to more than 62 kDa (Fig. 40, green arrow). KLH2-*c* is situated as expected at a molecular mass of 50,592 Da (brown arrow). The induced cul-ture of KLH2-*d* has an analogous strong protein band, as for KLH2-*c*, but of slightly

larger mass, namely 52,170 Da (blue arrow). The calculated theoretical mass of the recombinantly expressed KLH2-*e* is 59,687 Da (red arrow), when it is taken into account that the reverse primer, used in the RT-PCR reaction to create the corresponding cDNA, had 184 bp bound within KLH2-*f*. While the expected mass of KLH2-*f* itself is ~ 53,380 Da (seagreen arrow), KLH2-*g* is 52,066 Da (purple arrow). As shown in Figure 40, the induced cultures of KLH2-*a* and -2*h* have only very weak and blurred bands at the expected masses of 55,276 Da (orange arrow) and 64,435 Da (turquoise arrow), respectively. The positive control protein, in contrast, exhibits a very strong protein band at 80 kDa (black arrow).

9.2 Immunological analysis

The weak expression of the FUs KLH2-*a* and -2*h* was counterchecked by Western blot analysis. Therefore, all samples were applied two more times to large SDS-PAGEs and the electrophoresed proteins were electro-transferred onto nitrocellulose membranes. After a Ponceau S staining of both membranes (Fig. 41A and 42A), the first was incubated with a monoclonal α-His-antibody (Fig. 42B), while the proteins of the second membrane were detected by a polyclonal α-KLH2-antibody (Fig. 42B).

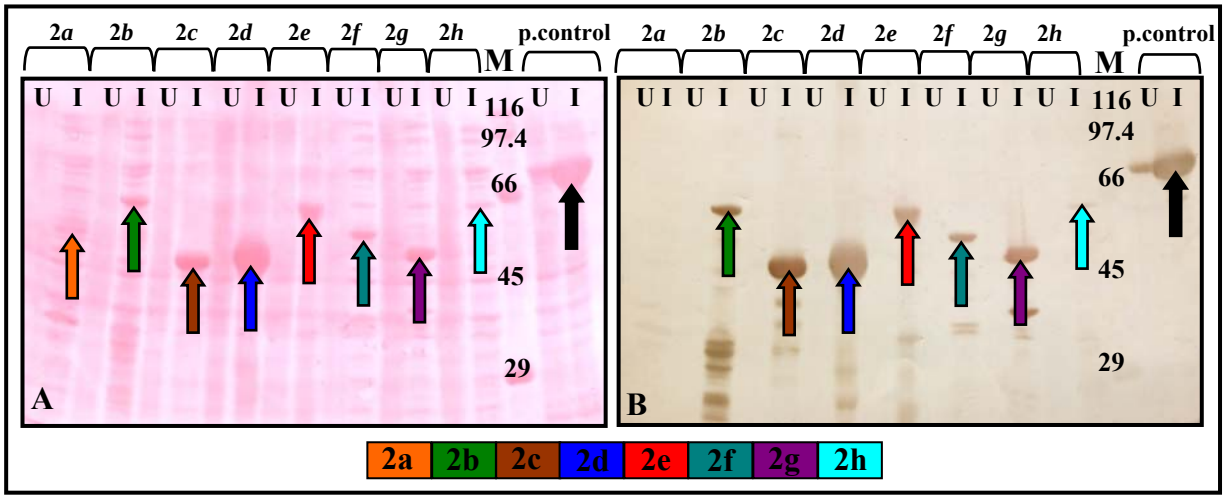


Fig. 41: Western blot. **A**: Ponceau S staining of the transferred proteins of KLH2. **B**: Detection of the induced (I) and uninduced (U) cultures of the single FUs with a monoclonal α-His-antibody. Except for KLH2-*a*, each induced culture shows a band of recombinant protein at the expected mass. M represents the marker, in kDa.

The Ponceau S staining (Fig. 41A) and the corresponding detection with the monoclonal α-His-antibody (Fig. 41B) of the induced cultures (I) compared to the respective uninduced (U) cultures show an interesting result. For the FUs KLH2-*b* to -2*h* as well as for the positive control protein β-glucuronidase, each induced culture shows a distinct band of recombinantly expressed protein at the expected mass, which is not present in the

uninduced culture. One exception remains: KLH2-*a*. No detection occurs at the expected mass of 55 kDa on the membrane, but only very weak bands of smaller degradation products are seen.

The above mentioned second nitrocellulose membrane incubated with the α-KLH2-antibody shows a slightly different result (Fig. 42A and B).

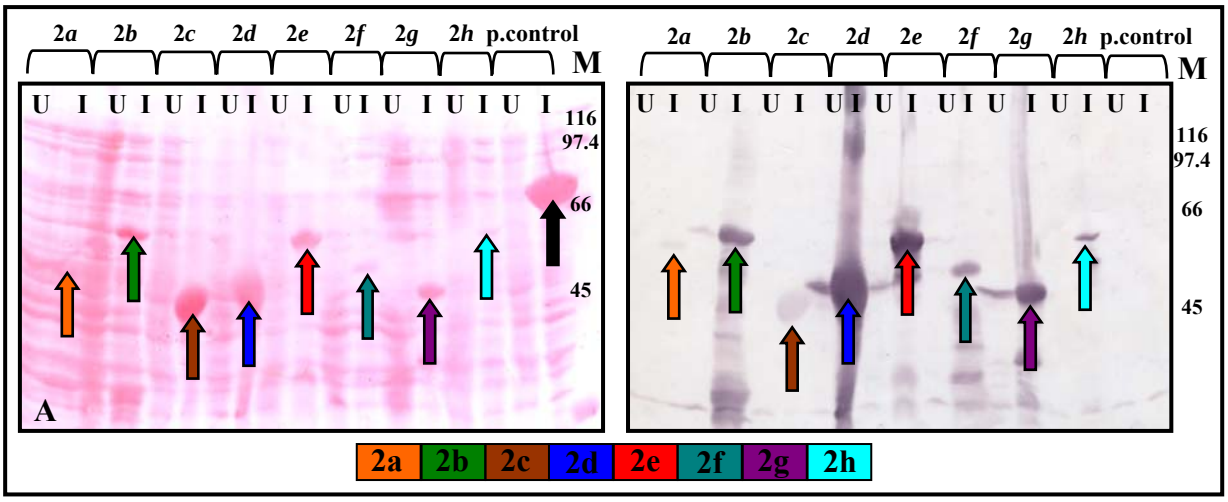


Fig. 42: Western blot. **A:** Ponceau S staining of transferred proteins. **B**: Specific detection using a polyclonal α-KLH2-antibody, confirming the result, that each clone of KLH2 expresses a recombinant protein band at its expected mass. M represents the marker, in kDa.

The Ponceau S staining in Fig. 42A, reveals for all FUs of KLH2 and for the positive control protein β-glucuronidase defined protein bands at the expected mass within the induced culture (I) which are not present in the uninduced (U) culture. Interestingly, specific detection by the α-KLH2-antibody occurs for each of the eight induced cultures of KLH2 expressing clones and at the expected mass. While the recombinant protein bands of the FUs KLH2-*a* and -2*c* are represented by very weak precipitation products (Fig. 42B, orange and brown arrows, respectively), the remaining FUs of KLH2 are detected in a very specific way by the α-KLH2-antibody (KLH2-*b*, green arrow; -2*d*, blue arrow; -2*e*, red arrow; -2*f*, seagreen arrow; -2*g*, purple arrow and -2*h*, turquoise arrow). As expected, the positive control is not recognized by the α-KLH2-antibody (Fig. 42B).

10 Recombinant expression of a KLH1 substructure containing two consecutive FUs: KLH1-*ef*

Knowing now that the expression of single FUs of KLH1 and KLH2 works well, the expression of clones encoding two consecutive FUs was attempted. The first one was KLH1-*ef*. The expression clone was transformed into the expression bacterial strain BL21-AI™. The expression of the protein was monitored by comparing induced cultures with uninduced cultures, as mentioned before. Aliquots of both cultures were taken and

applied to an SDS-PAGE. Coomassie staining, however, showed that only a small protein band at 26 kDa appeared within the induced cultures, which was not present in the uninduced cultures. A protein band at the expected calculated theoretical molecular weight of 102,318 Da for the two FUs KLH1-*ef* was not present (data not shown).

11 Recombinant expression of KLH1-*bc*

11.1 Expression experiment

The cDNA of KLH1-*bc* was expressed following the above mentioned protocol and aliquots of both cultures were applied on an SDS-PAGE (Fig. 43).

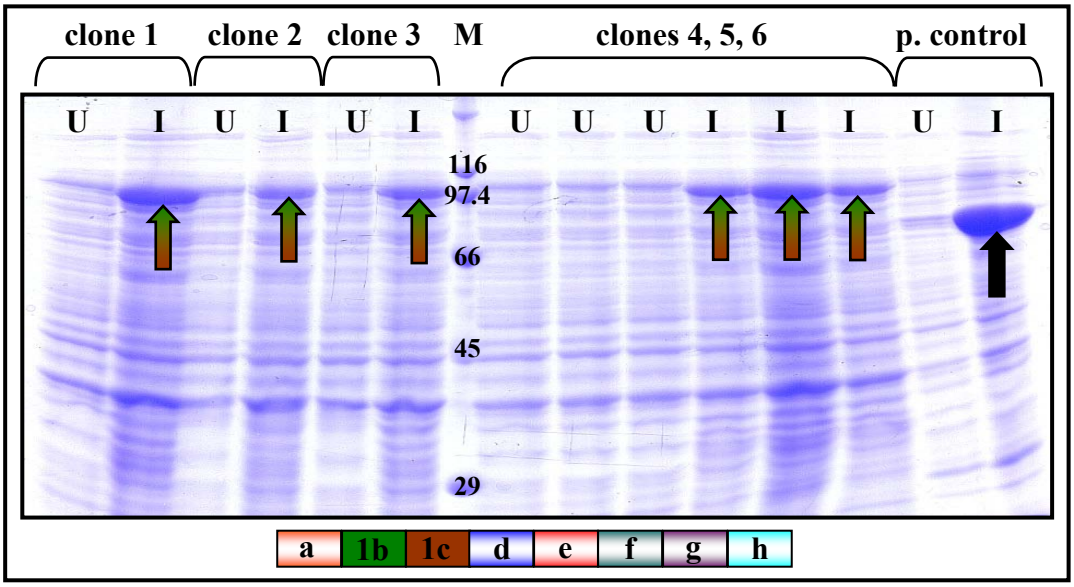


Fig. 43: Coomassie staining of the recombinant expression of KLH1-*bc*. All induced (I) cultures show a distinct band at approximately 100 kDa (green-brown arrows), which is not present in the uninduced (U) cultures. The positive control protein β-glucuronidase shows the expected band at 80 kDa (black arrow).

As can be seen in Fig. 43, the expression of the clones comprising KLH1-*bc* produces each a discrete band at the expected mass of approximately 100 kDa (green-brown arrows; the exact mass is 101,568 Da), which is only visible in the induce (I) cultures. The positive control protein β-glucuronidase also shows a specific protein band of 80 kDa, which is only present in the induced culture (black arrow).

11.2 Immunological analyses

To check the specificity of the seen bands, the proteins were electrophoresed twice and each transferred onto nitrocellulose membranes, which were incubated once with an α-His-antibody and once with a polyclonal α-KLH1-antibody. The results are shown in Fig. 44 (A and B) and 45 (A and B), respectively.

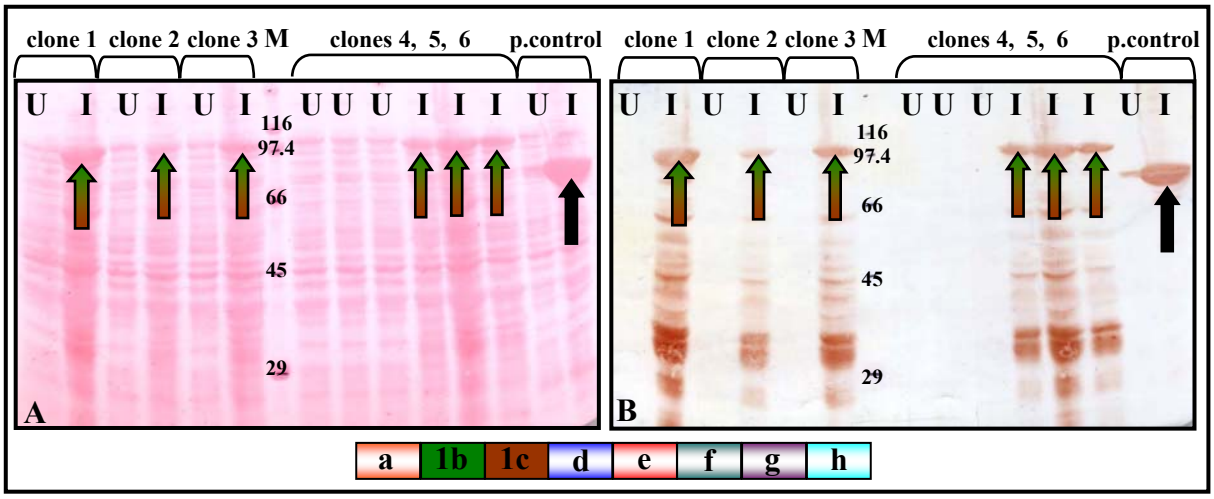


Fig. 44: Western blot of KLH1-*bc*. **A:** Ponceau S staining of the blotted proteins from the KLH1-*bc* expression. **B:** The α-His-antibody detects not only the recombinant protein band of KLH1-*bc* at 100 kDa (green-brown arrows), but also smaller degradation products, as well as the positive control protein at 80 kDa (black arrow).

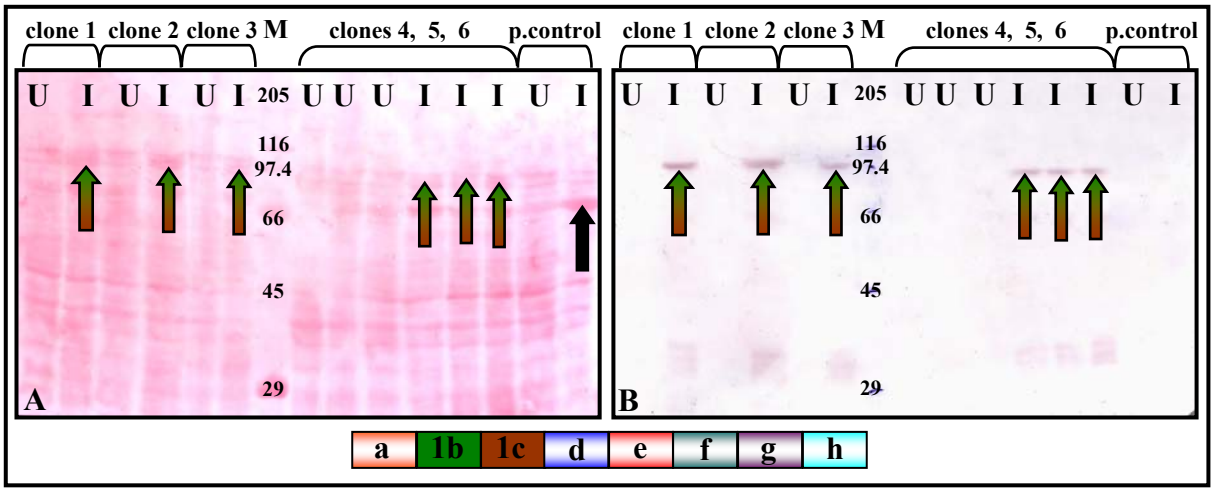


Fig. 45: Western blot of KLH1-*bc*. Detection of the recombinantly expressed protein (green-brown arrows) at 100 kDa by a polyclonal α-KLH1-antibody (**B**) and its appropriate Ponceau S staining (**A**).

The Western blots of the KLH1-*bc* expression (Figs. 44A and B, and 45A and B) shows that both antibodies recognize not only the expected protein band at 100 kDa (green-brown arrows), but also many smaller degradation products, especially for the α-His-antibody. However, these are only present in the induced (I) cultures of the expressing clones. The positive control protein β-glucuronidase is also recognized by the α-His-antibody, due to its coexpressed N-terminal 6xHis-tag (black arrow, Fig. 44B), but not by the α-KLH1-antibody (Fig. 45B).

12 Generation of additional clones coding for two consecutive FUs of KLH1 by SOE-PCR and expression

12.1 SOE-PCR

Although the expression of the cDNA of KLH1-*ef* did not work, the successful expression of KLH1-*bc* showed that the failure was not due to the mass of the recombinant protein. Therefore, generation of additional cDNAs encompassing two consecutive FUs of KLH1 by using the method of SOE-PCR (**S**plicing by **O**verlap **E**xtension, refer to: B Materials and methods, *5.4 Polymerase chain reaction and its variations,* - SOE-PCR) and their recombinant expression was attempted.

During SOE-PCR cycles, two genes or cDNAs can be fused together, without the use of restriction enzymes. One indispensable prerequisite, therefore, is that both DNA fragments must have overlapping, complementary sequences to each other, at opposite ends. These overlapping sequences can already be present, or they can be introduced by primers, by a previous PCR.

Knowing this at the beginning of the present work, the gene-specific primers used for RT-PCR to generate the KLH encoding cDNAs were chosen, to bind shortly before the 3' end of the previous or closely after the 5' end of the adjacent FU. The KLH cDNAs encompassing one or two FUs, thereby, had in fact always a small portion of the previous and following FU added at the 5' and 3' end, respectively. This can now be taken as an advantage for the SOE-PCR. In general, all previously generated clones were taken as templates to amplify the wanted FUs by PCR using only one attachment primer in combination with the corresponding gene-specific primer. After checking the right mass of the resulting PCR products by agarose gel, the PCRs were either applied directly or the cDNAs were purified from remaining nucleotides and primers by PCR clean-up and taken as templates for the subsequent SOE-PCR. This one was then performed using the outer attachment primer pair which thereafter enables the cloning of the newly fused gene.

Two additional cDNAs were generated by this way: KLH1-*ab* and KLH1-*cd*, by fusing the cDNAs encoding KLH1-*a* with KLH1-*b* and the cDNA of KLH1-*c* with KLH1-*d*, respectively (data not shown). The subsequent agarose gel showed that the occurring cDNAs had the expected mass corresponding to a cDNA encoding two consecutive FUs of KLH1, including the coamplified attachment sites, required for cloning. The cDNAs

were cloned into the vector pDONR™221, checked for the right insert mass and the isolated plasmids were subcloned into the expression vector pDEST™17.

12.2 Expression experiment

The clones encoding KLH1-*ab* and -1*cd* were transformed into the expression bacterial strain BL21™-AI and the expression took place as mentioned above. Aliquots of the induced and uninduced cultures were taken and analysed by SDS-PAGE. The Coomassie staining, however, revealed that no recombinant expression had taken place: no additional protein band was visible within the induced culture at the expected mass. This was observed for the expression of KLH1-*ab* as well as for the expression of KLH1-*cd* (data not shown).

13 Recombinant expression of KLH1-*gh*

13.1 Expression experiment

The clone encoding KLH1-*gh* was transformed by [redacted] into BL21™-AI, and five clones were picked and cultured under the same conditions as mentioned above. The result is depicted in Figure 46.

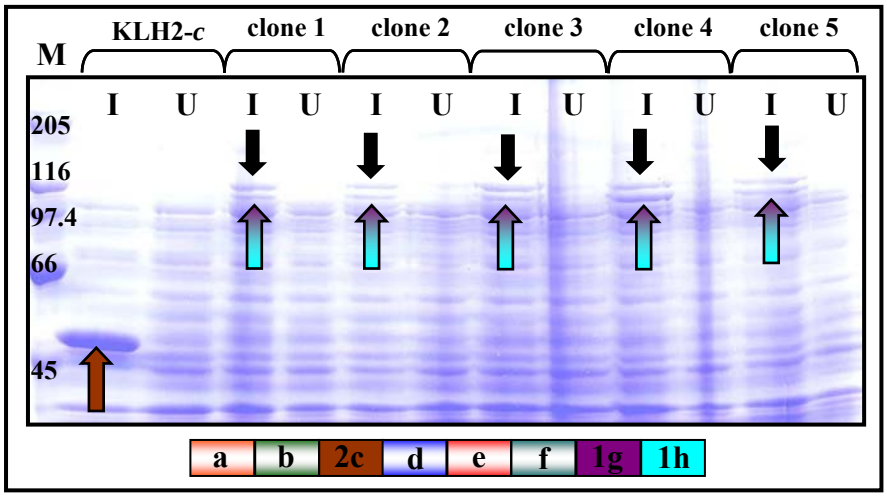


Fig. 46: SDS-PAGE of the expression of the FUs KLH1-*gh*. Two additional bands are shown in the induced (I) cultures, one at the expected mass of 110 kDa (purple-turquoise arrows) and a second one at approximately 120 kDa (black arrows). As a positive control, the simultaneous expression of KLH2-*c* is shown (brown arrow).

The expression experiment of KLH1-*gh* shows a weak band at the expected mass of 110 kDa (exactly 109,914 Da, including the 6xHis-tag), which is only present in the induced (I) cultures of the expressing clones (Fig. 46, purple-turquoise arrows). Interestingly, a second additional band at approximately 120 kDa appears also in the induced cultures, which is not present in the uninduced cultures (black arrows). As a

positive control, the single FU KLH2-*c* is simultaneously expressed, to check the quality of the expression. The resulting protein band of 50 kDa corresponds to the expected mass (brown arrow).

13.2 Immunological analysis

To elucidate if the second appearing band of 120 kDa was derived from the recombinant expression, the samples were applied again on a large SDS-PAGE and the electrophoresed proteins were transferred by Western blot onto a nitrocellulose membrane. This membrane was then incubated with an α-His-antibody (Fig. 47).

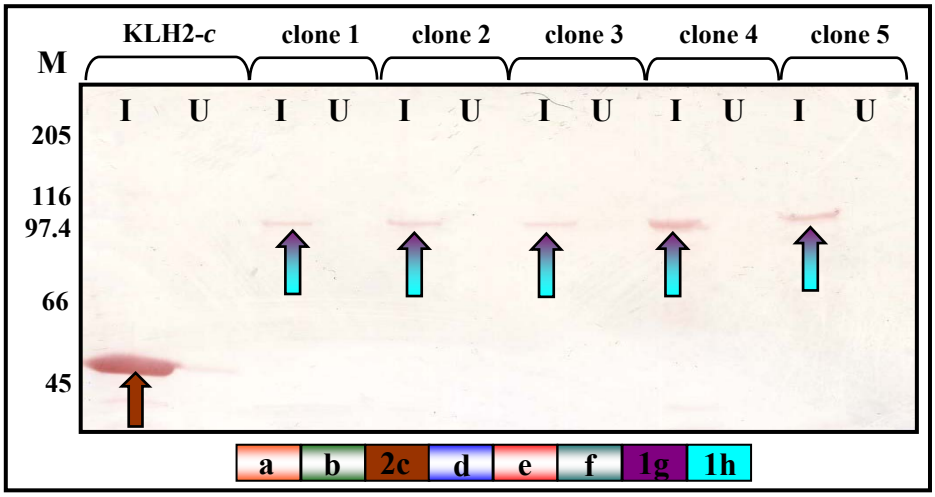


Fig. 47: Western blot of KLH1-*gh*. The protein band at 110 kDa within the induced cultures of the KLH1-*gh* expressing clones is detected by the α-His-antibody (purple-turquoise arrows). The positive control KLH2-*c* is also present (brown arrow).

The incubation of the membrane with the monoclonal α-His-antibody (Fig. 47) reveals in a specific way one of the two additional protein bands, namely the expected one at 110 kDa (purple-turquoise arrows). While the Coomassie staining in Fig. 46 shows two additional bands in the induced cultures of the KLH1-*gh* expressing clones, the immunoblot clarifies the fact, that only the smaller protein represents the recombinantly expressed substructure KLH1-*gh*.

All experiments concerning the expression and blotting of KLH1-*gh* were kindly done by [redacted] Although none of the trials to reproduce these results succeeded, it was achieved by [redacted] to express these clones in parallel in eukaryotic insect cells, wherefore the repetition of the expression was not further proceeded.

14 Testing the recombinant protein expression capacity of *E. coli* using a clone encoding for three FUs: KLH1-*abc*

14.1 Generation of the cDNA of KLH1-abc, and cloning

During this work it was already possible to express proteins recombinantly in *E. coli* of 100 kDa and 110 kDa (KLH1-*bc* and KLH1-*gh*, respectively). To test where the limit lies, a clone was generated encompassing three consecutive FUs, which will result in a protein of about 150 kDa. After the first two successful SOE-PCRs generating KLH1-*ab* and KLH1-*cd*, the next one was done to fuse the cDNA encoding KLH1-*a* with the cDNA encompassing KLH1-*bc*, to generate KLH1-*abc*. After amplifying each cDNA from its respective entry clone and checking the result by agarose gel electrophoresis, both PCR reactions were applied directly for the SOE-PCR. Once the cDNA was generated (Fig. 48A), it was cloned into the vector pDONR™221 to create the appropriate entry clone. Overnight grown clones were checked by clone-PCR using the vector specific primers M13 forward and reverse, so that the resulting fragment will contain 292 additional base-pairs (Fig. 48B).

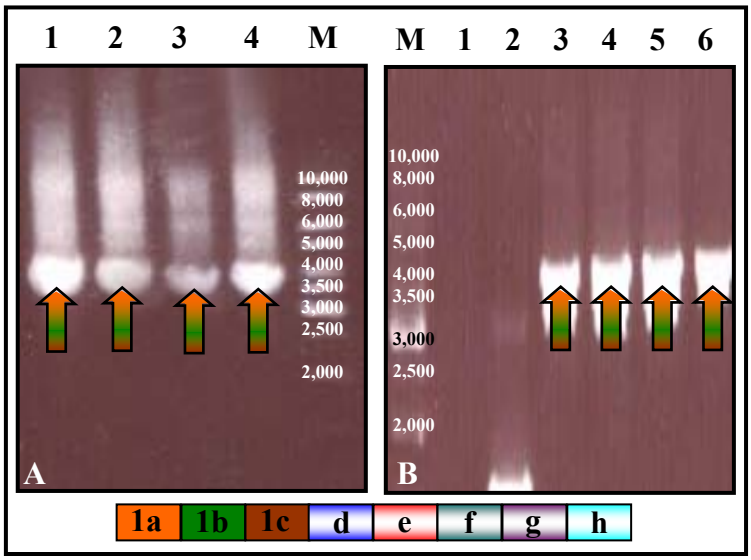


Fig. 48: A: Agarose gel of the SOE-PCR to fuse the cDNAs of KLH1-*a* and -1*bc* together, generating a cDNA of approximately 3,800 bp encoding KLH1-*abc* (orange-green-brown arrows). **B**: Clone PCR confirming the correct mass of the insert in pDONR™221, creating the appropriate entry clones (orange-green-brown arrows).

The gel electrophoresis depicted in Fig. 48A, shows that the SOE-PCR was successful in generating a cDNA of the expected approximately 3,800 bp (orange-green-brown arrows). This cDNA fragment represents KLH1-*abc*. The direct cloning of the PCR reaction led to several clones, and the clone-PCR (Fig. 48B) shows that four of them contain

an insert at the expected 4,100 bp (orange-green-brown arrows). Sequencing using vector and gene specific primers confirmed the obtained results.

14.2 Expression experiment of KLH1-abc

After a plasmid preparation, the cDNA encoding KLH1-*abc* was subcloned into pDEST™17 to create the appropriate expression clone. The following steps were done by [redacted]: after transformation of the expression clone into BL21™-AI, five clones were picked and the expression was done under standard conditions as mentioned before. Aliquots of the induced and uninduced cultures were taken and analysed by SDS-PAGE (Fig. 49).

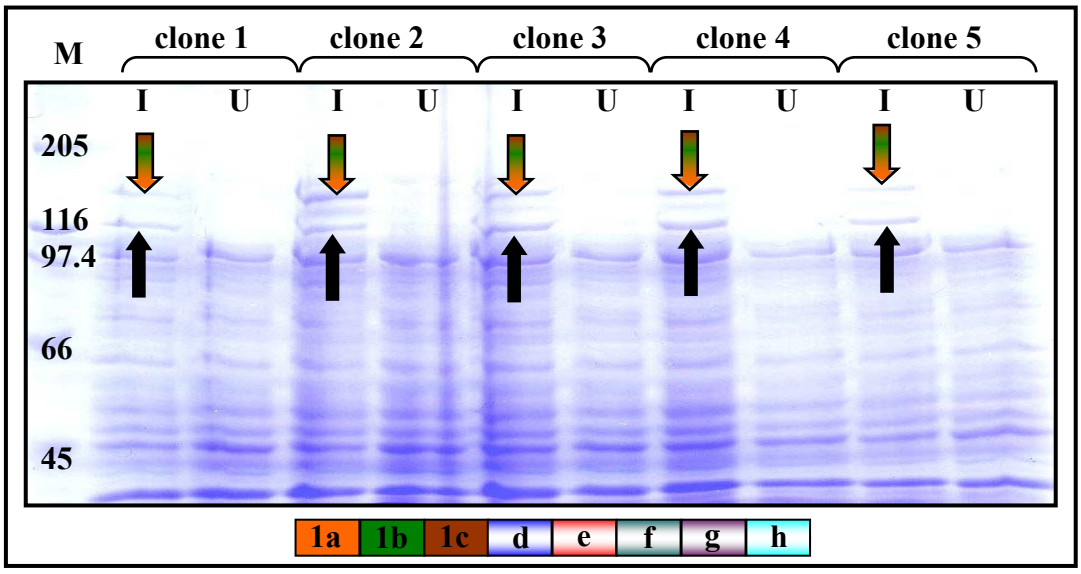


Fig. 49: SDS-PAGE of the expression of KLH1-*abc*. Beside a band of 120 kDa (black arrows), the induced (I) cultures also show a weak band at the expected mass of approximately 150 kDa (orange-green-brown arrows), which is not present in the uninduced (U) cultures. M represents the marker, in kDa.

Although it is well known that prokaryotic systems like *E.coli* have a limited expression capacity, a weak band appears in the induced cultures in the SDS-PAGE shown in Fig. 49 at the expected mass of approximately 150 kDa (orange-green-brown arrows). The calculated theoretical molecular weight of a protein containing the three FUs KLH1-*abc* and including the 6xHis-tag is about 147,503 Da. Interestingly, an additional band, having a mass of 120 kDa, is also present in the induced cultures, which is not apparent in the uninduced cultures (black arrows).

14.3 Immunological analysis

To check which one of the two additional protein bands seen in the Coomassie staining corresponded to the recombinantly expressed KLH1-*abc*, a Western blot was done using

monoclonal α-His-antibodies, which shall detect the coexpressed N-terminal 6xHis-tag of the recombinant protein. The results are shown in Fig. 50.

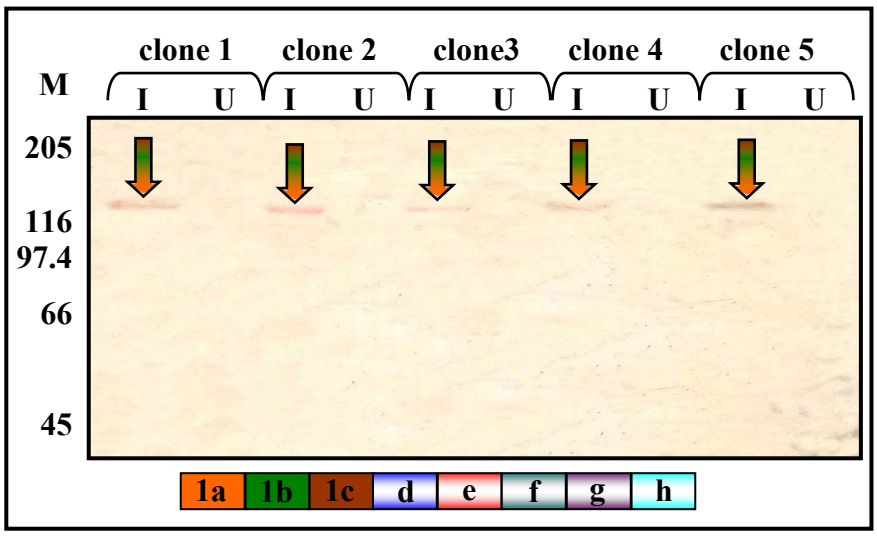


Fig. 50: Western blot of the expression of the KLH1-*abc* encompassing clones, incubated with a monoclonal α-His-antibody. Specific detection of a protein band at approximately 150 kDa occurs in the induced (I) cultures (orange-green-brown arrows).

The Western blot analysis of the KLH1-*abc* expression shows that only the band at 150 kDa is detected by the α-His-antibody (Fig. 50, orange-green-brown arrows) in contrast to the band at 120 kDa which is no longer seen. All the expression experiments concerning KLH1-*abc* were kindly done by [redacted]. Although none of the trials to reproduce these results succeeded, it was achieved by [redacted] to express these clones in parallel in eukaryotic insect cells, wherefore the repetition of the expression was not further proceeded.

15 Overview of all recombinantly expressed KLH1 and KLH2 substructures in *E. coli*

Figure 51 shows an overview of all the KLH1 (A) and KLH2 (B) FUs that were expressed recombinantly in *E. coli*.

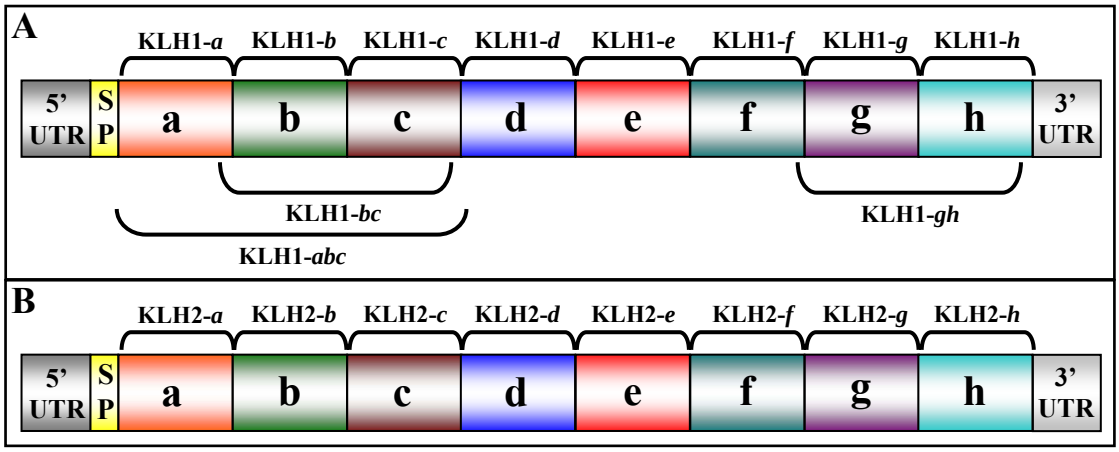


Fig. 51: Schematic overview of all the recombinantly expressed FUs of KLH1 (**A**) and KLH2 (**B**) in *E. coli*.

The following Table 11 shows an overview of all the KLH1 and KLH2 cDNAs, generated by RT-PCR and PCR, as well as the corresponding clones, in combination with the expressed proteins. One exception is KLH1-*abc* which was generated by SOE-PCR.

KLH1	1*a*	1*b*	1*c*	1*d*	1*e*	1*f*	1*g*	1*h*	1*ab*	1*bc*	1*cd*	1*de*	1*ef*	1*fg*	1*gh*	1*abc*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+
clone	+	+	+	+	+	+	+	+	-	+	-	-	-	+	+	+
expr.	+	+	+	+	+	+	+	+	-	+	-	-	-	-	+	+

KLH2	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	2*g*	2*h*	2*ab*	2*bc*	2*cd*	2*de*	2*ef*	2*fg*	2*gh*	2*abc*
cDNA	+	+	+	+	+	+	+	+	-	+	-	+	-	-	-	-
clone	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
expr.	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-

Table 11: Overview of all the KLH1 and KLH2 cDNAs, generated by RT-PCR and PCR, combined with the respective clones and the corresponding expressed proteins. One exception is KLH1-*abc* which is generated by SOE-PCR ("+": available; "-": not available).

16 Subsequent purification of the recombinantly expressed proteins

All KLH1 and KLH2 cDNAs encoding single FUs and larger KLH substructures had been subcloned into the destination vector pDEST™17. The presence of the N-terminal 6xHis-tag in this vector allows affinity purification using a nickel-chelating resin, such as Ni-NTA (**N**itrilo**T**riacetic **A**cid, 4 binding sites). Therefore, the protocol described in the handbook "The Qiaexpressionist™", from Qiagen (Hilden), was followed. The purification can thereby be done under native or under denaturing conditions.

The first experiment was done using the expression culture of KLH2-*c* in BL21™-AI. Aliquots of the supernatant, the flow-through, both wash steps and both elution steps were taken and analyzed by SDS-PAGE. While no protein bands were visible within both elution steps of the native purification, very weak bands were detectable by Coomassie staining in both elution fractions of the denaturing purification (data not shown). However, it was not possible to reproduce these results.

Therefore, Ni-IDA (**I**mino**D**iacetic **A**cid, 3 binding sites) and Ni-TED (**T**ris-carboxymethyl **E**thylene **D**iamine, 5 binding sites) spin columns from Macherey-Nagel (Düren) were used for native purification of the expression culture of KLH1-*bc*. But the SDS-PAGE containing samples of each purification step showed that the specific bands of recombinant protein as well as other bacterial proteins were found within the flow-through and the first wash fraction (data not shown).

Hence, the purification of KLH1-*b*c was tried under denaturing conditions using again the Ni-NTA spin columns from Qiagen (Hilden). The denaturing conditions should facilitate the binding of the 6xHis-tag to the Ni-NTA matrix: the tag should now be better available due to the unfolded protein structure. But the analysis by SDS-PAGE showed again that the wanted protein band representing the KLH1 substructure was found in the flow-through (data not shown).

16.1 Enhancing the protein solubility by coexpression of the fusion protein glutathione-S-transferase (GST)

The cDNAs encoding the FUs KLH1-*h* and KLH2-*c* as well as KLH1-*bc* were subcloned from the respective entry clones into the destination vector pDEST™15, which contains an N-terminal GST-tag. After transformation of the plasmids into BL21™-AI, the expression was done using the above mentioned standard protocol. Aliquots of the induced and uninduced cultures were taken and analyzed by SDS-PAGE (results not shown). The Coomassie staining of the SDS-PAGE showed that the expression experi-

ments worked well: each induced culture showed an additional distinct protein band at the expected mass which was not present in the uninduced cultures. The coexpression of the N-terminal GST-tag shifts the respective recombinant protein about 27.9 kDa, so that KLH2-*c* was represented by a band of approximately 80 kDa, and KLH1-*h* had now a molecular weight of nearly 90 kDa (data not shown). The expected mass for the two consecutive FUs KLH1-*bc* of approximately 100 kDa was also shifted, namely to nearly 130 kDa (data not shown).

16.2 Purification of N-terminal GST-tagged, recombinant KLH FUs using affinity chromatography

The fusion with GST enables the purification using affinity chromatography. Therefore, prepacked columns from Sigma (Deisenhofen) were used, whose resin consists of glutathione attached through the sulphur to epoxy-activated 4 % cross-linked beaded agarose. This permits mild, non-denaturing and highly selective purification of glutathione binding enzymes such as glutathione-S-transferase, glutathione peroxidase, and glyoxylase I.

The expression culture of KLH1-*bc* was taken and processed following the manufacturer's instructions. The bacterial cell pellet was lysed in the appropriate lysis buffer and, after a centrifugation step, the resulting supernatant was applied on a pre-equilibrated column. The flow-through was collected and after four wash steps, the proteins were eluted by three passages of elution buffer over the column. Samples of each step were collected and analyzed by SDS-PAGE. The Coomassie staining of this gel showed that the recombinant protein band, representing KLH1-*bc* of about 130 kDa, was present in the applied lysate, additionally to the bacterial proteins. However, after passing the column, nearly the complete protein fraction was found in the flow-through and within the first wash step (data not shown). While the first two elution steps revealed no proteins, the third and fourth elution steps contained few proteins of smaller mass, due to non-specific binding. These trials lead to the suggestion that the recombinant proteins must be located in inclusion bodies, thus the expression cultures of GST-tagged KLH1-*h*, KLH2-*c* and KLH1-*bc* were not further analysed.

17 Determination of target protein solubility

During recombinant expression of proteins in bacteria cells, many proteins are packed as inclusion bodies when they are expressed at high levels, while other proteins are tolerated well by the cells and remain in the cytoplasm in their native configuration.

To set up the best purification strategy, it is important to determine whether the protein is soluble in the cytoplasm or located in cytoplasmic inclusion bodies. Therefore a protocol from the handbook "The Qiaexpressionist™", from Qiagen (Hilden), was used and the instructions provided were followed. 50 ml of an expression culture of KLH1-*bc* were taken for this experiment. After centrifugation, the cells were resuspended in lysis buffer for native purification. Lysozyme was added and the sample incubated on ice. The lysate was then sonicated and a subsequent centrifugation step enabled separation of the supernatant (soluble protein) from the pellet (insoluble protein). This pellet was then also resuspended in lysis buffer. Aliquots of the induced (I) and uninduced (U) culture, as well as from the supernatant (SN) and from the pellet (P) were applied twice to an SDS-PAGE, first for Coomassie staining and second for Western blot analysis by specific detection using an α-KLH1-antibody. The result determines whether the recombinant protein is located within the soluble protein fraction (supernatant, SN) or within the insoluble fraction, the pellet (P).

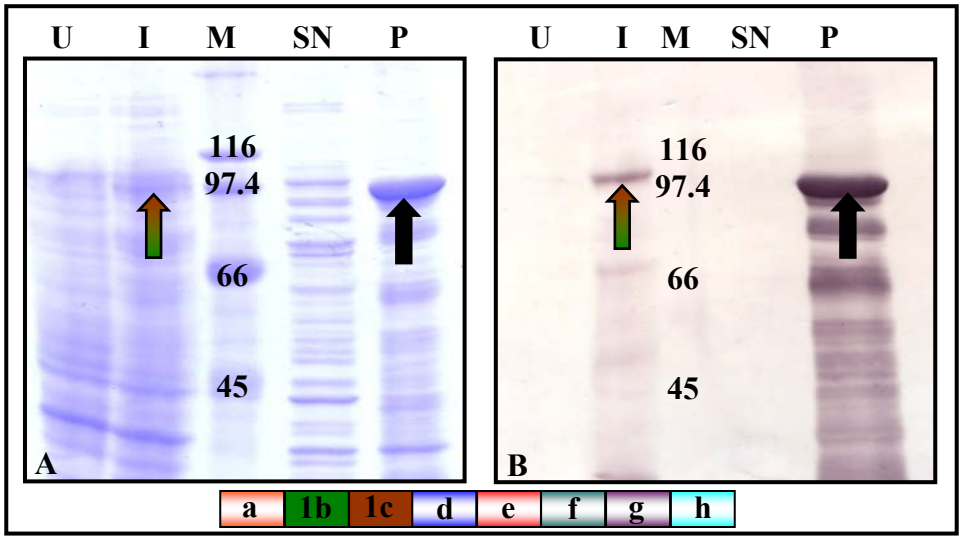


Fig. 52: Determination of target protein solubility using the expression culture of KLH1-*bc*. Samples of the induced (I) and uninduced (U) cultures as well as from the supernatant (SN) and from the pellet (P) were applied twice to an SDS-PAGE. **A:** Coomassie staining. **B:** Specific detection of the protein band at 100 kDa with an α-KLH1-antibody (green-brown arrows). The recombinant protein is located within the insoluble fraction (pellet, P, black arrow).

As seen in the Coomassie staining in Figure 52A, the recombinantly expressed protein, present in the induced culture as a blurred band at the expected mass of 100 kDa (green-brown arrow), represents KLH1-*bc*. After the treatment, including digestion with lysosyme, sonication and centrifugation, the applied aliquots of supernatant and pellet show that the recombinant protein is found within the pellet (black arrow). The Western blot analysis confirms this result: although the blurred band of recombinant protein of the

induced (I) culture was not visible in the Ponceau S staining (data not shown), it is clearly recognized by the polyclonal α-KLH1-antibody (Fig. 52B, green-brown arrow), as is the thick protein band located in the pellet (black arrow). The supernatant (SN), in contrast, contains many bacterial proteins, which are not detected by the antibody. This provides an indication of insolubility and of the location of the recombinant proteins in inclusion bodies.

18 Isolation of proteins from inclusion bodies

18.1 Establishing the method using different protocols

To be able to purify the recombinant protein from inclusion bodies, different approaches were applied, such as variation of the lysozyme concentration at different incubation temperatures, the use of native lysis buffer under repeated addition of Tween-20 and EGTA, different sonication durations and repetitions, the use of denaturing conditions like 6 M Guanidinuim hydrochloride (GuHCl) with varying incubation times, different concentrations of Triton X-100, SDS, DTT, alone and in combination, respectively, and also the alkaline lysis using the E.Z.N.A plasmid Miniprep Kit I from peqlab, Erlangen, was tried. However, the band of 100 kDa representing the recombinantly expressed KLH1-*bc* remained located in the pellet.

So, two additional protocols were tested:

- **Protocol 1:** Purifying proteins from inclusion bodies (from: http://structbio. vanderbilt.edu/chazin/wisdom/labpro/inclusion.html)
- **Protocol 2:** Purification of active eukaryotic proteins from the inclusion bodies in *E. coli* (from: http://www.dwalab.com/labman/op5.html)

Both protocols were based on the denaturing forces of 8 M urea. The expression culture of KLH1-*bc* was up-scaled to 50 ml, without changing any additional parameter, and was treated following the instructions mentioned at the respective web sites.

The aliquots of the arising supernatants and pellets produced during the different purifications were applied onto different SDS-PAGEs and stained with Coomassie (Fig. 53, purification using protocol 1; Fig. 54, purification using protocol 2).

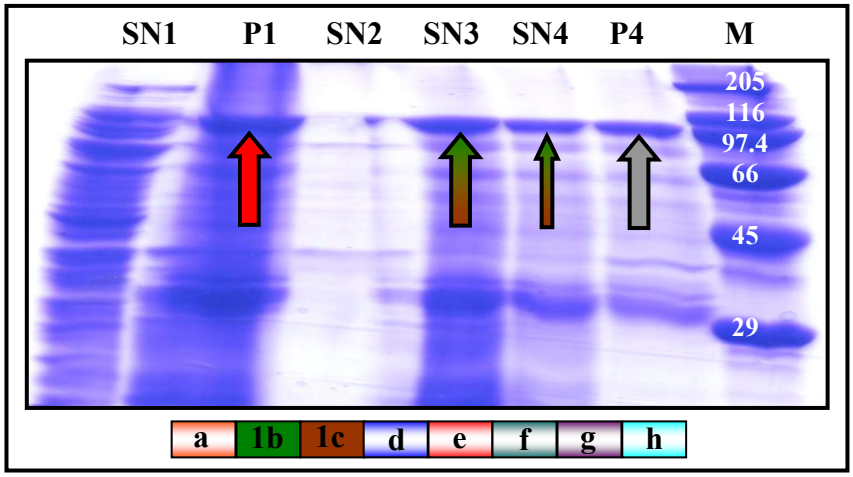


Fig. 53: SDS-PAGE of KLH1-*bc*, isolated from inclusion bodies using protocol 1. Aliquots of the supernatants (SN) and the pellets (P) were applied and stained with Coomassie. While the recombinant protein is first still located in pellet no 1 (P1, red arrow), the incubation with 8 M urea leads to solubilisation of the protein (SN3, thick green-brown arrow). Further incubation in 8 M urea shows that still some material dissolves (SN4, thin green-brown arrow) from the recombinant protein, still present within the pellet (P4, grey arrow). M represents the marker, in kDa.

While the expected protein band of 100 kDa in Figure 53, representing KLH1-*bc*, is still present in the pellet after the first isolation step (P1, red arrow), along with other prokaryotic proteins, the second supernatant (SN2) contains almost no proteins. As it can be seen in Figure 53, the step of protocol 1 producing SN3 (addition of 20 ml 8 M urea with 1 mM DTT to the pellet containing the inclusion bodies) solubilises the recombinant protein. This can be identified by the clear appearing band at approximately 100 kDa which represents KLH1-*bc* (SN3, thick green-brown arrow). The addition of 8 M urea with 1 mM DTT followed by centrifugation was repeated until no further pellet dissolved (SN4, thinner green-brown arrow) but some protein still remains in the pellet (P4, grey arrow).

The isolation using protocol 2 showed no solubilisation of the inclusion bodies during the first four steps (SN1 - SN4 and P1 - P4, data not shown). Then, supernatant 4 was split and centrifuged, and one part was resuspended in 4 M urea (P5) and the other one in 8 M urea (P6). After an incubation of one hour at room temperature, the dissolved pellets were centrifuged, producing SN7 and SN8, respectively (Fig. 54).

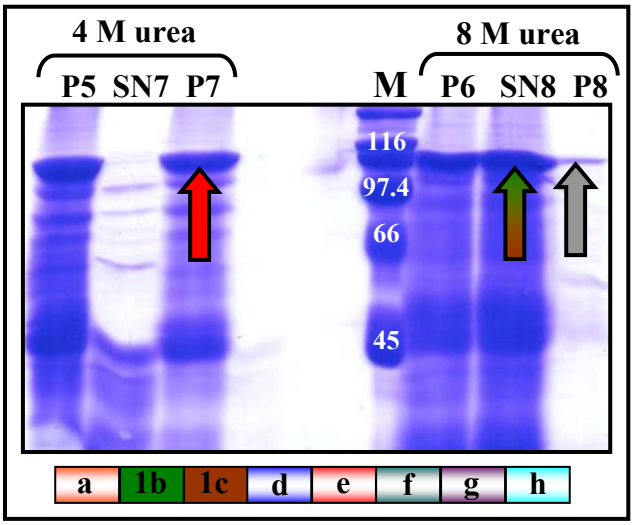


Fig. 54: SDS-PAE of aliquots of the final supernatants and pellets of the isolation of KLH1-*bc*, using protocol 2. Only the treatment with 8 M urea dissolves the recombinant protein from the inclusion bodies (SN8, thick green-brown arrow). By treatment with 4 M urea, the protein remains within the insoluble fraction P7 (red arrow). M represents the marker, in kDa.

The SDS-PAGE in Figure 54 shows that only the incubation with 8 M urea solubilises the inclusion bodies, so that the recombinant protein of 100 kDa is now soluble in the supernatant (SN8, thick green-brown arrow). The pellet P8 shows that some protein still remains undissolved (grey arrow). The incubation with 4 M urea was insufficient: the recombinant protein remains in the pellet (P7, red arrow).

In summary, both protocols worked well, if 8 M urea is applied. Because the first protocol is shorter than the second one, while giving the same satisfying results, it is used for all further purification steps.

18.2 Isolation of recombinantly expressed proteins from inclusion bodies using protocol 1, and immunological analysis: KLH2-c and KLH1-h

Knowing now that the inclusion bodies can be cracked and that the recombinantly expressed proteins can be dissolved, protocol 1 was applied for the isolation of KLH2-*c* and KLH1-*h*. The expressing cultures were also up-scaled to 50 ml and the isolation was done following the given instructions. The supernatants and the pellets arising were tested on an SDS-PAGE to check the quality of the isolations.

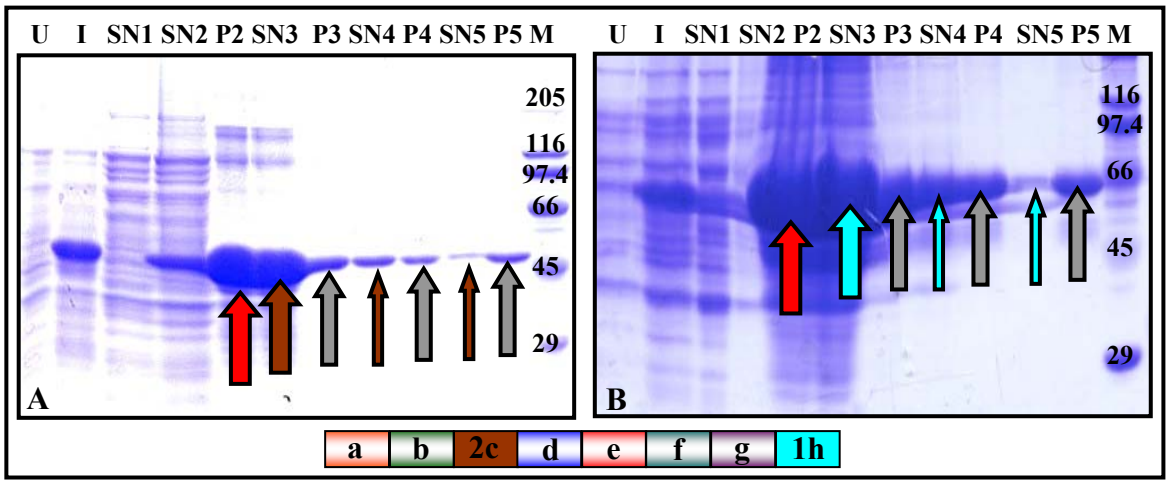


Fig. 55: SDS-PAGE of the isolation of proteins (**A:** KLH2-*c*; **B:** KLH1-*h*) from inclusion bodies using protocol 1. Aliquots of the supernatants (SN) and the pellets (P) of each purification step were applied and stained with Coomassie. The recombinant proteins are still located within the pellets P2 (red arrows), but the addition of 8 M urea leads to solubilisation (SN3, thick brown and turquoise arrows, respectively). Each further incubation step dissolves more protein (SN4 and SN5, thin brown and turquoise arrows), while some still remains within the pellets (P4 and P5, grey arrows). M represents the marker, in kDa.

The Coomassie staining of the isolation of the recombinant proteins from inclusion bodies shows that again the incubation of the pellet with 8 M urea (SN3) results in a thick soluble protein band of approximately 52 kDa for KLH2-*c* (Fig. 55A, thick brown arrow) and 62 kDa for KLH1-*h* (Fig. 55B, thick turquoise arrow), which was previously present in the pellet (P2, red arrows). The thickness of the bands is a good indication to what extent the recombinant protein was over-expressed in the bacterial cells, which leads to the formation of the inclusion bodies. The following pellets show that there is still some insoluble protein remaining (P3, P4 and P5, grey arrows), which is further dissolved during the following repeated steps (SN4 and SN5, small brown and turquoise arrows, respectively).

In parallel, a Western blot was done with the samples of both isolations (KLH2-*c* and KLH1-*h*) and incubated with a monoclonal α-His-antibody (Fig. 56).

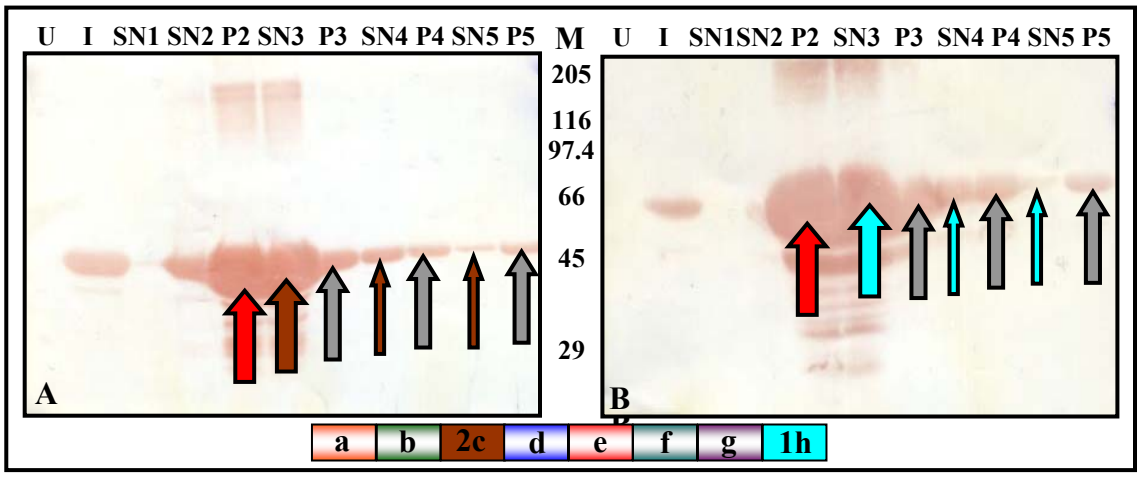


Fig. 56: Western blot analysis of the isolations of recombinantly expressed KLH2-*c* (**A**) and KLH1-*h* (**B**) from inclusion bodies both incubated with a monoclonal α-His-antibody. All bands representing the recombinant proteins are detected by the antibody. M represents the marker, in kDa.

The incubation of the proteins isolated from inclusion bodies with the α-His-antibody shows each that the appropriate protein bands in SN3 are His-tagged proteins and represent the FUs KLH2-*c* (A) and KLH1-*h* (B), respectively. The large brown and turquoise arrows in Fig. 56A and B, show the recombinant proteins, now soluble in SN3. Some additional protein still remains in the pellets (P3, P4 and P5, grey arrows, respectively), while some proteins dissolves in the following supernatants (SN4 and SN5, small brown and turquoise arrows).

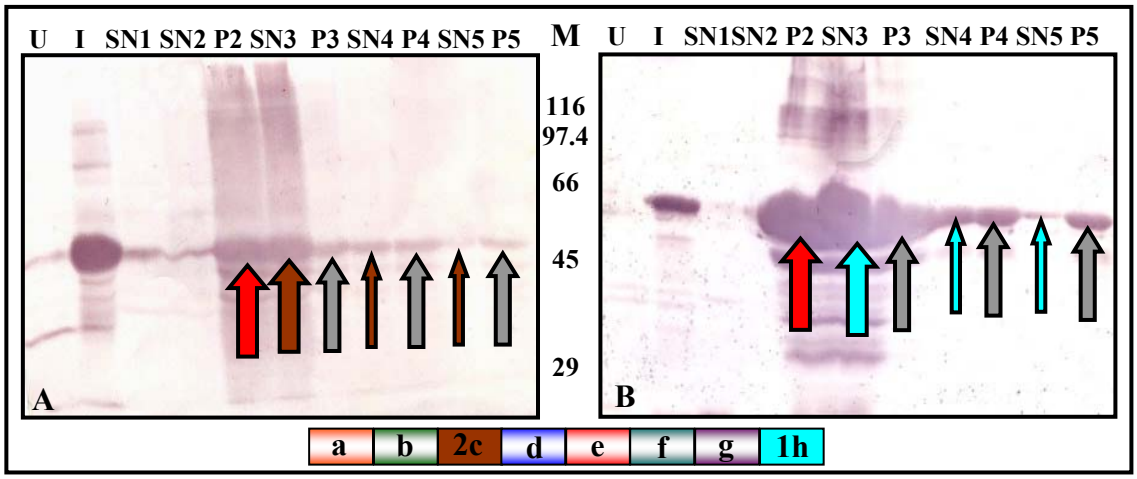


Fig. 57: Western blot analysis of the isolation of recombinantly expressed KLH2-*c*, incubated with polyclonal α-KLH2*c*-antibodies (**A**), and KLH1-*h* incubated with polyclonal α-KLH1-antibodies (**B**). All bands representing recombinant proteins are detected by the respective antibody. M represents the marker, in kDa.

The detection using on one hand polyclonal α-KLH2*c*-antibodies (Fig. 57A, KLH2-*c*) and on the other α-KLH1-antibodies (Fig. 57B, KLH1-*h*) also shows that both recombinantly expressed FUs are well recognized during Western blot analysis (soluble frac-

tions, SN3, SN4 and SN5, brown and turquoise arrows; insoluble protein, P3, P4 and P5, grey arrows, respectively).

19 Repeated purification of recombinantly expressed KLH substructures after the isolation from inclusion bodies

19.1 Standard conditions

Now that the recombinantly expressed proteins were obtained from the inclusion bodies, the purification using affinity chromatography was repeated. The resulting SN3 of KLH2-*c* was taken, where the most isolated protein was found (Fig. 55A). The supernatant was dialyzed for 2 days against 2 litres of a 50 mM Tris-HCl buffer solution at pH 8.5 using a < 10,000 Da molecular mass cut off dialysis tube, with repeated exchange of the buffer, to remove the 8 M urea from the protein solution. The Ni-NTA spin columns (Qiagen, Hilden) were additionally equilibrated with the same buffer to optimize the binding conditions. Aliquots of the supernatant (SN), the flow-through (FT), both wash (W1 and W2) and elution steps (E1 and E2) were applied on an SDS-PAGE and stained with Coomassie (Fig. 58).

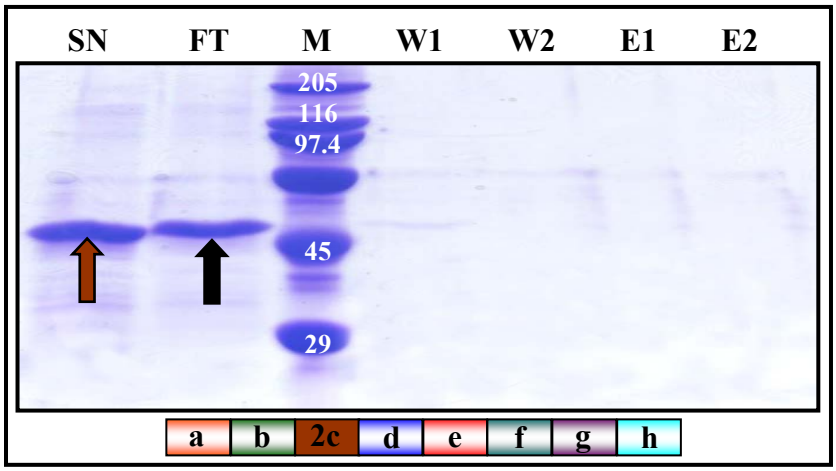


Fig. 58: Aliquots of the supernatant (SN), the flow-through (FT), both wash (W1 and W2) and elution steps (E1 and E2) of the purification of recombinantly expressed KLH2-*c*, isolated from inclusion bodies and dialysed against 50 mM Tris buffer, pH 8.5. While the recombinant protein is first located within the applied supernatant (SN, brown arrow), it is found in the flow-through after passing the column (black arrow). M represents the marker, in kDa.

The recombinantly expressed protein KLH2-*c*, first seen among other bacterial proteins in the supernatant (Fig. 58, SN, brown arrow), is found, after passing the Ni-NTA column, within the flow-through (FT, black arrow). The washing steps as well as the elution steps reveal no specific protein bands.

19.2 Applying different conditions for a better binding of the proteins to the column

Regarding the previous result, the pH of the dialysis buffer (pH 8.5), which possibly did not fit the recommended maximum pH range of the columns (pH 8.4 should not be exceeded) was lowered to 8.0. The same amount of protein as previously used was applied to the column and then washed and eluted. In parallel, the salt concentration of the protein sample was adjusted to the one contained within the recommended lysis, wash and elution buffers of the native purification using the Qiagen spin columns, namely 50 mM NaH_2PO_4 and 300 mM NaCl. Again the resulting protein fraction was applied onto a column and washed and eluted. The results were visualized by SDS-PAGE. Although the pH had slightly been adjusted (8.0 instead of 8.5) within the first experiment, the recombinant protein was still found within the flow-through (data not shown).

The second attempt with the adjusted salt concentrations showed that while the same amount of recombinant protein was clearly present in the applied supernatant, it was neither found in the flow-through, nor in any of the wash or elution steps (data not shown). Now it seemed that the recombinant protein may still be bound onto the column. Different approaches were tried to elute the protein: first with a lowered salt concentration (1:10 dilution of the buffers), followed by an augmented imidazole concentration (500 mM), which should compete with the 6xHis-tag for the Ni-NTA binding sites. A consecutive decrease in pH by application of 50 mM Tris, pH 6.8, was thought to remove the proteins from the column, followed by a final addition of 50 mM Tris, pH 3.2, to be sure to elute all remaining proteins. All different eluates were checked by SDS-PAGE for the recombinant protein KLH2-*c*, but all lanes remained empty (data not shown). Thus, it was not possible to elute the protein from the column.

The same experiment was performed with the same amount of purified proteins of KLH1-*bc*, but before the dialysis step, so that the proteins remained denatured, to ensure the fully exposed 6xHis-tag. This time 100 mM of NaH_2PO_4 and 10 mM Tris were added and an adjustment of the pH to 8.0 was done (conditions of the denaturing lysis buffer). While one approach was done using the just mentioned conditions, a second was done with the addition of 20 mM NaCl (Fig. 59).

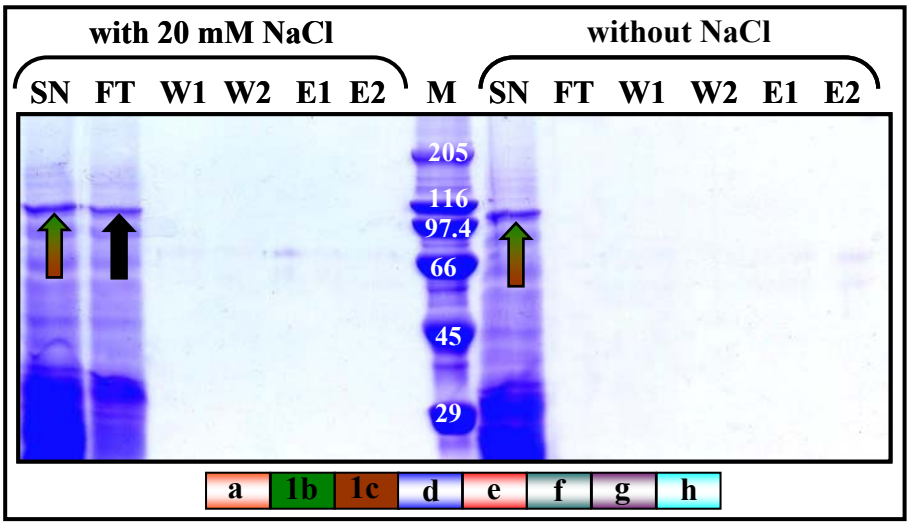


Fig. 59: SDS-PAGE of the purification experiment of KLH1-*bc* using adapted denaturing conditions, once with 20 mM NaCl and once without. The buffer containing NaCl results therein that the protein is found in the flow-through (FT, black arrow), the one without NaCl leads to no elution at all.

So, independently which conditions were chosen to purify the recombinantly expressed proteins by affinity chromatography, either the proteins did not bind to the matrix or it was not possible to elute them any more.

20 Biochemical analysis of native and recombinantly expressed KLH1-*h*: immunological characterization and absorption spectra

20.1 Native KLH1-h

The native equivalent of the FU KLH1-*h* was kindly isolated and donated by [redacted] A mixture of KLH1 and KLH2 was provided by Biosyn, Fellbach, and dialysed against 2 % (v/v) ammonium-molybdate buffer, pH 5.7, and centrifuged for 4 hours at 29,000 x g. Isoform 1 is now present in the pellet, while dissociated KLH2 remains in the supernatant. The pellet is dissolved in stabilizing buffer, pH 7.4, and dialysed against 0.13 M glycine buffer, pH 9.6, which dissociates the KLH1 didecamers into single subunits. After enzymatical cleavage by repeated addition of 2 % V8 protease for 4 hours at 37 °C, the resulting mixture of FUs was applied on an anion-exchanger Q-Sepharose. The collected fractions of the first appearing peak were analysed by SDS-PAGE, crossed-line and two-dimensional immunoelectrophoresis (data not shown). Fraction no. 50 was identified to contain the most and purest amount of KLH1-*h*. Therefore it was chosen for all further analyses (Fig. 60).

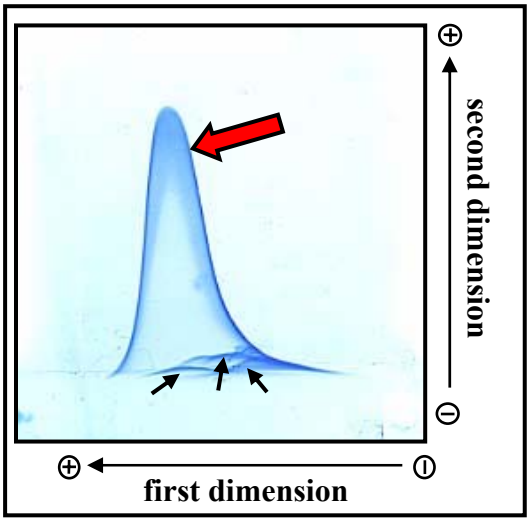


Fig. 60: Two-dimensional immunoelectrophoresis of fraction no 50. A distinct peak of precipitated protein is shown which represents the native KLH1-*h* (red arrow), together with several impurities (black arrows).

The major peak of precipitated protein in the two-dimensional immunoelectrophoresis of fraction 50 (Fig. 60) represents the native KLH1-*h* (red arrow), together with several impurities (black arrows). With this fraction an absorption spectrum was produced from 220 nm to 600 nm, also kindly done by [redacted]. For native hemocyanin, two characteristic absorption maxima occur, at 280 nm and 340 nm, respectively.

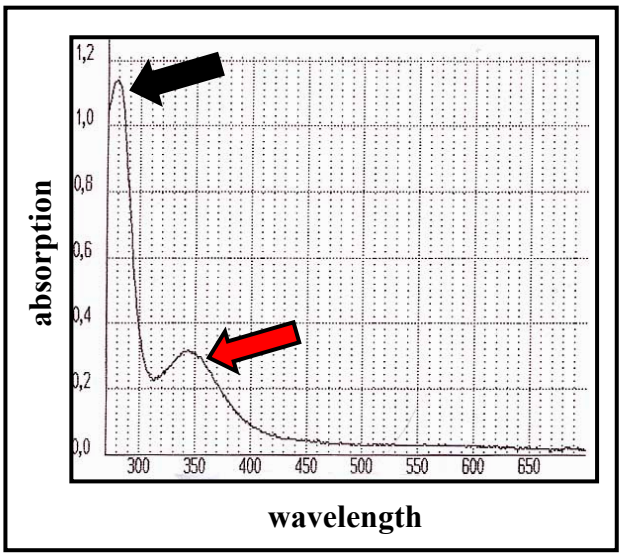


Fig. 61: Absorption spectrum of native KLH1-*h*, produced from 220 nm to 600 nm. Two characteristic absorption maxima are shown, one at 280 nm (black arrow) and one at 340 nm (red arrow).

The first peak in Fig. 61 (black arrow) is typical for proteins, due to the absorption of aromatic amino acids such as tryptophane and tyrosine. The oxygenated active site of the hemocyanin (Cu-O_2-Cu-complex) has an absorption maximum at 340 nm (second peak) and therefore, the UV spektrum provides evidence for the quantitative presence of

hemocyanin, when a maximum occurs at this wavelength (red arrow). It originates from the binding of oxygen to the active site of hemocyanin.

20.2 Recombinantly expressed KLH1-h from E. coli

Although the most recombinant protein appeared within the third supernatant (SN3) of the purification using protocol 1 (refer to: 18 *Isolation of proteins from inclusion bodies*) of the KLH1-*h* expressing culture, SN5 was taken for further analyses because of the fewer contaminations. The supernatant was dialyzed for 2 days in 2 litres of a 50 mM Tris-HCl buffer solution at pH 8.5 using a < 10,000 Da molecular weight cut off dialysis tube, with repeated exchange of the buffer, to remove the 8 M urea from the protein solution. 1, 2, 4 and 6 µl, respectively, of the dialysed protein solution were taken for one-dimensional immunoelectrophoresis (so called "rockets") against the polyclonal α-KLH1-antibody and stained with Coomassie (Fig. 62).

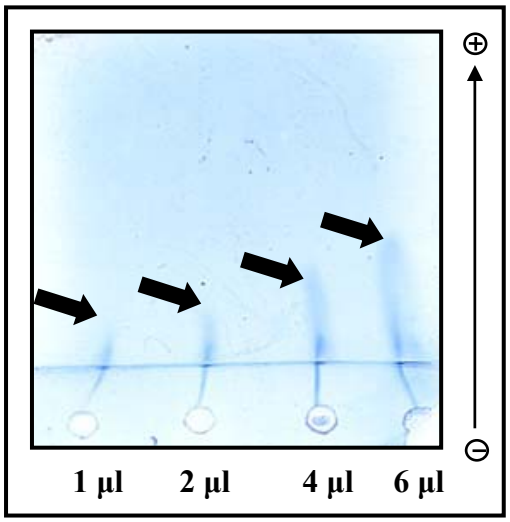


Fig. 62: One-dimensional immunoelectrophoresis of the dialysed SN5 of the recombinantly expressed and purified KLH1-*h* against the polyclonal α-KLH1-antibody. Each different applied amount shows only a blurred precipitation smear against the antibody.

The rocket in Figure 62 shows no typical peak of precipitated protein within the α-KLH1-antibody-containing agarose. Only smear-like formations are observed, with increasing height dependent of the applied amount of recombinant protein (black arrows).

Therefore an absorption spectrum was produced (Fig. 63), like previously done for native KLH1-*h*.

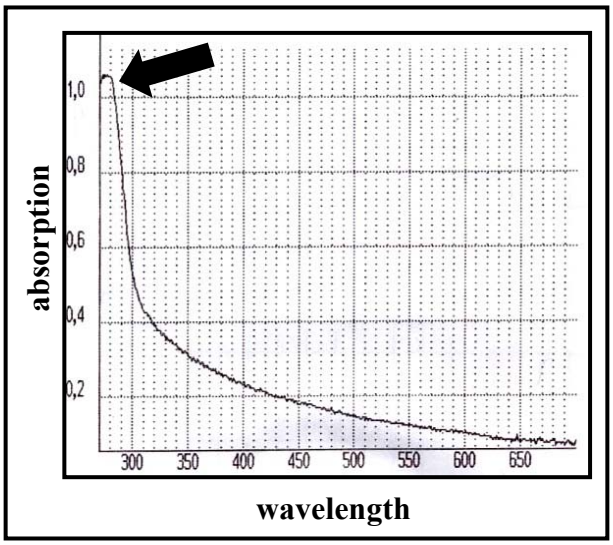


Fig. 63: Absorption spectrum of dialysed SN5 of recombinantly expressed KLH1-*h*. Only the peak at 280 nm is present. The second, characteristic for hemocyanin, at 340 nm, is missing.

The absorption spectrum of the dialysed SN5 of the recombinantly expressed KLH1-*h*, produced from 220 nm to 600 nm, shows only the first peak at 280 nm, which is typical for proteins (Fig. 63, black arrow). However, the expected second peak, at 340 nm, which is characteristic for active hemocyanin, is not present. The dialysed recombinant KLH1-*h* was previously treated with denaturing 8 M urea, to dissolve the *E. coli* inclusion bodies. Regarding the results above it seems that the protein had not refolded correctly during dialysis.

21 Application of the same conditions used for recombinantly expressed KLH1-*h* to its native counterpart: a comparison

21.1 Denaturation of native KLH1-h by treatment with 8 M urea

To be able to explain the observed behaviour of the recombinant protein during precipitation against the α-KLH1-antibody, a sample of the native KLH1-*h* from fraction 50 was taken and treated with 8 M urea. After an overnight incubation, followed by dialysis against 50 mM Tris-HCl, immunorockets were produced by [redacted], to compare the native and the denatured KLH1-*h*, again using the α-KLH1-antibody (Fig. 64).

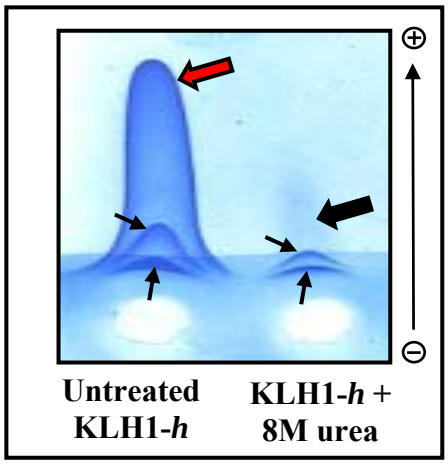


Fig. 64: One-dimensional immunoelectrophoresis of native and 8 M urea-treated KLH1-*h*. While a distinct peak of precipitated protein is observed for the untreated sample (left side, red arrow), together with several impurities (small black arrows), the denaturation with 8 M urea leads to complete disappearance of the major peak (big black arrow, right side).

The denaturing conditions during the overnight incubation of the native KLH1-*h* in 8 M urea, shown by the rocket (Fig. 64), indicate that the protein is no longer recognized by the α-KLH1-antibody. The major peak, seen for untreated KLH1-*h* (red arrow), together with some impurities (small black arrows) disappears and a faint smear-like formation occurs within the antibody containing agarose (big black arrow).

21.2 Carbohydrate digestion of native KLH1-h

It is known that KLH is a highly glycosylated protein which plays an important role in antibody recognition. Prokaryotic bacterial cells lack the possibility of post-translational modification, so that the recombinantly expressed protein has no carbohydrate side chains. To check if the weak recognition by the antibody is due to these circumstances, native KLH1-*h* was deglycosylated and subsequently tested with the antibody.

PNGase F is a glycopeptidase, which cleaves asparagine-linked high mannose as well as hybrid and complex oligosaccharides from glycoproteins. Following the protocol of PNGase F (Sigma, Deisenhofen), three consecutive steps are applied preceding the de-glycosylation, to maximize the recovery. The first one is the addition of 10 µl phosphate buffer, pH 7.5, to up to 200 µg glycoprotein, followed by the incubation with PNGase F overnight at 37 °C. The second step includes the first and the addition of 2.5 µl of 2 % SDS with 1 M β-mercaptoethanol, before the incubation at 37 °C. The third step includes the two previously mentioned followed by an addition of 2.5 µl 15 % Triton-X 100 and by a denaturation step for 5 minutes at 95 °C, which should lead to the highest recovery of deglycosylated protein. Then the samples of each step (Fig. 65, dot no. 2 to dot no. 4) are applied drop wise onto a nitrocellulose membrane beside the negative control pro-

tein, creatinase, which is not glycosylated, to determine the background. As a positive control, untreated KLH1-*h* is also applied to the membrane.

After a reversible Ponceau S staining of the proteins (Fig. 65A), the results were visualized by testing the applied proteins with the DIG glycan detection Kit (Roche, Mannheim), which is based on an enzyme immunoassay. This detection of carbohydrates was done following the manufacturer's protocol (Fig. 65B).

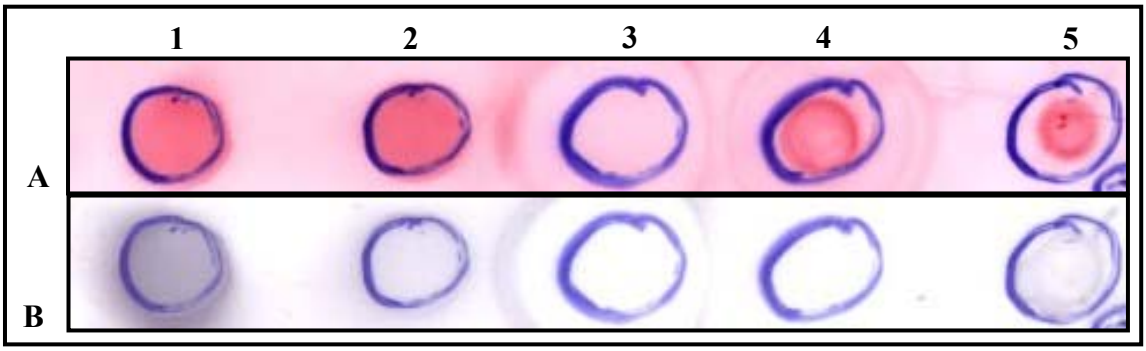


Fig. 65: Deglycosylation of native KLH1-*h* following different treatments (dot no. 2: + 10 µl phosphate buffer, pH 7.5; dot no. 3: + 2.5 µl of 2 % SDS with 1 M β-mercaptoethanol; dot no. 4: + 2.5 µl 15 % Triton-X 100 followed by 5 minutes at 95 °C), each with a subsequent incubation with 2 µl PNGase F overnight at 37 °C. Dot no. 1 represents the untreated KLH1-*h*; dot no. 5 the negative control protein creatinase, which is not glycosylated. **A**: Ponceau S staining. **B:** Specific carbohydrate detection using the DIG glycan detection Kit (Roche, Mannheim).

The nitrocellulose membrane in Figure 65 shows, at the left side, the untreated native KLH1-*h*, which is applied as a positive control (dot no. 1, A and B). The FU contains carbohydrates, apparent due to the clear blue central precipitation product of the conjugated alkaline phosphatase. The dot marked as no. 2 (A and B) represents the glycoprotein treated only with phosphate buffer prior to the incubation with PNGase F overnight at 37 °C. The Ponceau S staining shows for both dots (Fig. 65A, no. 1 and no. 2) that the same amounts of proteins are applied to the nitrocellulose membrane. The immunoassay reveals that, compared to dot no. 1, only a very weak blue coloration is visible in dot no. 2, which leads to the conclusion that most of the applied KLH1-*h* on dot 2 should be deglycosylated. The protein applied on dot no 3 was additionally treated with SDS and β-mercaptoethanol before the incubation with the glycopeptidase, but the Ponceau S staining shows that no protein has bound to the membrane (Fig. 65A), so detection through the alkaline phosphatase also failed (B). In contrast, the Triton-X 100 treated and heat-denatured KLH1-*h*, which is represented by dot no. 4, is clearly visible in the Ponceau S staining (A), but no precipitation occurs during the detection with the DIG glycan detection Kit (B). This treatment prior to the incubation with PNGase F produces the best results: the removal of the carbohydrate side chains seems to be completed. The

negative control protein creatinase (dot no. 5) which is well represented by a red protein dot within the Ponceau S staining (Fig. 65A) shows no detection in part B, as expected.

22 Analysis of comparably treated native and *E. coli* derived FU-1*h*

Because KLH1-*h* has to be kept in its native conformation for the subsequent experiments, all additions were avoided, although this led to the best recovery of deglycosylation. Therefore the incubation at 37 °C was extended to the maximum (72 hours), to ensure the best deglycosylation of the protein by PNGase F.

22.1 SDS-PAGE and immunological analysis

To visualize on one hand the impact of the deglycosylation and on the other the consequence of the denaturation of the protein, native KLH1-*h* (Fig. 66A and B, lane 1), deglycosylated (lane 2) as well as denatured one (lane 3) were applied to an SDS-PAGE and stained with Coomassie (Fig. 66A). As a comparison, the recombinantly expressed KLH1-*h* from *E. coli* was also applied (lane 6). To check if there were immunological differences between the native and the recombinantly expressed protein, the samples were applied a second time to an SDS-PAGE and the electrophoresed proteins were blotted and incubated with an α-KLH1-antibody (Fig 66B).

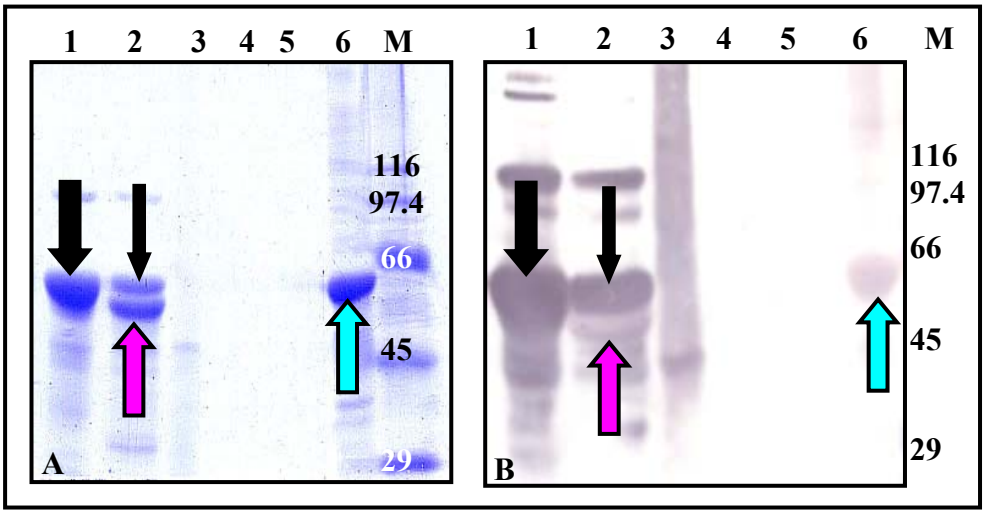


Fig. 66: Different KLH1-*h* samples: the native FU (lanes 1, thick black arrow), the deglycosylated one (lanes 2, black and pink arrow), denatured (lanes 3). Lanes 6: *E. coli* derived KLH1-*h* (turquoise arrow). **A**: Coomassie staining. **B**: Detection using α-KLH1-antibodies. M represents the marker, in kDa.

As can be seen from the Coomassie staining in Figure 66A, the native KLH1-*h* in lane 1 is shown by one strong protein band (thick black arrow). The second lane, containing the deglycosylated sample of KLH1-*h* shows two distinct bands, one at the same migration level as the untreated KLH1-*h* in lane 1 (thin black arrow), and a second protein band

directly beneath it (theoretical calculated molecular weight: 57 kDa, pink arrow) which is quantitatively greater and represents the deglycosylated protein.. The denatured KLH1-*h* in lane 3 is not visible. The recombinantly expressed one, however, is shown by a thick protein band at the expected 62.5 kDa (lane 6, turquoise arrow). The detection of the four just mentioned different samples of KLH1-*h* by the polyclonal α-KLH1-antibody produced an interesting result (Fig. 66B). The protein band in lane 1 (thick black arrow) as well as the upper band in lane 2 (thin black arrow) were strongly recognized by the antibody, apparent by the precipitation product of the alkaline phosphatase. The lower protein band in lane 2 (pink arrow) was bound by the antibody to a lesser extent, although the Coomassie staining (Fig. 66A) shows that this band is quantitatively greater than the upper one. Lane 3 reveals only a smear of degradation products. The recombinant KLH1-*h* shows, in lane 6, a pronounced protein band at 62.5 kDa (turquoise arrow), but recognition by the antibody is weak, although a comparable amount of protein had been applied, as for its native counterpart in lane 1.

22.2 Native PAGE and immunological analysis

During native PAGE, untreated proteins are separated electrophoretically due to their molecular mass combined with their net charge and conformation (tertiary structure). To determine the migration pattern of these protein samples within a native PAGE, regarding at one hand the glycosylation status and on the other the denaturation status of the proteins, native KLH1-*h* (Fig. 67A and B, lane 1), the deglycosylated (lane 2) as well as the denatured FU (lane 3) were applied to a minigel, under native conditions. The migration behaviour of the prokaryotic KLH1-*h* was also analyzed in this PAGE (Fig. 67A and B, lane 6). To determine the binding ability of the polyclonal α-KLH1-antibody to the different KLH1-*h* samples within a native PAGE, the proteins were applied to an additional gel and blotted onto a nitrocellulose membrane (Fig. 67B).

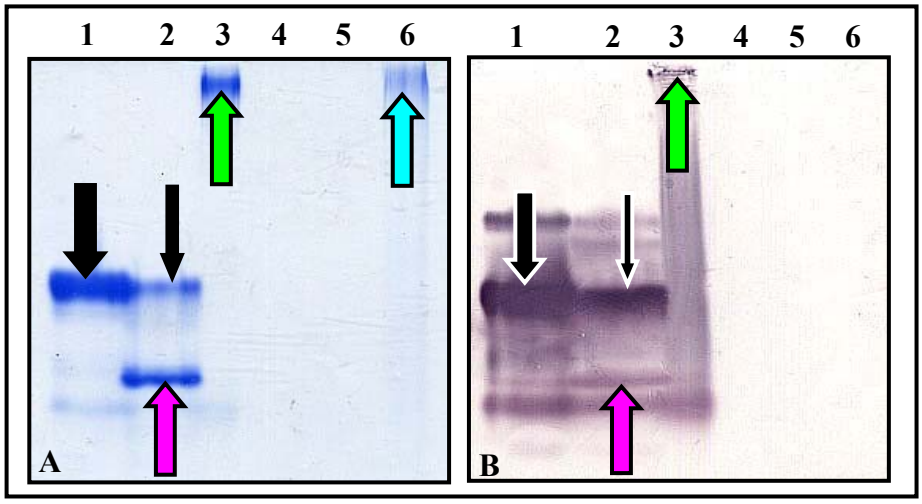


Fig. 67: Native PAGEs of untreated (lane 1, black arrow), deglycosylated (lane 2, black and pink arrow), denatured (lane 3, green arrow), and recombinantly expressed (lane 6, turquoise arrow) KLH1-*h*. **A:** Coomassie staining. **B**: Specific detection using an α-KLH1-antibody.

The Coomassie staining of the native PAGE (Fig. 67A) shows a similar result to the previous denaturing SDS-PAGE (Fig. 66A): the native KLH1-*h* (lane 1) has only one distinct protein band (thick black arrow), while the deglycosylated fraction applied in lane 2 shows two. The upper one (thin black arrow) has also in the native PAGE the same migration pattern than the protein band of the untreated sample, while the second protein band (pink arrow), representing the deglycosylated protein, is positioned anodal to the first mentioned band. The denatured fraction applied in lane 3 produces a smear at the upper edge of the gel (green arrow), which is also observed for the recombinantly expressed KLH1-*h*, in lane 6 (turquoise arrow).

The Western blot analysis of the native PAGE shows that the native, untreated FU-1*h* is strongly recognized by the α-KLH1-antibody (Fig. 67B, lane 1, thick black arrow), as well as smaller degradation products. The deglycosylated sample of KLH1-*h* which is applied in lane 2, shows the same result as the antibody incubation of the SDS-PAGE (Fig. 67B): while two protein bands appear, the upper one (thin black arrow) is quantitatively less than the lower band (pink arrow). However, the upper band is more strongly recognized by the polyclonal α-KLH1-antibody than the anodic protein band. The denatured FU which is applied in lane 3 and represented by a small protein band at the top of the gel, as seen in the Coomassie staining (Fig. 67A), is weakly bound and detected by the antibody (B, green arrow). Lane 6 contains the recombinantly expressed KLH1-*h* which gives no reaction with the antibody (Fig. 67B).

23 Investigations on native and recombinantly expressed KLH1-*h*: application of the same conditions to insect cell- and *E. coli*-derived FU-1*h*

23.1 Expression experiment

Eukaryotic organisms have a more complex protein expression machinery than prokaryotic cells and they do introduce post-translational modifications such as glycosylation. The recombinant expression of KLH in a eukaryotic system, like insect cells, should give new insights and possibilities of comparison.

All recombinant expression experiments of KLH substructures in insect cells were kindly done by [redacted], using the Bac-to-Bac protocol for baculovirus expression (Invitrogen, Karlsruhe). The entry clone of the FU KLH1-*h*, generated during the present work, was taken and the cDNA was first subcloned into the insect cell-specific destination vector pDEST™10, which contains an N-terminal 6xHis-tag. During transformation of this clone into the bacterial strain DH10Bac, which exhibits bacmids containing the viral genome of the baculo virus, the insert is integrated into the bacmid. The isolation of this construct is done during a bacmid preparation. The transfection of Sf9 insect cells occurs with the help of an artificial polymer, termed Nanofectin (PAA), which allows the passage of the DNA through the membrane into the cells. After 4 days, the supernatant is harvested, containing the first virus. This one is amplified at least two times before virus infected cells can be harvested for protein recovery.

After the first expression experiments it was seen that the insect cells do express the recombinant KLH1-*h*. During the following purification steps, however, it was found that the recombinant protein was located as insoluble aggregates within the cells, which always pelleted within the cellular debris after cell lysis.

To enhance the protein solubility, a new construct of KLH1-*h* was generated by PCR, starting at the direct beginning of the FU, without overlapping sequence to FU-1*g*, but having at the 5' end the KLH specific signal peptide and a 6xHis-tag at the 3' end. After generation of an entry clone, this cDNA was subcloned into the destination vector pDEST™8, having neither N- nor C-terminal tag. The recombinant expression in Sf9 insect cells took place in the same way, as described above. The monitoring of the expression by SDS-PAGE analysis showed that some soluble recombinant protein was now found in the supernatant of the insect cell culture, but the larger part was still found within the cell pellet.

At this time point, samples of the pellet and the supernatant of 50 ml KLH1-*h*-expressing insect cell culture were kindly made available to me for the following analyses. While the pellet was resuspended in 5 ml PBS, pH7.5, and stored at 4 °C, the proteins of the supernatant were precipitated overnight in 20 % ammonium sulphate. Concentration of the sample was obtained by a centrifugation step for 20 minutes at 6,000 x g at 4 °C, followed by resuspension of the pelleted proteins in 500 µl PBS, pH 7.5.

23.2 Analysis of the migration pattern by SDS-PAGE and immunological analysis

To be able to compare the insect cell-derived FUs representing KLH1-*h* with its native counterpart, aliquots of the pellet and the supernatant were taken and treated in the same way. While one sample of each was incubated with PNGase F for 72 hours at 37 °C, a second aliquot was treated with 8 M urea during the same time period. The first insect cell-derived KLH1-*h* recombinantly expressed without signal peptide was taken as a comparison, as well as its denatured fraction. The native counterpart, its deglycosylated and its denatured fraction as well as the *E.coli*-derived KLH1-*h* were also taken into account. The results were visualized by SDS-PAGE (Fig. 68).

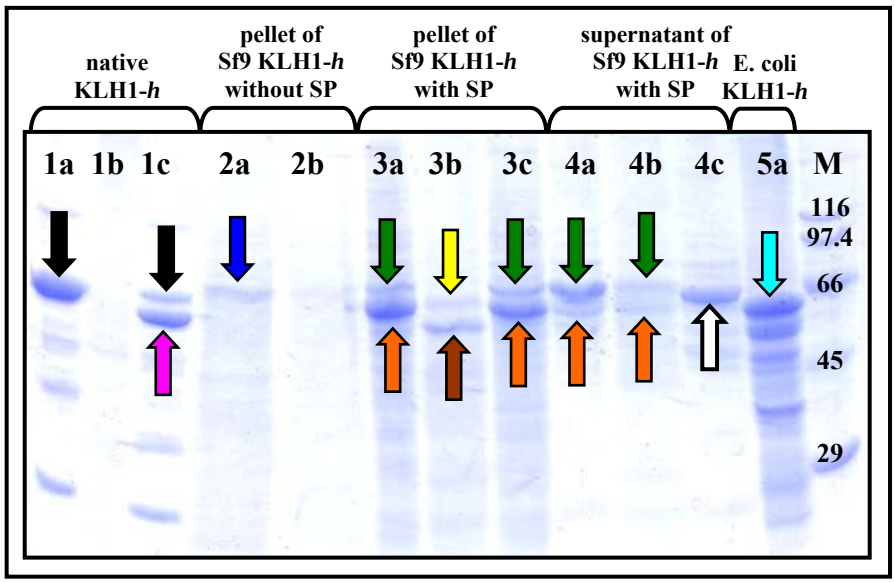


Fig. 68: SDS-PAGE of different KLH1-*h* samples: the native FU, the insect cell-derived FU (Sf9) without signal peptide (SP), the insect cell-derived FU with signal peptide (pellet and supernatant) and KLH1-*h* from *E. coli*. The lanes labelled "a" represent the respective untreated sample, "b" the denatured one and "c" the deglycosylated one, respectively. M represents the marker, in kDa.

The SDS-PAGE in Fig. 68 demonstrates the different occurring bands of KLH1-*h*, depending on the origin and the treatment of the samples. As already seen in Figure 66, the

native FU is represented by a thick single protein band, at the expected mass of approximately 63,5 kDa (lane 1a, black arrow). The denaturation of this FU leads to a disappearance of almost all protein bands (lane 1b), while its deglycosylation using PNGase results in an additional band, at approximately 57 kDa (lane 1c, pink arrow). At the same time, the upper band, which is also present in the untreated fraction, becomes weaker (lane 1c, black arrow). The recombinant KLH1-*h* from insect cells (Sf9), which is expressed without signal peptide and therefore located within the cellular pellet, shows in lane 2a a weak protein band (blue arrow), slightly larger than the native counterpart (64 kDa), due to the N-terminal 6xHis-tag of pDEST™10 and the coexpressed proportion of FU-1*g*. Its denaturation, seen in lane 2b, shows again that nearly all protein has disappeared, due to the treatment with 8 M urea. The pellet-forming KLH1-*h*, although expressed in insect cells with the KLH1 specific signal peptide (SP), has a similar migration pattern (lane 3a) to the native, deglycosylated fraction: while a small protein band migrates at 57 kDa (orange arrow), a thicker band was found directly beneath (green arrow). The denaturation (lane 3b) and also the deglycosylation (lane 3c, green and orange arrow) of this sample looks quite similar to the untreated one, except that the denatured protein bands are slightly reduced in mass and look weaker (lane 3b, yellow and brown arrow, respectively). The recombinantly expressed, soluble KLH1-*h*, found within the supernatant of the insect cell culture, is also present as two distinct protein bands, one at 57 kDa (lane 4a, orange arrow) and one above it (green arrow), but the last one is quantitatively greater. The denatured counterpart of this sample (lane 4b) also shows two very weak protein bands of the same mass (green and orange arrow, respectively). The deglycosylated sample (lane 4c) contains only one protein band around 63 kDa (white arrow). The recombinant prokaryotic KLH1-*h* expressed in bacterial cells applied in lane 5a is represented by a thick protein band, at 62.5 kDa (turquoise arrow), together with smaller degradation products.

The same samples were applied to an additional SDS-PAGE and transferred onto a nitrocellulose membrane. The Ponceau S staining of the different, blotted KLH1-*h* samples showed that all protein bands, previously seen in the Coomassie staining (Fig. 68), had been transferred onto the membrane (data not shown). The incubation with the α-KLH1-antibody, however, showed a different result (Fig. 69).

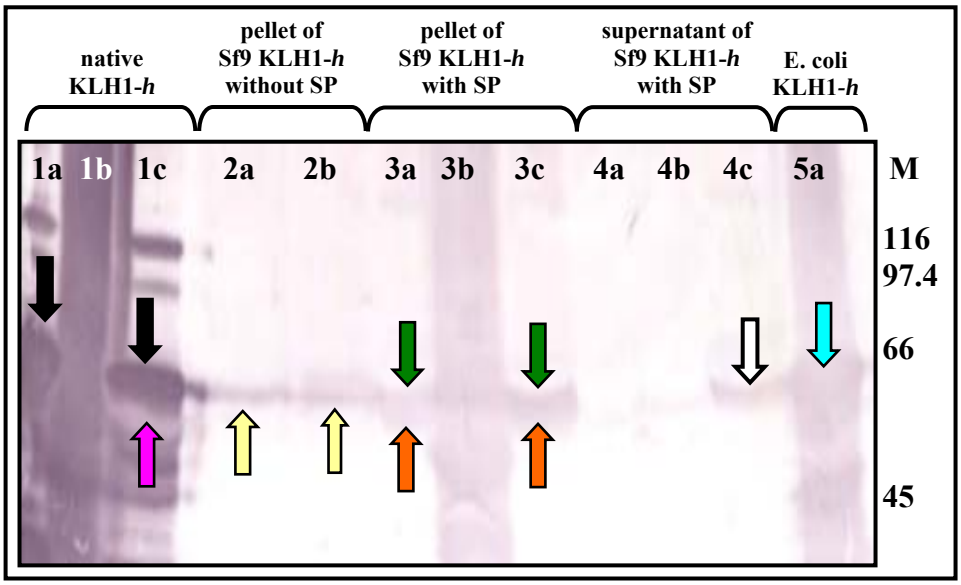


Fig. 69: Western blot analysis of the different KLH1-*h* samples, incubated with an α-KLH1-antibody: the native FU, the insect cell derived FU (Sf9) without signal peptide (SP), the insect cell-derived FU with signal peptide (pellet and supernatant) and KLH1-*h* from *E. coli*. The lanes labelled "a" represent the untreated sample, "b" the denatured and "c" the deglycosylated samples, respectively. M represents the marker, in kDa.

The native FU (Fig. 69, lane 1a), its denaturation (lane 1b) and its deglycosylation (lane 1c) shows the same results than before (Fig. 66B): The upper protein band (black arrow) is much better recognized than the lower protein at 57 kDa (pink arrow), although this one is quantitatively greater. The denatured FU shows only a smear. All other protein bands are definitely less well recognized by the antibody. KLH1-*h* from the pellet of the insect cells (Sf9), expressed without signal peptide (SP), as well as the denatured fraction (lanes 2a and 2b, respectively) show a weak recognition, but the mass of the detected bands (yellow arrows) do not correspond to the protein bands observed in the Coomassie staining (Fig. 68) and Ponceau S staining (data not shown): the Coomassie stained band marked with a blue arrow is slightly larger than the native KLH1-*h* (Fig. 68), the detected bands in Fig. 69 are not (yellow arrows). The detection of the samples applied in the lanes 3a and 3c (expressed in Sf9, with SP, pellet) has two protein bands, as seen in the previous Coomassie staining (Fig. 68, green and orange arrow), but their recognition is weak. The corresponding denatured sample (lane 3b) shows only smear. Although the Ponceau S staining (data not shown) shows that the protein bands has been transferred onto the membrane from the lanes 4a and 4b, no detection takes place. In contrast, the protein band in lane 4c, which represents the deglycosylated fraction of the signal peptide containing soluble KLH1-*h* from insect cells, is revealed by the antibody (white arrow). The recombinantly expressed KLH1-*h* from *E. coli* shows a smear along lane 5a, but a protein band at the expected 62.5 kDa seems to be labelled more inten-

sively by the antibody (turquoise arrow). Nevertheless, the recognition is again less strong than for the native counterpart (lane 1a).

23.3 Analysis of FU migration pattern using native PAGE, and immunological analysis

Again, the proteins were applied on a large, native PAGEwhich was stained with Coomassie (Fig. 70).

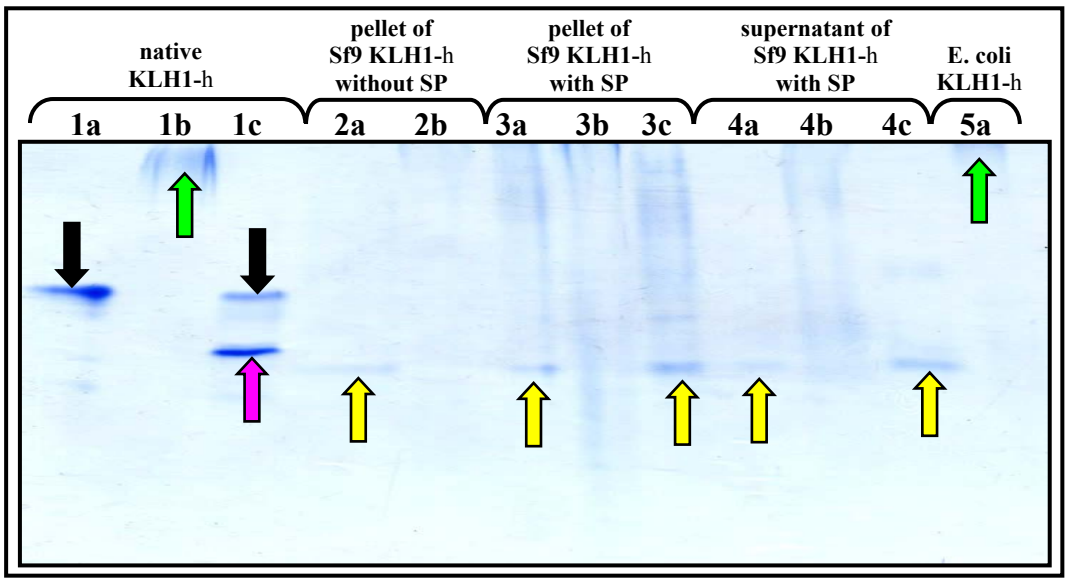


Fig. 70: Native PAGE of different KLH1-*h* samples: the native one, the insect cell-derived (Sf9) without signal peptide (SP), the insect cell-derived with signal peptide (pellet and supernatant) and the *E.coli*-derived KLH1-*h*. The lanes labelled "a" represent the untreated samples, "b" the denatured and "c" the deglycosylated samples, respectively.

The Coomassie staining of the native PAGE containing the different samples of KLH1-*h* (Fig. 70) shows the same migration pattern for the native FU, as for its denatured and deglycosylated counterpart, as already seen in Figure 67A. The native untreated fraction applied in lane 1a shows a single distinct protein band (black arrow), while the denatured one shows a smear at the upper bottom of the gel (lane 1b, green arrow). The deglycosylated sample (lane 1c) again shows two protein bands: one with the same migration pattern than the native protein (black arrow), although it was quantitatively less represented, and one additional band beneath it (pink arrow). The lanes 2a, 3a, 3c, 4a and 4c which represent all untreated and deglycosylated fractions of the eukaryotic, insert cell-derived KLH1-*h* samples all show only one weakly visible protein band, which has always the same migration pattern (yellow arrows). Within all denatured fractions of these samples (lanes 2b, 3b and 4b) only a smear was revealed by the Coomassie staining. The same

was observed for the prokaryotic, *E.coli* derived KLH1-*h* (lane 5a), but this smear was only visible at the upper bottom of the gel (green arrow).

To analyse the detection ability of the antibody towards the different KLH1-*h* samples separated by native PAGE, the proteins were again applied to a native gel and the electrophoresed proteins were transferred onto a nitrocellulose membrane. The reversible Ponceau S staining of the differently originated FUs KLH1-*h* shows that all bands had been transferred, although some are just weakly stained (data not shown). The detection, shown in Figure 71, was done using an α-KLH1-antibody.

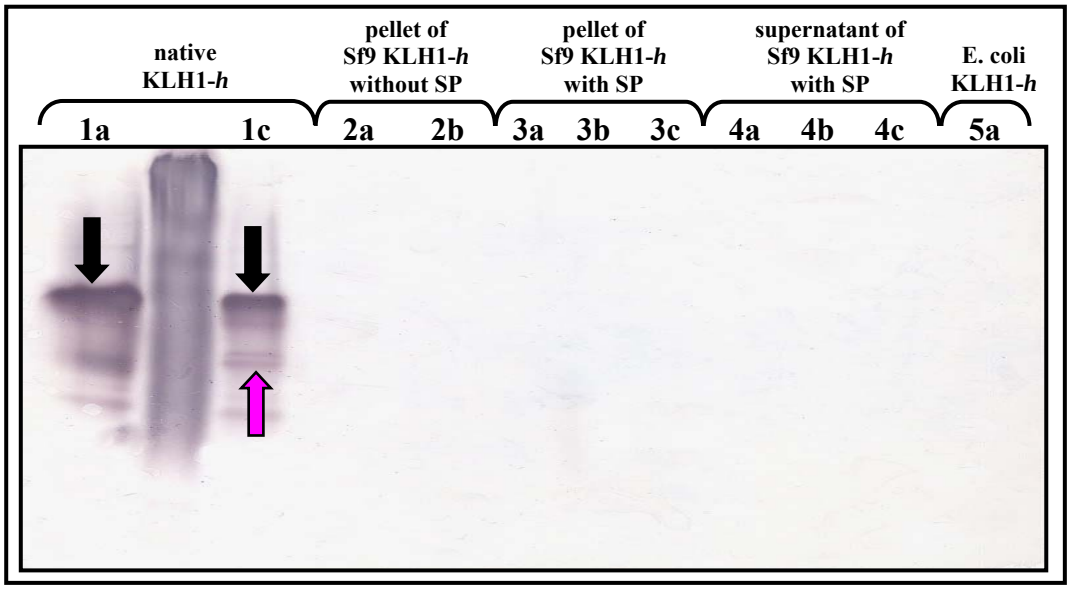


Fig. 71: Specific detection using an α-KLH1-antibody of the KLH1-*h* samples from the animal (lane 1a), the eukaryotic cells (lane 2a, without signal peptide (SP); lanes 3a and 4a expressed with SP, pellet and supernatant, respectively) and *E. coli* (lane 5a) and their denatured (lanes with "b") and deglycosylated (lanes with "c") respective counterparts.

The Western blot analysis of the above mentioned blotted proteins shows an interesting result (Fig. 71): only the native FU KLH1-*h*, extracted from the snail, as its denatured and deglycosylated fraction, react with the α-KLH1-antibody. While the FU treated with 8 M urea (lane 1b) shows only a smear of blue precipitation product along the lane, the untreated one shows a distinct protein band (lane 1a, black arrow) at the same position as in the Coomassie (Fig. 70) and Ponceau S staining (data not shown). The specific detection of the deglycosylated fraction applied in lane 1c shows the same result as the Western blot of these samples previously applied to a native PAGE (Fig. 67B). Although the upper protein band (lane 1c, black arrow) is quantitatively less represented than the lower one (lane 1c, pink arrow) it is much better recognized by the α-KLH1-antibody. In contrast, none of the recombinantly expressed proteins are detected. Any

further attempts to detect the recombinantly expressed proteins applied to native PAGEs also failed.

24 Generation of larger KLH1 substructures up to the whole subunit by SOE-PCR and cloning for future experiments in eukaryotic expression systems

Now that the method of SOE-PCR was established and had worked well for the generation of cDNAs encompassing two and three consecutive FUs of KLH1 (KLH1-*ab*, -1*cd* and KLH1-*abc*), the final goal of the present work was to fuse the existing clones together to create larger substructures, up to the whole subunit. Although it was thought that the recombinant expression in *E. coli* will not be possible due to the mass of the resulting proteins, these clones will serve for the future experiments in eukaryotic expression systems.

24.1 Splicing four FUs: KLH1-abcd and KLH1-efgh

While the splicing of the cDNAs containing two and three FUs each were done using only one approach, the generation of larger substructures, like the cDNA encoding the four FUs KLH1-*abcd*, required several trials. Because the fusion of the cDNA of KLH1-*ab*c with the one of KLH1-*d* constantly failed, the two cDNAs of KLH1-*ab* and KLH-*cd* were amplified from their respective vector, with one attachment and one gene-specific primer, respectively. Then, both were fused together by SOE-PCR (Fig. 72A). The expected mass was 5,005 bp. The cloning was done by [redacted] (data not shown).

I generated the cDNA for the remaining four FUs, namely KLH1-*efgh*, by fusing the cDNA of KLH1-*ef* together with the one coding for KLH1-*gh*. The resulting cDNA should be slightly larger than the one coding for the first four FUs (KLH1-*abcd*), due to the 3' extension of the cDNA of KLH1-*h* of approximately 300 bp and a coamplified portion of the 3' UTR (81 bp), namely 5,382 bp. The cloning was done using pDONR™221 to create the appropriate entry clone. A clone-PCR should give evidence if the right cDNA had been inserted into the vector (Fig. 72B).

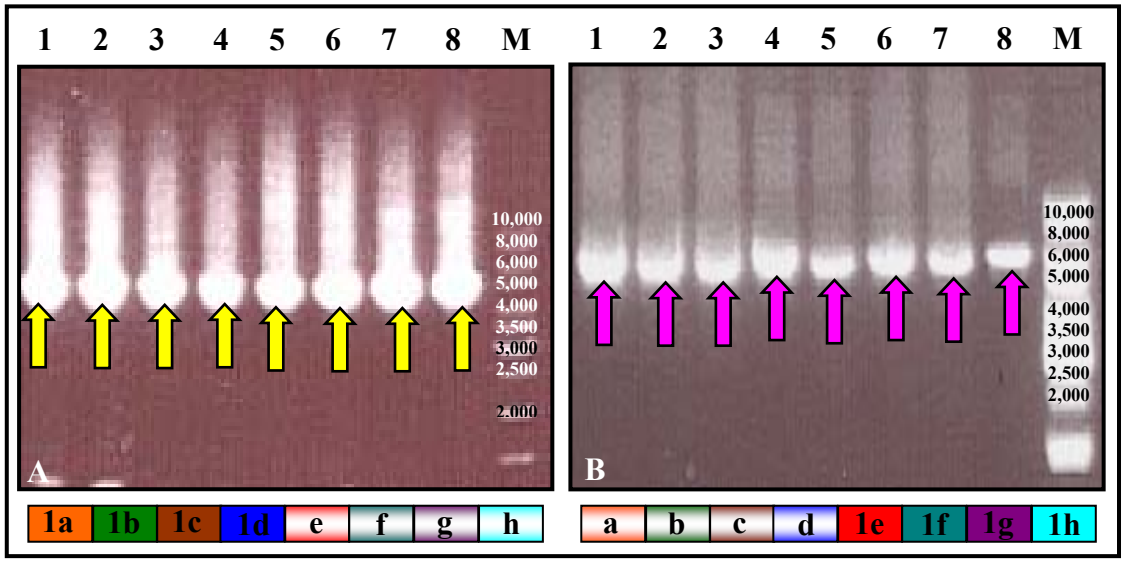


Fig. 72: A: SOE-PCR generating the cDNA coding for KLH1-*abcd*. All reactions lead to a distinct band of 5,000 bp (yellow arrows). **B:** Clone-PCR to check the right insert mass: the bands at approximately 5,600 bp (pink arrows) represent the insert KLH1-*efgh*, including the coamplified vector arms. M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

As shown in Fig. 72A, the SOE-PCR generating the cDNA of KLH1-*abcd* as well as the cloning of the previously generated cDNA of KLH1-*efgh* (Fig. 72B) is successful. Both agarose gels reveal the right mass of cDNAs: the expected 5,005 bp for the first (yellow arrows) and around 5,600 bp for the second one (pink arrows, exact calculated mass: 5,674 bp, including the vector arms). The cloning of the cDNA of KLH1-*abcd* lead to several clones having the right insert mass (data not shown). The sequencing of one clone of each showed that no stop codon occurs within the reading frame.

24.2 Generation of clones encoding six consecutive FUs

- KLH1-*cdefgh*

Starting with these clones, I tried to produce cDNAs encompassing six FUs. The cDNA encoding KLH1-*cdefgh* was generated by SOE-PCR, by fusing the cDNAs KLH1-*cd* with -1*efgh*. The resulting cDNA of 8,000 bp was cloned and thereby the corresponding entry clone was created. However, the sequencing showed that three mutations spoiling the open reading frame had been introduced within the coding sequence, wherefore the data are not shown and the clone was not further processed.

- Generation of KLH1-*abcdef* and necessary mutagenesis-PCR

Because each attempt to fuse both previously mentioned cDNAs encoding four consecutive FUs (KLH1-*abcd* and KLH1-*efgh*) failed, I tried to splice different cDNAs of varying length together to be able to combine the clones bit by bit, always keeping in mind

that the final goal of the present work the generation and cloning of the complete subunit of KLH1 was.

The SOE-PCR worked well for the following three to seven FUs-encoding cDNAs: KLH1-*cde*, -*def*, -*bcde*, -*cdef*, -*abcde*, -*bcdef*, -*defgh*, -*abcdef* and KLH1-*bcdefgh*. Several attempts were needed to clone at least four of them: KLH1-*def*, -*cdef*, -*defgh* and KLH1-*abcdef*, but the sequencing revealed that at least one mutation (insertion, deletion or point mutation, generating a stop codon or a shift within the open reading frame) in each clone spoiled the open reading frame.

The last mentioned clone was generated by fusing the cDNA of KLH1-*abcd* with the cDNA encoding -*ef*, generating the cDNA of KLH1-*abcdef*, which codes for the six wall-forming FUs of the cylindrical KLH1 molecule. The sequencing of this clone showed, however, that only one mutation has occurred: a deleted cytosine at the end of the cDNA coding for KLH1-*e* shifted the open reading frame. Therefore two mutating primers were generated, one for forward and one for reverse application, having the missing cytosine within its sequence. Starting from the defective clone, two mutagenesis PCRs were performed. The first one, using the attachment primer of KLH1-*a* in combination with the reverse mutating primer, gave a PCR product of the expected 6,150 bp (KLH1-*abcde*, data not shown). The forward mutating primer in combination with the reverse attachment primer of KLH1-*f* resulted during a second mutagenesis-PCR in a cDNA of approximately 1,450 bp (KLH1-*f*, data not shown). Both cDNAs were then again fused together by SOE-PCR (Fig. 73A) and cloned (B).

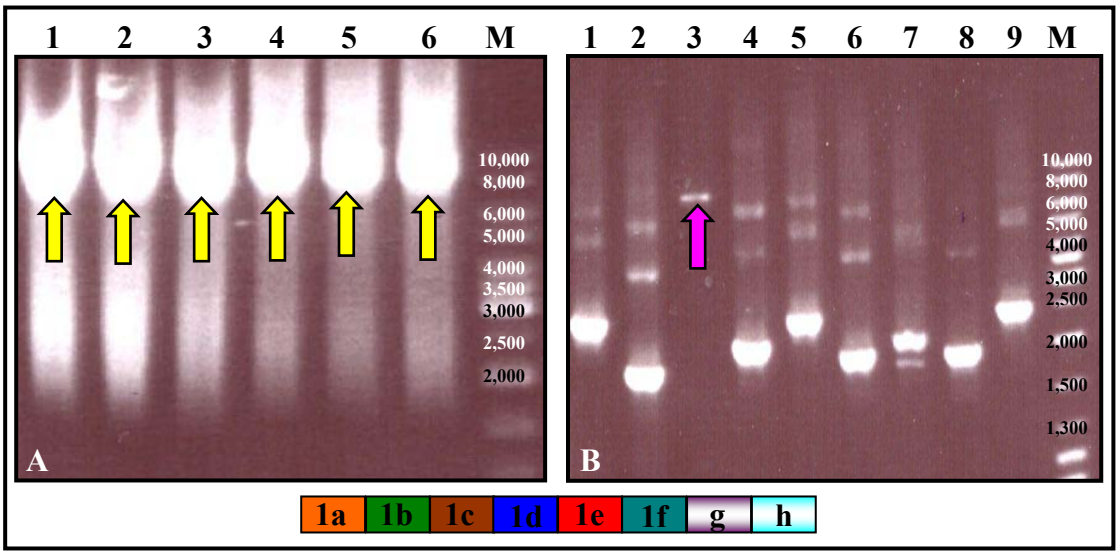


Fig. 73: A: SOE-PCR fusing two cDNAs of approximately 6,150 bp and 1,450 bp, generating a cDNA of 7,519 bp, representing KLH1-*abcdef* (lanes 1-6, yellow arrows). **B:** Corresponding clone-PCR having one clone with the right insert mass of approximately 7,800 bp, including the vector arms (lane 3, pink arrow). M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

The cloning results in one single plasmid, having an insert of the right mass (Fig. 73B, lane 3, pink arrow, 7,800 bp). The sequencing of this clone reveals that the missing cytosine has correctly been incorporated into the coding sequence and that no further mistake has been introduced.

24.3 Building up the complete subunit of KLH1 and cloning

The above mentioned cDNA encoding KLH1-*abcde* of 6,150 bp was further fused with an additional cDNA: KLH1-*fgh*, which was previously amplified from the clone encoding KLH1-*efgh* by applying the mutagenesis forward primer in combination with the reverse attachment primer of KLH1-*h*. The overlapping sequences of both cDNAs were thus located around the missing cytosine, which had been inserted by both mutagenesis primers. The SOE-PCR of these cDNAs led to the first successful generation of a cDNA comprising all eight consecutive FUs of KLH1 (data not shown). After several trials to clone directly the PCR reaction as well as the corresponding PCR clean up and the respective gel extraction, one single clone appeared to have the right mass of approximately 10,200 bp (data not shown). The sequencing however revealed that three mistakes spoiled the open reading frame.

So, different combinations of cDNAs again were tried to create the cDNA encoding the complete subunit, such as the cDNA of clone KLH1-*abcdef* with the one coding for KLH1-*fgh*, amplified from the clone KLH1-*efgh*, and the cDNA KLH1-*abcde* with KLH1-*efgh*.

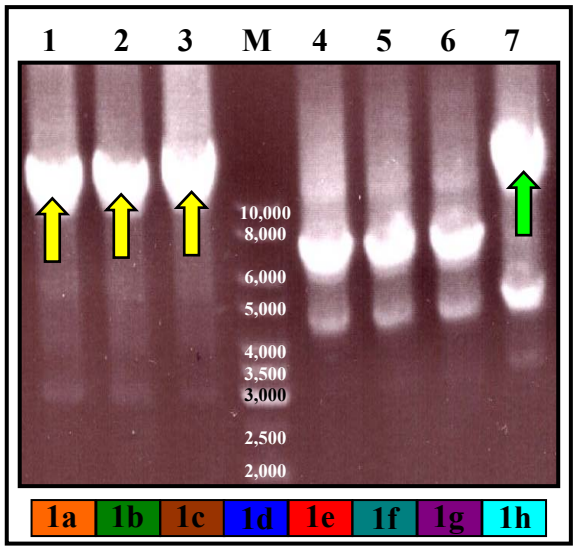


Fig. 74: Electrophoretic separation of SOE-PCRs revealing cDNA bands of approximately 10,200 bp, generated by fusing the cDNAs of KLH1-*abcdef* with KLH1-*fgh* (lanes 1 to 3, yellow arrows) and KLH1-*abcde* with KLH1-*efgh* (lane 7, green arrow), respectively. The lanes 4 - 6 show too small DNA bands, due to unspecific primer hybridization. M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

The agarose gel of these SOE-PCRs reveals strong bands at the expected 10,200 bp (Fig. 74, lanes 1 to 3, yellow arrows). Also, the cDNA KLH1-*abcde*, generated with the reverse mutagenesis primer, in combination with the cDNA of the clone 1-*efgh* leads to a distinct band which corresponds to the eight FUs (Fig. 74, lane 7, green arrow). In contrast to the previous SOE-PCRs, these cDNAs were fused using newly designed attachment primers: the forward primer including the KLH1 specific sequence for the signal peptide and the reverse primer binding at the direct end of KLH1-*h* with an additional 6xHis-tag followed by a stop codon. The lanes 4 - 6 give only too small DNA bands, due to unspecific primer hybridization.

For the cloning of the first fused fragment (Fig. 74, yellow arrows), the PCR clean up of the SOE-PCR of lane 3 was applied and this resulted, amongst others, in two clones, having a very weak band at the expected insert mass (data not shown). The second SOE-PCR (Fig. 74, lane 7, green arrow) was applied twice to an agarose gel and the method of cloning without UV damage was used (refer to: B Materials and methods, *5.14 Cloning*, - Cloning without UV damage). The DNA within the excised gel slice which had not been exposed to UV light, was subsequently cloned, which resulted in only three clones. The clone-PCR showed for all three clones a band at the expected mass of approximately 10,200 bp, but the bands were again very weak (data not shown). Hence, these five clones were cultured overnight and after a plasmid preparation, a test-PCR using the vector specific primers was performed to check the insert mass (Fig. 75).

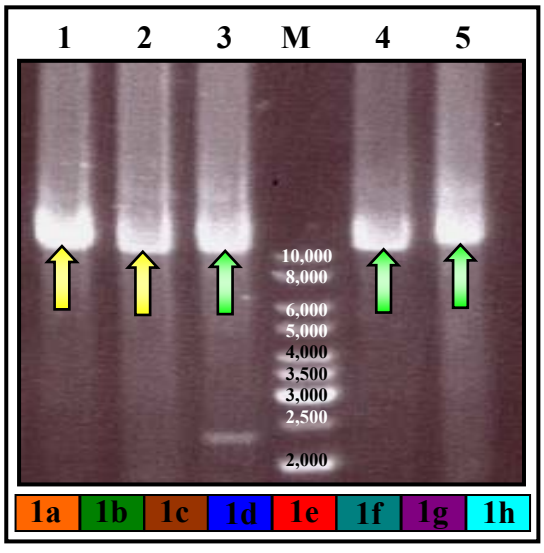


Fig: 75: Agarose gel of the test-PCR, confirming the insert mass of the five previously selected clones. All five clones have an insert mass corresponding to the cDNA of the eight FUs of KLH1 (10,558 bp), generated by fusing on the one hand the cDNAs KLH1-*abcdef* with KLH1-*fgh* and cloning of the PCR clean up (lanes 1 and 2, light yellow arrows) and on the other hand by fusing the cDNA of KLH1-*abcde* with -1*efgh* and cloning of the gel extraction without UV light exposure (lanes 3, 4 and 5, light green arrows). M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

The agarose gel (Fig. 75) reveals that all five clones contain an insert of slightly more than 10,000 bp (yellow and light green arrows, respectively), which fits the exact expected mass of 10,245 bp for the coding nucleotide sequence of the whole KLH1 subunit, including the coamplified signal peptide, the 6xHis-tag and the vector arms (total mass of 10,558 bp).

The sequencing of the first clone (Fig. 75, lane 1) showed that it is in the right reading frame with the forward attachment site and the 6xHis-tag at the 3' end, and that it contains no stop codon within the coding sequence.

So, the main goal of this work, the generation and cloning of a cDNA encompassing the complete subunit (FUs *a* to *h*) of KLH1 was successfully achieved.

25 Modification of the SOE-PCR conditions leads to an accelerated fusion of KLH2-cDNAs

Now that the method of SOE-PCR had worked well for the generation of larger KLH1 substructures up to the whole subunit, this was also tried for the other isoform, KLH2. All at this time point available cDNAs encoding KLH2 encompassed only single FUs (KLH2-*a*, -*b*, -*c*, -*d*, -*e*, -*f*, -*g* and -*h*). Each had been generated with overlapping sequences to the previous and following FU, due to the chosen gene-specific primers. To minimize polymerase-related mutations during repeated PCR cycles, the number of PCRs to fuse these cDNAs together was reduced. Therefore I tried to fuse not only two but several cDNAs together, in one PCR step. First two cDNAs were melted in one SOE-PCR approach, always using a proofreading polymerase and following the manufacturer's instructions. One difference, however, was the omission of the primers. So, both cDNAs melt throughout the denaturation step, and during 10 cycles, the complementary strands anneal to each other, and the occurring free 3' OH groups serve the polymerase as binding site. In this way, the first two cDNAs can be fused together without exponential amplification. Now, the third and consecutive cDNA was added to the previous mixture, including the outer primer pair and during the next cycles, this cDNA can be fused with the overlapping sequence of the 3' end of the already fused gene. With the help of the outer primer pair the newly fused cDNA, encoding three FUs of KLH2, can be amplified.

25.1 Consecutive fusion of three KLH2 cDNAs in one SOE-PCR step

This technique was tried for the FUs KLH2-*a*, -*b* and -*c*, as well as the combination of -2*d* with -*e* and -*f*. The remaining two FUs KLH2-*g* and -*h* also should be fused (Fig. 76).

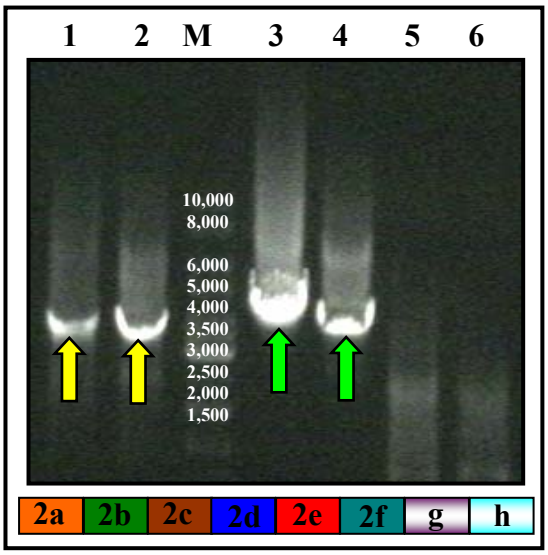


Fig. 76: SOE-PCR fusion of three consecutive FUs in one PCR step. Lanes 1 and 2: cDNA of KLH2-*abc* at approximately 3,850 bp (yellow arrows, 50 °C and 60 °C, respectively). Lanes 3 and 4: cDNA of KLH2-*def* (green arrows, 3,808 bp, 50 °C and 60 °C, respectively). The lanes 5 and 6 should reveal KLH2-*gh*, but this failed.

As shown in Figure 76, fusion of three different cDNAs during one single PCR step worked well. The fused cDNAs coding for the FUs KLH2-*abc* are represented by distinct cDNA bands visible in the agarose gel at the expected mass of 3,836 bp (lane 1 and 2, 50 °C and 60 °C, respectively, yellow arrows). In the lanes 3 and 4 the SOE-PCRs coding for the FUs -2*d*, -*e* and -*f* are applied which give a cDNA band of 3,808 bp. The green arrows indicate the location at the approximate right mass. Interestingly, the fusion of the two consecutive FUs KLH2-*g* and KLH2-*h* fails (lanes 5 and 6).

25.2 Fusion of four consecutive FUs of KLH2 by different approaches

The first approach used the same strategy than before: two cDNAs were first mixed, annealed to each other to allow the hybridization of the overlapping sequences during 10 cycles and then the third cDNA was added during the next ten cycles. Finally, the last and fourth cDNA in combination with the outer primer pair was added. This was tried for the four cDNAs KLH2-*a*, -*b*, -*c* and -*d* which should lead to the cDNA of KLH2-*abcd*, as well as for the cDNAs encoding KLH2-*efgh* (from KLH2-*e*, -*f*, -*g* and -*h*). In parallel, the fusion of the respective mentioned cDNAs was also tried by mixing them all directly together at the beginning of the SOE-PCR. After ten cycles of denatura-

tion, annealing and elongation, the corresponding outer primers were added. This change can, additionally, shorten the cycling duration.

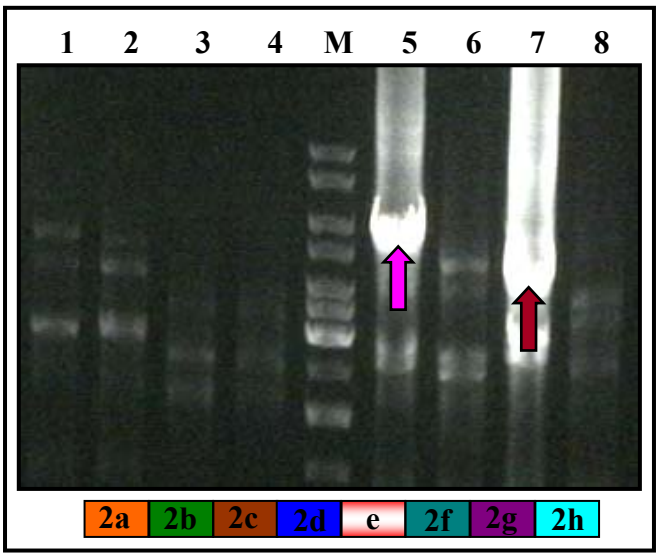


Fig. 77: SOE-PCR fusion of four cDNAs of KLH2. Lanes 1 to 4: cDNAs mixed consecutively; lanes 5 to 8: four cDNAs added all together (lanes 1, 2, 5 and 6: KLH2-*a*, -*b*, -*c* and -*d*; lanes 3, 4, 7 and 8: KLH2-*e*, -*f*, -*g* and -*h*). Lanes 1 - 4: no distinct band. Lane 5: a distinct cDNA band representing KLH2-*abcd* (5,026 bp, pink arrow). Lane 7: the present cDNA is only about 4,200 bp (dark red arrow).

Both attempts to fuse four cDNAs of KLH2 by consecutive addition fail for the first four FUs (Fig. 77, lanes 1 and 2, KLH2-*abcd*) as well as for the last four (lanes 3 and 4, KLH2-*efgh*), although it was found that it worked well when fusing only three FUs together (Fig. 77). Interestingly, the mixing of the respective four consecutive FUs all together directly at the beginning of the SOE-PCR and adding the outer primer pair after 10 cycles work well: lane 5 shows a clear band at the expected mass of 5,026 bp (pink arrow) representing the cDNA of KLH2-*abcd* (annealing temperature of 50 °C). The trial at 60 °C (lane 6) fails. The same is observed for the generation of the cDNA encoding KLH2-*efgh*. Only at an annealing temperature of 50 °C (lane 7) is a distinct DNA band visible. But the mass of this band is too small: approximately 4,200 bp (dark red arrow) instead of 5,441 bp. After checking the applied SOE-PCR components, it was found that a wrong upstream primer has been used: the primer used binds at the end of KLH2-*e*, therefore generating a cDNA encoding only for KLH2-*fgh*, whose mass corresponds to the band appearing in lane 7.

25.3 Generation of cDNAs encoding six and seven FUs of KLH2

To reach the final goal, the splicing of the complete subunit, different approaches were tried following the above mentioned variation of SOE-PCR. The generated cDNA en-

compassing KLH2-*abc* was fused together with the cDNA encoding KLH2-*def*. The corresponding *att*-site PCR was performed in parallel for the subsequent cloning of the resulting fragment. Additionally, the fusion of the four mentioned cDNAs with a third was tried, by adding the cDNA encoding KLH2-*g* after ten cycles, in combination with the corresponding outer primer pair. This generates a cDNA encoding KLH2-*abcdefg*.

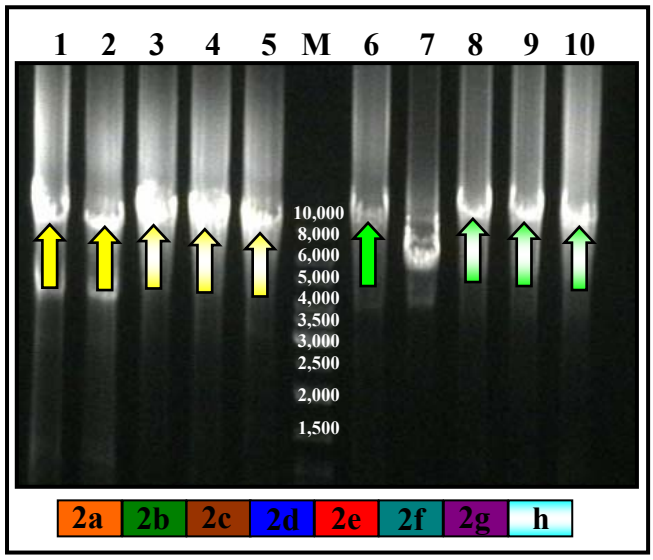


Fig. 78: SOE-PCR generating the cDNA of KLH2-*abcdef* (7,534 bp, lanes 1 and 2, yellow arrows) and its corresponding *att*-site PCR (lanes 3, 4 and 5, light yellow arrows) and KLH2-*abcdefg* (8,774 bp, lane 6, green arrow) with its respective *att*-site PCR (lanes 8, 9 and 10, light green arrows). M represents the marker GeneRuler™ DNA Ladder Mix, in bp.

The UV light shows the results of the different SOE- and *att*-site PCRs (Fig. 78): while the SOE-PCR generating the cDNA KLH2-*abcdef* works well (lanes 1 and 2: cDNA band at the expected mass of 7,534 bp, yellow arrows, 50 °C and 60 °C), the SOE-PCR fusion of the cDNAs -2*abc* with -2*def* and additionally -2*g* works only at an annealing temperature of 50 °C (8,774 bp, lane 6, green arrow). The corresponding *att*-site PCRs work well for both fragments at the three different annealing temperatures: 50 °C, 60 °C and 70 °C (light yellow and light green arrows, lanes 3, 4 and 5 as well as lanes 8, 9 and 10, respectively). However, the cloning of the PCRs and the PCR clean-ups of these cDNAs failed.

In a different approach, the further addition of the cDNA of KLH2-*h* with ten more cycles was tried. The agarose gel showed that the consecutive splicing of the cDNAs up to the whole subunit of KLH2 did not lead to the expected band of 10,406 bp (data not shown). Each further trial to generate a cDNA encompassing the whole subunit of KLH2 failed.

26 Resuming overview of the cDNAs and clones of KLH1 and KLH2, generated in the present work

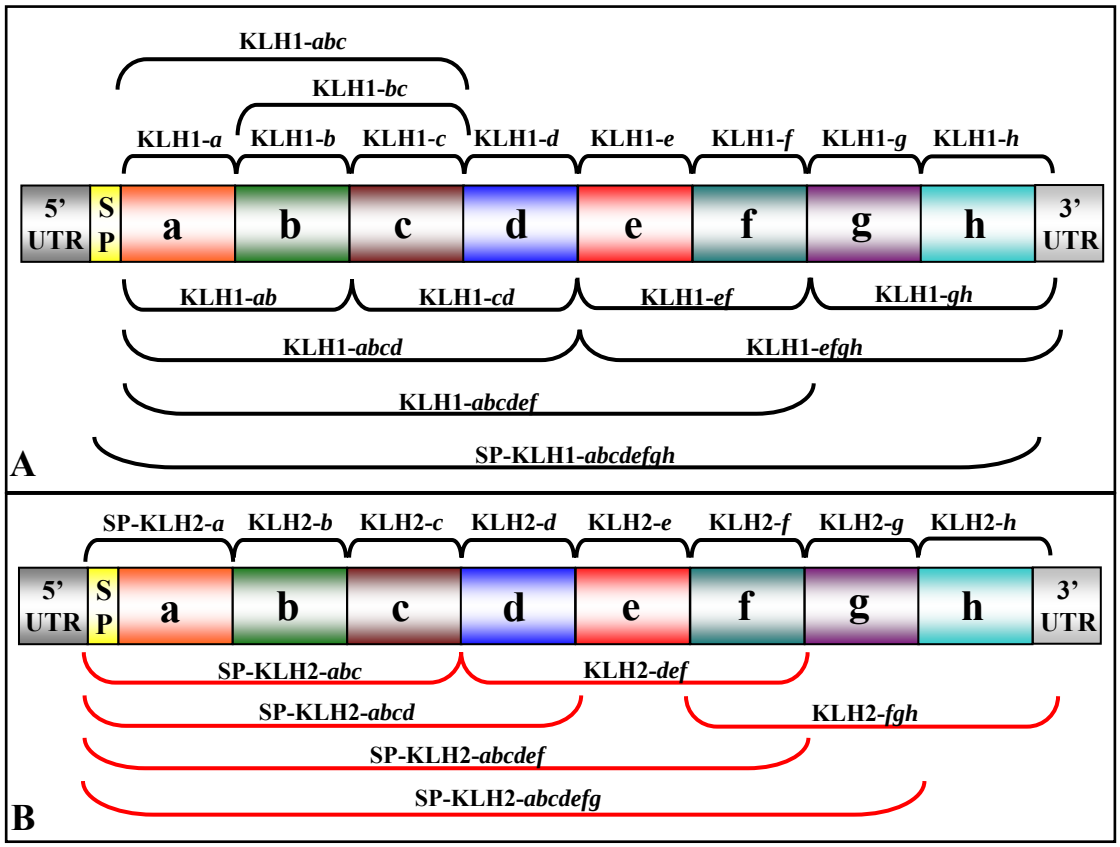


Fig. 79: Schematic overview of the clones (black brackets) and cDNAs (red brackets) of KLH1 (**A**) and KLH2 (**B**), generated during the present work (SP: signal peptide).

D Discussion

The focus of the present research lays emphasis on the recombinant expression of KLH, as an attempt to establish procedures for future easy, cost- and time-effective production of this highly potent immuno-stimulator.

1 Applicability of the Gateway™ technology

1.1 att-site PCR

The cloning procedure, applying the Gateway™ technology, is based on a recombination reaction between specific attachment (*att*) sites. To generate PCR products suitable for use as substrates in a Gateway™ BP recombination reaction with a donor vector, these *att* sites need to be incorporated into the PCR products. Due to the resulting length of the corresponding primers (31 - 49 bp for *att*B1 plus 18 - 25 gene specific nucleotides; 30 - 32 bp for *att*B2 site, plus 18 - 25 gene specific nucleotides) it is recommended, following the manufacturer's instruction, to design two primer pairs for two different PCRs: the first for the template specific amplification and a second to install the complete *att*B sequences. After several trials using this strategy, which constantly resulted in unexpectedly small cDNAs, it was found that the installation of the complete *att*B sites using only two long primers worked much better. Apart from the positive results obtained by this change, the by-pass of the second PCR is an additional bonus concerning time and potential polymerase-related inserted mutations into the coding sequence.

1.2 BP and LR recombination reactions

After the accomplishment of numerous cloning reactions during the present work, it became clear that the BP recombination reaction is the weak spot of the whole method. Although the Gateway™ BP Clonase™ II enzyme mix catalyzes in a very specific reaction, the *in vitro* recombination between an *att*B-PCR product and an *att*P-containing donor vector to generate an *att*L-containing entry clone, each clone-PCR following a BP-reaction revealed almost 50 % of false positive or inserts of incorrect size. This result becomes even worse with increasing insert size.

A drastic amelioration was obtained by using the method of cloning without UV light exposure of the DNA. The consequences of UV damage to DNA are single strand breaks within the double stranded DNA molecule and the generation of smaller degradation products. This, in turn, can lead to the fact that these smaller degradation products will

be cloned more efficiently into the vector than the desired molecule. Using cloning via site-specific recombination this can normally not occur because the smaller DNA molecules do not have the specific attachment sites. However, already the first assay using the cloning method without previous exposure of the DNA to UV light lead to a 100 % recovery of positive clones, all having the right insert size. So, UV damage of DNA molecules must be considered as a high risk factor for the failure of cloning procedures, even for recombination reactions.

But once the cDNA has been cloned successfully into one of the Gateway™ donor vectors, the transfer into one of the destination vectors to create the corresponding expression clone worked in nearly every case. Two exceptions remain: KLH1-*de* and KLH1-*fg*. Although it was possible to clone both cDNAs into the donor vector pDONR™221, the correct subcloning into pDEST™17 repeatedly failed. While the clone-PCR indicated that the transferred inserts had the right size of about 2,500 bp, the sequencing of the clones, especially resulting in KLH1-*fg*, showed that recombination had not only occurred at the *att*-sites but also within the insert, changing the coding sequence of the KLH1-clones and the open reading frame. So, the sequence coding for KLH1-*fg* was aligned with the *att*B1 site as well as with the *att*B2 site (relative alignments, http://xylian.igh.cnrs.fr/bin/lalign-guess.cgi). The first one reveals that an identity of 76.5 % is given over 17 nucleotides at a position around the first 50 nucleotides of this cDNA. The second alignment shows that an even higher score of identity was found at the end of this cDNA sequence, namely between the nucleotides 2550 - 2584. Here, the identity over 34 nucleotides is about 67.6 %. As mentioned by Argos and co-workers as long ago as 1986, conservative, site-specific recombination by the enzyme integrase (Int) is restricted to specific sites on both partner molecules and requires only very short regions of DNA homology. So, the high sequence identities between the cDNA encoding for KLH1-*fg* and both attachment sites can be an explanation for the wrong binding and catalyze of the recombination enzyme integrase. Although even higher scores are found when aligning the remaining KLH1 sequence with both *att*B sites, the highest scores for each binding site where not found in the right 5′ - 3′ direction to allow correct binding of the enzymes for directed recombination reaction.

1.3 Recombinant expression using different Gateway™ destination vectors

- pDEST™14

During the first expression experiments using pDEST™14, no additional bands of recombinant protein were found within the induced cultures of KLH2-*c* expressing clones, but within the uninduced ones. The sequencing of the expression clones showed, indeed, that both the Shine-Dalgarno consensus sequence and the starting methionin at the 5′ end were in the right reading frame with the coding sequence of KLH2-*c*. A basal expression level, as described in the handbook "*E. coli* expression system with Gateway™ technology" from Invitrogen, Karlsruhe, can explain the additional band within the uninduced cultures appearing even without addition of the inducing L-arabinose. This can be repressed by the addition of glucose. However, basal expression levels cannot explain the missing protein band in the induced cultures.

- pDEST™24

The subcloning of the KLH2-*c* encoding insert into pDEST™24 and the consecutive induction of the clones showed that now, two additional protein bands of 48 kDa and 44 kDa can be observed within the induced cultures which were not present in the uninduced samples (data not shown). The first one, at approximately 48 kDa could correspond to KLH2-*c,* whose calculated molecular mass is about 48,031 Da, if the expression starts at the first methionine, previously inserted by the forward attachment primer. Looking at the amino acid sequence of KLH2-*c*, one can see that the next methionine within the coding sequence is located only 31 amino acids downstream of this first one. If the translation starts only at this second methionin, however, the calculated molecular mass of the resulting protein band should be about 44,488 Da. This would fit exactly the second additional protein band. So, if both methionines are used simultaneously as starting codons for translation due to their close localisation, two bands of recombinant protein, one at 48 kDa and one at 44 kDa, could appear.

- pDEST™17

First the subcloning of the KLH2-*c* cDNA into the destination vector pDEST™17 led to the successful recombinant expression, as shown in Fig. 27. The next section will deal with further details.

2 Recombinant expression of single functional units of KLH1 and KLH2

2.1 Starting with KLH2-c

During the first successful recombinant expression of a KLH substructure in *E. coli*, namely KLH2-*c* in combination with the destination vector pDEST™17 (Fig. 27), very strong protein bands were seen at the expected mass of approximately 50 kDa. The masses shifted from the above mentioned one (48,031 Da) due to the coexpressed N-terminal 6xHis-tag of this vector (exact mass: 50,592 Da). As shown in the Figures 28, B, and 29, B, the recombinant protein bands were detected in a very specific way by both the monoclonal α-His-antibody and the polyclonal α-KLH2c-antibody, respectively. Interestingly, small protein bands of the same size were also present in the uninduced cultures. This could be due to basal expression levels in the bacteria cells having the plasmid, but without the addition of the inducer L-arabinose.

2.2 Single FUs of KLH1

The first recombinantly expressed KLH1 cDNA was FU-1*h*, which led to a distinct protein band of the expected 62 kDa (Fig. 30). The Western blot analyses confirmed this result by strong hybridisation of the specific α-His-antibody (Fig. 31) and α-KLH1-antibody (Fig. 32). So, to rule out whether the remaining FUs of KLH1 could be expressed recombinantly in *E. coli* in the same easy way, all cDNAs were cloned into pDEST™17. The expression of the functional units KLH1-*a*, -1*c*, -1*d*, -1*e*, -1*f* and -1*h* in the *E. coli* host BL21™-AI worked very well (Fig. 34) and resulted in clear recombinant protein bands at the expected mass for the corresponding FU, while the FUs KLH1-*b* and KLH1-*g* showed no recombinant protein band. The expression of these two FUs was repeated by changing different expression conditions: the culture medium (NZY instead of LB-medium), the expression duration (4 hours instead of overnight expression) and also the L-arabinose concentration was varied (ranging from 0.01 to 5 %) but the expression remained unsuccessful. Only the change of the expression bacterial strain from BL21™-AI to BL21™(DE3)pLysS led to the successful expression of KLH1-*b*, revealing a recombinant protein band of 52 kDa (Fig. 37). The BL21-AI™ *E. coli* strain offers a good regulation of expression for production of toxic proteins using the T7 promoter. Upon induction, the *araBAD* promoter exhibits strong expression, resulting in high-level protein production. This (possibly too) strong expression may hinder the correct transla-

tion of difficult recombinant proteins. In contrast, BL21™(DE3)pLysS *E. coli* carries both the DE3 lysogen and the plasmid pLysS. This latter constitutively expresses low levels of T7 lysozyme, which reduces basal expression of recombinant genes by inhibiting basal levels of T7 RNA polymerase. This system is specially designed for an improved folding of recombinantly expressed proteins due to a tight regulation, so that the proteins can fold more slowly. Nevertheless, the expression of KLH1-*g* remained unsuccessful. So, the expression temperature was in addition changed for these clones and the expression course was monitored. The decrease of this parameter after induction from 37 °C to 28 °C can additionally enable a slow protein translation and correct folding. As shown in Figure 38, the apparition of the additional recombinant protein band of KLH1-*g* at the expected size of 50 kDa began after 4 hours.

So, each single functional unit of KLH1 was eventually expressed in *E. coli*, with some requiring different conditions and additional time for successful expression.

2.3 Single FUs of KLH2

It is known that the corresponding FUs of the two different isoforms are more closely related than FUs of one isoform to each other (Miller *et al*., 1998; Lieb *et al*., 2000). For example, there is 62.7 % nucleotide identity between KLH1-*a* and KLH2-*a,* compared to only 53.2 % identity between KLH1-*a* and KLH1-*b*. Based on this knowledge, I attempted to express also all eight single functional units of KLH2 first in BL21™-AI, to determine if the expression difficulties of the single FUs observed for KLH1-*b* and KLH1-*g* are due to specific sequence motives. If this is the case, KLH2-*b* and KLH2-*g* would exhibit the same expression problems. Interestingly, the expression worked well for every single FU of KLH2 (Fig. 41). However, FU-2*b* was expressed with almost 300 bp of FU-2*c* coexpressed at the C-terminus, due to the downstream primer localisation. This could have led to additional stabilization of FU-2*b*, in contrast to FU-1*b*. Only KLH2-*a* and KLH2-*h* had only weak and blurred protein bands and while FU-2*a* was not detected by the α-His-antibody (Fig. 42, B), the protein was weakly detected by the α-KLH1-antibody (Fig. 43, B). KLH2-*h* was detected by both antibodies, but they also gave only weak immuno-precipitation products. So, the hypothesis that the corresponding FUs of KLH2 will behave in the same way than the FUs of KLH1 concerning recombinant expression in *E. coli* was not correct. Nevertheless, each single FU of KLH2 is successfully expressed in *E.coli*.

3 Influence of a eukaryotic signal peptide on the recombinant expression of proteins in a prokaryotic system

Several studies have been undertaken on the prokaryotic signal peptide (SP) amino acid sequence determination and its influence of protein localisation within the bacterial cell (Perlman & Halvorson, 1983; Sjöström *et al.*, 1987). The synthesis of proteins usually takes place in the cytoplasmic compartment, either on free or membrane-associated ribosomes. But, final protein localisation can be at the inner membrane, the periplasmic space, the outer membrane or the proteins can be secreted (like toxins), depending on the signal peptide sequence. Characteristics for pro- and eukaryotic signal peptides are the N-terminal position and the sequence of approximately 13 - 36 amino acids. It begins always with a methionine, which serves simultaneously as a start codon for translation, and presents a hydrophobic core, flanked by hydrophilic amino acids. The signal peptides of KLH1 and KLH2 me*et al*l these demands (Fig. 80A and B).

A **MLSVRLLIVVLALANA**	**1 - 16**
B **MWTILALLTATLLFEGAFS**	**1 - 19**

Fig. 8: A: Signal peptide amino acid sequence of KLH1. **B:** Signal peptide amino acid sequence of KLH2. The start methionines are marked in green, hydrophilic, flanking amino acids in red and the hydrophobic core is indicated in blue. Preferred cleavage sites are underlined.

As early as 1983, Perlman and Halvorson determined that Ala-X-Ala is the most frequent sequence preceding the signal peptidase cleavage site; Ala stands for the amino acid alanine (A) and X can be any amino acid. This motif is found in KLH1 (Fig. 80A, underlined: ANA). They further proposed the existence of a general signal peptidase recognition sequence A-X-B. Following their claims, alanine residues would be likely to participate in A or B in peptidase cleavage, and position A could also include larger aliphatic amino acids, such as leucine (L), valine (V) and isoleucine (I). The residues alanine, glycine (G) or serine (S) are generally found at position B. A sequence motif confirming this proposed peptidase cleavage sequence is found in KLH2 and is underlined in Fig. 80B: AFS. Proteins naturally synthesized as precursor proteins (preproteins) with an N-terminal extension known as a leader or signal peptide, are mostly translocated across lipid bilayers (Black, 1993). Leader peptidases (LPs) are membrane bound proteolytical enzymes present in both prokaryotes and eukaryotes which specifically catalyze the hydrolysis of the peptide bond between the N-terminal leader and the mature sequence of preproteins (San Millan *et al.*, 1989; Wickner *et al.*, 1991). The pro-

karyotic peptidase responsible for this activity, LP, is known to consist of a single polypeptide chain (Wolfe, *et al*., 1983), unlike the eukaryotic equivalents, which consist of between two and six polypeptides (Baker & Lively, 1987, Greenburg *et al*., 1989; YaDeau *et al*., 1991).

The cDNA encoding for the single functional unit KLH1-*a*, generated by RT-PCR (Fig. 23) was first reverse transcribed not only with an overlapping sequence to FU-1*b* at the 3′ end but also with the KLH1-specific signal peptide at the 5′ end (SP-KLH1-*a*). The first expression experiments in BL21™-AI, however, led to no visible additional protein band, when applying aliquots of the pelleted expressing bacterial cells onto an SDS-PAGE (data not shown). So, the cDNA was amplified using a gene-specific primer cutting off the signal peptide. Now, the expression of KLH1-*a* worked well and resulted in a strong band of recombinant protein at the expected mass of 50 kDa (Fig. 33) which was also well recognized by both the α-His-antibody and the α-KLH1-antibody (Fig. 34 and 35). The first idea was that the translation of the recombinant protein can be hindered by the preceding signal peptide.

The recombinant expression of SP-KLH2-*a* showed a similar result and gave new insights: the recombinant protein band seen in the Coomassie staining was only weak and blurred (Fig. 40) and no specific recognition took place using the α-His-antibody (Fig. 41). The α-KLH2-antibody, instead, gave a weak detection of a protein band at the expected size (Fig. 42). This leads to the conclusion that the KLH-specific coexpressed signal peptide potentially permits the secretion of the recombinantly expressed protein into the surrounding medium. The weakly detected protein can be membrane-bound or a translocated protein. This would additionally argue for the subsequent cleavage of the signal peptide by the leader peptidase, so that the 6xHis-tag is cleaved at the same time, due to its location at the N-terminus of the signal peptide. Therefore, it would not be recognized by the α-His-antibody. This idea is plausible and does not preclude the more specific recognition of the KLH2-*a* protein by the α-KLH2-antibody, shown in Figure 43. If the signal peptide is cleaved with the 6xHis-tag, only the application of a KLH-specific antibody would detect the recombinant protein, whether located within the cell or secreted into the surrounding medium.

4 Successive recombinant expression of larger KLH1 substructures for testing of the expression capacity limit of *E. coli*

4.1 Success and failure to express two consecutive FUs

The first experiment trying to express two consecutive functional units of KLH was done with the clone coding for KLH1-*ef*. However, this led only to a protein band of 26 kDa instead of the expected ~100 kDa.

The first successful expression of two FUs was achieved with the clone encoding for KLH1-*bc*. Clear additional protein bands at the expected size of 100 kDa were found in the Coomassie staining (Fig. 43) as well as in the detection by both the α-His-antibody (Fig. 44B) and the α-KLH1-antibody (Fig. 45B). Surprisingly, the expression of FU-1*b* in combination with KLH1-*c* showed no expression problems, as for the single FUs. Structural properties of the KLH1-*c* protein and/or its mRNA may stabilize and help the correct translation and subsequent folding of KLH1-*b* within the bacterial cell. However, both antibodies revealed that large quantities of smaller degradation products were also present, only in the induced cultures. Especially the α-His-antibody detected these additional protein bands, indicating that the 6xHis-tag is still present and that the recombinant protein is maybe degraded from the C-terminus.

So, after generation of two additional substructures of KLH1 by SOE-PCR, namely KLH1-*ab* and KLH1-*cd*, I also tried to express them recombinantly in *E. coli*. Although FU-1*b* and FU-1*c* had been successfully expressed together in KLH1-*bc*, the expression of these two additional substructures, each containing one of both mentioned FUs, failed.

Only one further KLH substructure was expressed successfully: KLH1-*gh*. This was kindly been performed by [redacted] (Fig. 46) and interestingly, two additional protein bands appeared within the induced cultures, one at the expected 110 kDa and a second one at ~120 kDa, which were both not present in the uninduced cultures. The monoclonal α-His-antibody used in the Western blot analysis revealed that the expected protein band of 110 kDa represents KLH1-*gh*, giving a weak brown precipitation product, while the second protein of 120 kDa disappeared (Fig. 47).

At this time point, it is not possible to decide whether the expression capacity limit of *E. coli* BL21™-AI had already been reached, due to the different expression behaviour of the clones encompassing two consecutive functional units of KLH1. The expression of KLH1-*bc* resulting in the 100 kDa protein was achieved, but others of the same size

failed (KLH1-*ab* and KLH1-*cd*). Although the expression of KLH1-*gh* was initially successful (110 kDa), this result was not reproducible.

4.2 Is it possible to express KLH components above 110 kDa in E. coli?

The generation of a cDNA by SOE-PCR encompassing three consecutive functional units, namely KLH1-*abc*, can give new insights concerning the expression limits of *E. coli,* considering the size of the recombinant protein product. After the successful cloning into the expression vector, the protein was expressed in BL21™-AI by [redacted]. This effectively generated a protein of 150 kDa (Fig. 49). Interestingly, the additional second protein band of 120 kDa, already found for KLH1-*gh*, was again visible in the Coomassie staining. The Western blot analysis using the α-His-antibody, did, indeed, again show that the protein of 150 kDa represents the recombinant protein KLH1-*abc*, while the band at 120 kDa disappeared (Fig. 50).

So, the answer to the question asked in the section heading is definitely *yes*, however, again this result was not reproducible.

4.3 Proteins around 150 kDa may be the limit

As the method of SOE-PCR was working well, two cDNAs encompassing four consecutive FUs of KLH1 were generated: KLH1-*abcd* and KLH1-*efgh*. A protein of ~ 200 kDa was expected from the expression of KLH1-*abcd* (194,642 Da), but the induced cultures showed no additional protein band at this size (data not shown). The substructure KLH1-*efgh*, coding for the second half of the subunit, is slightly larger than -1*abcd,* due to the C-terminal extension of FU-*h*, namely 205,237 Da. A very strong additional protein band was found within the induced cultures, which was not present in the uninduced cultures. But the mass of 26 kDa was too small (data not shown).

Looking at all SDS-PAGEs of the present work applying the *E. coli* BL21™-AI expression cultures, one can see that the largest expressed self-proteins of this bacterial strain do not exceed a mass of ~ 170 kDa. The expression capacity concerning the mass of recombinant proteins could also be around this upper limit. However, the failure to express the KLH substructures resulting in proteins around 200 kDa does not provide evidence that *E. coli* is unable to express larger proteins.

5 Recombinant protein localisation analysis within the bacterial cells and subsequent isolation steps

The accumulation of inactive protein within a host cell after over expression of recombinant proteins can lead to the formation of intracellular inclusion bodies. Although there are several different theories on inclusion body formation, incorrect protein folding seems to be the most important. Newly synthesized proteins can initially form several different intermediates of partially folded structures before achieving their final native conformation. Some of these folding intermediates have hydrophobic areas on their surface, which may later be located inside of the native protein. Extremely high local protein concentration, due to over expression, can mean that these hydrophobic regions interact with each other and lead to aggregation of the protein, with deposition in inclusion bodies.

Reasons can therefore be:

- high concentration of the protein, broadly beyond the physiological limit
- the absence of other coexpressed proteins required for correct folding, such as chaperons and protein-disulfide-isomerases
- the absence of post-translational modification, such as glycosylation, which can assist the correct folding and assist solubilisation of the protein
- that the recombinant protein is normally secreted or located in specific compartments of the cell
- the intracellular *milieu* of the host cell is different (pH, redox potential, ion content)

For the recombinantly expressed substructures of KLH, several of these parameters are likely to be important: considering the amount of recombinantly expressed protein in contrast to the other bacterial proteins, it is clear that KLH is over expressed. Prokaryotic chaperons may catalyse protein folding in a different way to the eukaryotic folding pathway and therefore can lead to incorrect protein folding. KLH is also thought to have three disulfide bridges in every FU, which therefore requires the presence of protein-disulfide-isomerases. The numerous glycosylations are the most important post-translational modifications in KLH (van Kuik *et al*., 1990; Kamerling & Vliegenthart, 1997), which are not be introduced by prokaryotic cells due to the absence of ERs. Additionally, native KLH is an extracellular, secreted protein, freely soluble in the hemo-

lymph. Thus, four of the five above mentioned reasons for inclusion body formation are applicable to recombinant KLH.

After several unsuccessful attempts to purify the recombinantly expressed KLH substructures using affinity chromatography, all indications suggested that the proteins may be located within inclusion bodies. The experiment for determination of target protein solubility confirmed this result (Fig. 52): the recombinant protein was found to be located in the insoluble pellet.

The proteins of inclusion bodies are normally fully synthesized and some can, nevertheless, achieve a completely native conformation. In most cases, however, denaturation of the inclusion bodies and therefore also of the proteins is necessary for complete solubilisation. Chaotropic agents such as 8 M urea and reducing agents such as β-mercaptoethanol are often employed to produce solubilisation.

As shown in Fig. 53 and Figs. 55 to 57, the applied protocol 1 using 8 M urea led to satisfactory results: the three substructures KLH1-*bc*, KLH1-*h* and KLH2-*c* were successfully isolated from the inclusion bodies. The distinct bands of recombinantly expressed protein were clearly shown at the expected mass, within the soluble supernatant (SN) fractions.

The purification of the recombinant proteins from remaining bacterial contaminatory proteins using different NI-NTA columns and protocols, however, failed again. The buffer applied for solubilisation of the inclusion bodies was not compatible with the binding conditions needed for the columns. Therefore, the isolated proteins were dialyzed against 50 mM Tris/HCl to remove the detergent. Different salt concentrations and combinations were tested, according to the appropriate column protocols. But, without the addition of specific ions the protein did not bind to the column, and after addition of these ions, the protein could no longer be eluted.

Looking back on the consecutive protein isolation steps (Fig. 56) it is apparent that with every incubation step of the pellet in urea, the recombinant protein containing supernatant became progressively pure. So, since affinity purification was not possible, supernatant no 5 (SN5), having nearly no contaminations, was taken for the consecutive steps, thus enabling further analyses.

6 Poor antibody recognition of recombinant KLH1-*h* is due to both denaturation and the missing N-glycosylation

Several previous studies of our research group have shown that the native FU-1*h* is easily isolatable from the other FUs by proteolytical cleavage by repeated addition of 2 % V8 protease (Gebauer & Harris, 1999). Application to an anion exchanger followed by analysis of the resulting fractions by crossed-line and two-dimensional immunoelectrophoresis revealed the purest KLH1-*h* containing fraction. Due to current investigations on crystal structure, this FU was available and kindly donated by [redacted]. At that time point the clone encoding for KLH1-*h* had already been expressed in *E. coli* and the protein purified from the inclusion bodies.

The direct comparison by two-dimensional immunoelectrophoresis (Figs. 60 and 62) and the absorption spectra (Figs. 61 and 63) of these two KLH1-*h* preparations of different origin did, however, show differences: while the native FU-1*h* revealed a distinct peak of precipitated protein against the α-KLH1-antibody and the two characteristic peaks for biologically active hemocyanin, the recombinantly expressed protein possessed none of these properties. It was therefore concluded that the *E. coli*-derived KLH1-*h* had no biological activity. One reason therefore could be the absence of copper atoms within the bacterial cultivation medium. Oxygen cannot be bound reversibly to the hemocyanin FU if copper atoms are not present within the active site. Biological activity of the protein was thus impossible. The poor recognition by its specific antibody could be for two reasons: first, the denaturation by 8 M urea of the recombinantly expressed protein may lead to an incorrect protein folding so that specific structural epitopes may now be misfolded or located in an inner area of the protein; second, the recombinant KLH-1*h* does not have its characteristic glycosylation (due to the expression in a prokaryotic system), which is known to be required for antibody recognition.

To decide which explanation is more likely or if maybe both play a role, the native KLH1-*h* was first denatured by overnight incubation in 8 M urea. The subsequent rocket immuno-analysis showed that the distinct peak of precipitated protein, normally seen for native KLH1-*h* against the specific α-KLH1-antibody, had disappeared completely (Fig. 64). So, the denaturation of the recombinant protein and the subsequent incorrect conformation could indeed be one reason for the poor recognition by the antibody. Denaturation does not lead necessarily to an unfolded protein conformation, with all epitopes still accessible. Rather, the protein is only partially and therefore incorrectly folded, sometimes with altered epitopes. Granted that the two disulfide bridge bonds present in

native KLH-1*h* had been correctly formed in the recombinantly expressed protein, the addition of 1 mM DTT within the 8 M urea solution would split them. Hence, it is thought that, due to this incorrect folding, essential antibody-recognition motives, normally present on the surface of the molecule, are now altered or located in the inner protein molecule and inaccessible for the antibody.

As mentioned by Söhngen and coworkers in 1997, the approximate mass of KLH1-*h* is located around 60 ± 5 kDa. Keeping in mind, that the native FU, used in this work, was previously isolated by digestion with V8 protease, this enzyme is known to cleave preferentially after the amino acids glutamic (E) and aspartic acid (D) within the linker peptides located between the different FUs. Looking at the amino acid sequence of KLH1-*h*, five potential different cleavage sites (indicated by vertical arrows, 1 - 5) result from the linker peptide amino acid sequence (Fig. 81).

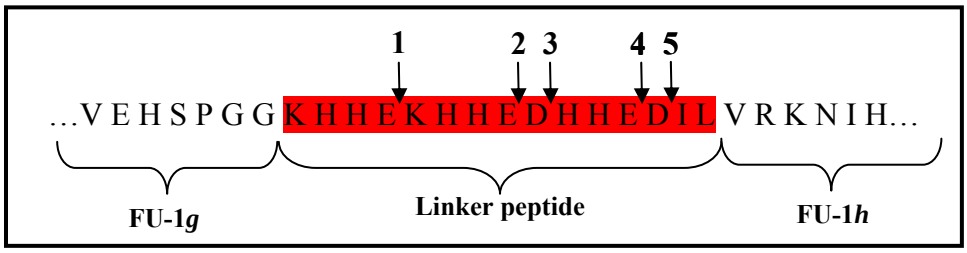


Fig. 81: Amino acid sequence of the linker peptide between FU-1*g* and FU-1*h* of KLH1 (highlighted in red). The five putative V8 protease cleavage sites, preferentially located after glutamic and aspartic acid (E and D, respectively), are indicated by vertical black arrows (1 - 5).

If all five potential cleavage sites within the amino acid sequence of the linker peptide between FU-1*g* and FU-1*h* are used by the V8 protease, five different proteins decreasing in mass can result: 57,956 Da, 57,425 Da, 57,310 Da, 56906 Da and 56,791 Da (theoretical calculated molecular weight).

Looking at the amino acid sequence of the complete FU of KLH1-*h* starting from the N-terminus, one site for N-glycosylation is present at position 370 with the sequence NLT. Numerous investigations on the glycosylation status of the different KLH FUs had been performed; Stoeva and coworkers found using Concanavalin (Con A)-binding that FU-1*h* indeed is N-glycosylated (Stoeva *et al.*, 1999). To exclude whether the missing glycosylation of the recombinant KLH1-*h* also plays an essential role in antibody recognition, the native FU was incubated with PNGase F during 72 hours at 37 °C. This glycosidase was able to cleave the carbohydrate side chain bound by N-glycosylation. The subsequent glycan detection test revealed that nearly all carbohydrate side chain bound by N-glycosylation had been removed (Fig. 65B, dot no 2).

The consecutive SDS-PAGE and its corresponding Western blot analysis using the α-KLH1-antibody showed a clear result (Fig. 66A and B): while the native KLH1-*h* was represented by one distinct protein band, the deglycosylated fraction showed two bands: one at the same size and one directly beneath. Considering the five potential cleavage sites within the linker peptide, each protein band can represent an appropriate mixture of five different proteins, all with very close related molecular masses to each other. Thus, the smaller protein will represent the deglycosylated protein due to the missing carbohydrate and has a molecular mass of 57 kDa (average of the five calculated masses mentioned above). The upper protein is the still glycosylated KLH1-*h*. Considering the co-expression of the N-terminal 6xHis-tag and the overlapping sequence to FU-1*g*, the recombinantly expressed FU from *E. coli* shall result in a protein of 62,528 Da (theoretical calculated molecular weight). The SDS-PAGE (Fig. 66) showed that its protein band is located between both observed proteins of the deglycosylated fraction of the native FU. This leads to speculations that the molecular mass of the glycosylated native KLH1-*h* must be located above this 62.5 kDa.

The detection with the antibody showed an interesting result: although the upper protein band was quantitatively lower than the smaller protein, the recognition was much stronger, shown by the intense blue precipitation product. The poor recognition of the smaller protein band was now comparable to the one observed for the recombinantly expressed KLH1-*h*. Although KLH1-*h* is known to have only one N-bound glycosylation, the impact of this glycosylation concerning the antibody recognition is highly significant. In addition to this, antibody hybridisation can also take place on O-linked carbohydrate side chains, but Stoeva *et al*. (1999) showed that this type of glycosidic linkage is only found in KLH2-*c*. Interestingly, the incubation of the native KLH1-*h* in 8 M urea led to an almost complete disappearance of the protein band and therefore no immuno-detection was possible.

This result also explains the two observed protein bands within the subsequent native PAGE and its Western blot (Fig. 67A and B). The strong antibody recognition of the upper protein band as well as the same migration pattern revealed that this protein represented the native and still glycosylated FU, while the protein band beneath it represented the deglycosylated proportion. The migration pattern of the denatured FU as well as the recombinantly expressed KLH1-*h* was completely different: both were represented by diffuse protein bands at the upper edge of the native PAGE and the weak antibody recognition of the denatured originated from the limpet showed again the importance of the

native conformation. Interestingly, the recombinantly expressed KLH1-*h* was not recognized by the α-KLH1-antibody, when applied to a native PAGE.

In conclusion, one can say that both the wrong protein conformation, due to denaturation in 8 M urea, and the missing carbohydrate side chain of the prokaryotic KLH1-*h* lead to a very poor antibody recognition, compared to the native counterpart. The absent glycosylation could in turn additionally prevent the correct protein folding.

7 Can recombinant expression of KLH1-*h* in eukaryotic insect cells lead to better results?

After translation, eukaryotic cells can modify their newly synthesized proteins (post-translational modification) so that disulfide bridges are created, carbohydrate side chain are bound at specific sequence motives, the N-terminal signal peptides (SP) are cut off and the protein is folded correctly with the help of chaperones. Some of these features are only partially or not possible within prokaryotic cells. So, due to its more complex protein expression machinery, recombinant expression of KLH1-*h* in eukaryotic Sf9 insect cells can lead to a correctly folded and fully glycosylated protein. The denaturation step with 8 M urea can additionally be omitted, because nothing is known about inclusion body formation in eukaryotic cells. During the present work, three different samples of insect cell-derived KLH1-*h* were generated by [redacted]: the first one was found to form pelletable insoluble aggregates and the missing signal peptide was thought to be the reason. Each eukaryotic secreted protein, like KLH, is naturally expressed with an N-terminal signal sequence to enable the translocation into the ER and post-translational modification. Furthermore, this sequence contains information about the final destination of the protein. The first KLH1-*h* expressed in insect cells had no such signal peptide, thus a new cDNA was generated by addition of the KLH1-specific signal peptide by extended PCR primers. The resulting clones were expressed in Sf9 and while the most recombinant protein was still located within the insoluble pellet fraction, some recombinant KLH1-*h* protein was now also found within the soluble supernatant.

To enable a comparison to be made with the denatured and unglycosylated *E. coli* derived KLH1-*h* and the equivalently treated native counterpart, the same conditions were also applied to the recombinant insect cell protein.

The following SDS-PAGE allows conclusions to be drawn about the putative glycosylation states of the different recombinantly expressed proteins (Fig. 68). To assist comprehension, Figures 68 is repeated as Figure 82.

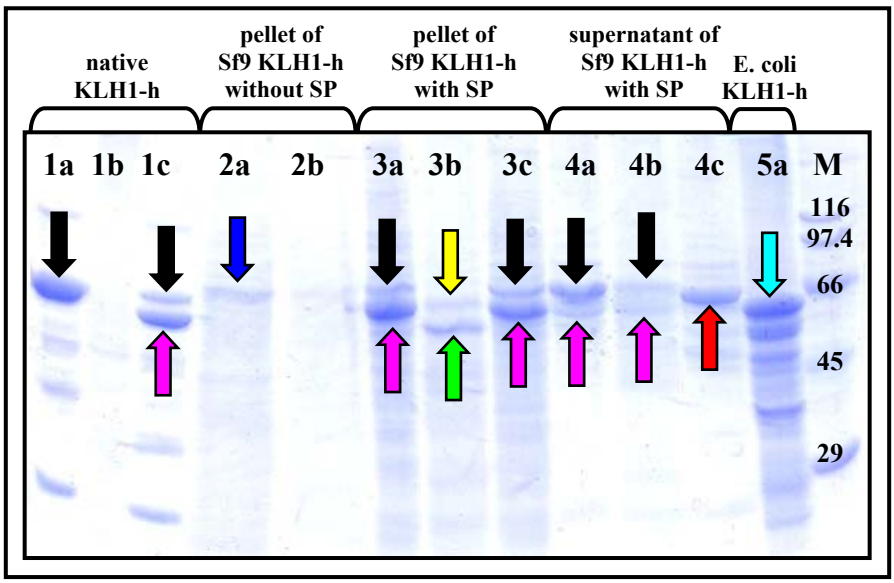


Fig. 82: SDS-PAGE of different KLH1-*h* samples: the native protein, the insect cell-derived (Sf9) KLH1-*h* without signal peptide (SP), the insect cell-derived KLH1-*h* with SP (pellet and supernatant) and KLH1-*h* produced in *E. coli*. The lanes named with "a" represent the respective untreated sample, "b" the denaturated one and "c" the deglycosylated one, respectively. M represents the marker, in kDa.

As already mentioned, the denaturation of native KLH1-*h* (Fig. 82, lane 1b) leads to disappearance of the normally occurring protein (lane 1a, black arrow). The deglycosylation, in contrast, contained a small additional protein of approximately 57 kDa (lane 1c, pink arrow), presumably the deglycosylated FU. The first insect cell-derived KLH1-*h* expressed without signal peptide is represented by one protein band, which, interestingly, is somewhat larger than the native FU (64 kDa) due to the coexpressed N-terminal 6xHis-tag and the overlapping sequence to FU-1*g* (lane 2a, blue arrow). KLH1-*h*, coexpressed with the KLH1-specific signal peptide in Sf9, was expected to contain glycosylated protein, at least within the soluble supernatant. The pellet forming Sf9 fraction of this expression experiment reveal two protein bands (lane 3a). The smaller protein band (lane 3a, orange arrow), quantitatively more present, is assumed to be still insoluble and unglycosylated protein. The calculated theoretical molecular weight of this construct is about 61 kDa which will fit this speculation. One can therefore conclude that the upper protein band (lane 3a, green arrow), which is quantitatively lower, can constitute a small proportion of glycosylated KLH1-*h,* due to the larger mass. The coexpressed signal peptide could have led to the correct insertion of the newly synthesized protein into the ER of the insect cells, resulting in a post-translational glycosylation. While the deglycosylation of this sample led to no change (lane 3c, green and orange arrows), neither in mass nor in quantitative distribution of both observed proteins, the denaturation led to a reduction of the mass of both protein bands (lane 3b, yellow and light green arrows). The su-

pernatant of the SP-KLH1-*h* expression showed a reverse behaviour with regard to the quantitative distribution of the two proteins: the upper protein band (lane 4a, green arrow), which can be assumed to be glycosylated due to mass comparison with the native protein, was greater then the protein beneath (orange arrow), probably representing a small quantity of unglycosylated protein. The denaturation of this sample (lane 4b) did not shift the protein mass, as observed for the pellet-forming proteins (lane 3b), but led only to a reduced amount of the upper protein (lane 4b, green arrow). The deglycosylation, in contrast, produced only one protein band (lane 4c, red arrow). Its size is located between both previously observed proteins (~ 63 kDa) so that conclusions drawn suggest that it consists maybe of a mixture of glycosylated and unglycosylated protein, as well as partially glycosylated intermediates. The *E. coli*-derived KLH1-*h* (lane 5a, turquoise arrow) is definitely smaller than this (62.5 kDa) and can be compared to the lower, unglycosylated protein band in the other samples.

The protein migration within this SDS-PAGE allows conclusions to be drawn indicating that the eukaryotic insect cells are able to produce glycosylated KLH1-*h*, when the signal peptide is coexpressed. The Western blot analysis using the α-KLH1-antibody was expected to show stronger recognition behaviour towards the recombinantly expressed, glycosylated proteins, compared to the unglycosylated ones. The detection, however, showed a different result (Fig. 69, again repeated in Fig. 83).

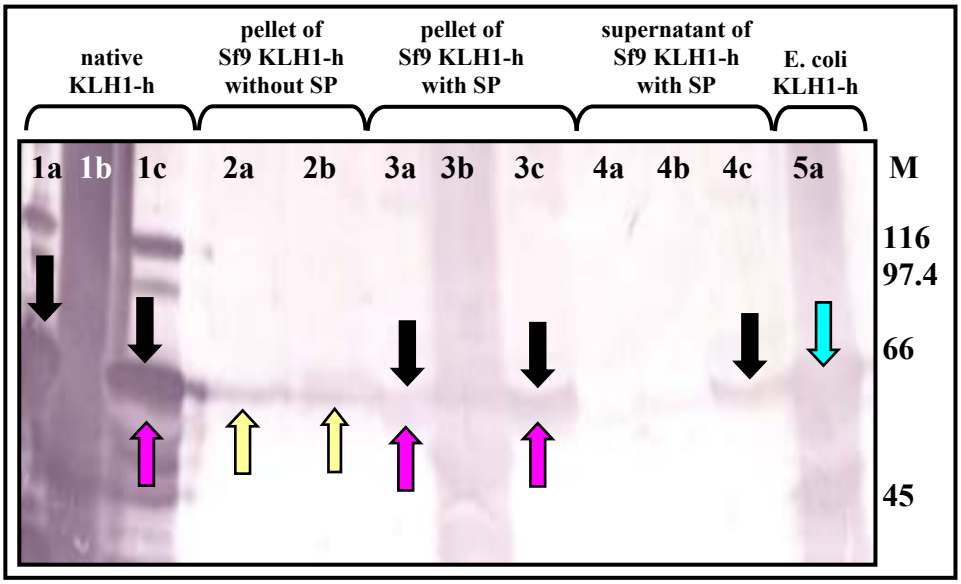


Fig. 83: Western blot analysis of the different KLH1-*h* samples, incubated with an α-KLH1-antibody: the native FU, the insect cell derived (Sf9) FU without signal peptide (SP), the insect cell derived FU with signal peptide (pellet and supernatant) and KLH1-*h*, produced in *E. coli*. The lanes named with "a" represent the respective untreated samples, "b" the denaturated and "c" the deglycosylated one, respectively. M represents the marker, in kDa.

While the native KLH1-*h* (lane 1a, black arrow) and the corresponding protein in its deglycosylated counterpart (lane 1c, black arrow) were strongly recognized, as already mentioned above, all insect cell-derived recombinant proteins were recognized less strongly than the prokaryotic protein. This was observed, irrespective of the origin, from the pellet (lanes 2a and 3a) or the supernatant (lane 4a), or if expressed with (lanes 3a and 4a) or without signal peptide (lane 2a) or if thought to be glycosylated (green arrows) or not (yellow and orange arrows). The weak antibody binding was comparable to the one observed for the deglycosylated, lower protein band of the native sample (lane 2c, pink arrow) and the one observed for the prokaryotic protein (lane 5a, turquoise arrow).

The corresponding native PAGE of these samples and its Western blot showed the same result as the one for the prokaryotic expression: while here only one protein band was apparent for all insect cell-derived recombinant proteins, all having the same migration pattern, but a different one from that of the native FU (Fig. 70), none of these products were recognized by the α-KLH1-antibody (Fig. 71).

So, although the different mass of the protein bands leads to the conclusion that some KLH1-*h*, expressed in insect cells with the signal peptide, can be glycosylated, the incubation with the α-KLH1-antibody shows that the detection is as weak as for the prokaryotic protein. The recombinant expression with the signal peptide can therefore neither guarantee the correct KLH1-*h* processing as in the organism of origin, nor the ability of the insect cells to add the correct carbohydrate side chains. Furthermore, the localisation of the signal peptide at the N-terminus of KLH1-*h*, the last FU within the subunit, is an artificial construct. The absence of the other FUs, KLH1-*a* to KLH1-*g*, could contribute to the correct processing and folding, not only for the whole protein but also for the single FU KLH1-*h*.

In conclusion, one can say that so far, the native KLH1-*h* isolated from *M. crenulata*, is most strongly recognized immunologically.

8 Establishing the method of SOE-PCR for the generation of larger KLH1 substructures and subsequent improvement of the method with KLH2

As already described, SOE-PCR is a PCR technique used for the splicing of two cDNAs or genes by hybridization of corresponding overlapping sequences. This method was used in the present work, starting with cDNAs encoding for single functional units of

KLH1, for the generation of larger KLH substructures. This was made possible by the selection of the gene-specific primers for RT-PCR, each binding to the previous or following FU of the one investigated. So, overlapping sequences were created from each FU to the surrounding ones; these were already available at the beginning of this work. Consequently, the generation of cDNAs encompassing two (KLH1-*ab*, KLH1-*cd*), three (KLH1-*abc*) and four consecutive functional units (KLH1-*abcd* and KLH1-*efgh*) was achieved.

Regarding the potential recombinant expression capacity of *E. coli*, it is obvious that proteins bigger than 150 kDa probably will not be expressed in bacterial cells. Also, due to the missing glycosylation in a prokaryotic expression system, the focus of future recombinant expression of KLH will be the improvement of recombinant expression with an appropriate eukaryotic system. Thus, the generation and cloning of even larger KLH1 substructures for future expression in eukaryotic cells was tried and achieved for two other cDNAs, containing six FUs: KLH1-*abcdef* and KLH1-*cdefgh*. In the time available, the method of SOE-PCR was optimized and the best cycling parameters were defined, according to the length of the expected cDNA. The final success was the realization of a cDNA encompassing all eight FUs *a* to *h* of KLH1, correctly by two approaches: the splicing of KLH1-*abcdef* with KLH1-*fgh* and of KLH1-*abcde* with KLH1-*efgh*, creating the whole subunit KLH1-*abcdefgh*. Anticipating the future expression experiment in eukaryotic cells, this was directly designed having the KLH1-specific signal peptide at the 5′ end, thus enabling post-translational modification by successful translocation into the ER, and introduction of a 6xHis-tag at the 3′ end, for purification using affinity chromatography. During these studies, I found out that the length of the overlapping sequence played a critical role. The length of the respective sequences between two FUs of KLH1 was around 100 bp (from 72 to 126 bp), due to the location of the chosen gene-specific primers. This distance was found to be optimal, irrespective of the proper length of the spliced cDNAs. Shorter distances were not tried, but for longer ones, even up to a complete FU (around 1200 bp), the hybridization worked well. Beyond this size, the splicing efficiency decreased.

After the optimal conditions had been established, several approaches were tried to further improve the method. Although a proofreading polymerase was used at every PCR step, repeated amplification of the cDNAs led to several sequence mutations within the different clones, the worst case leading to nucleotide deletion, insertion or mutation, resulting in a shift of the open reading frame or in a triplet coding for translation termi-

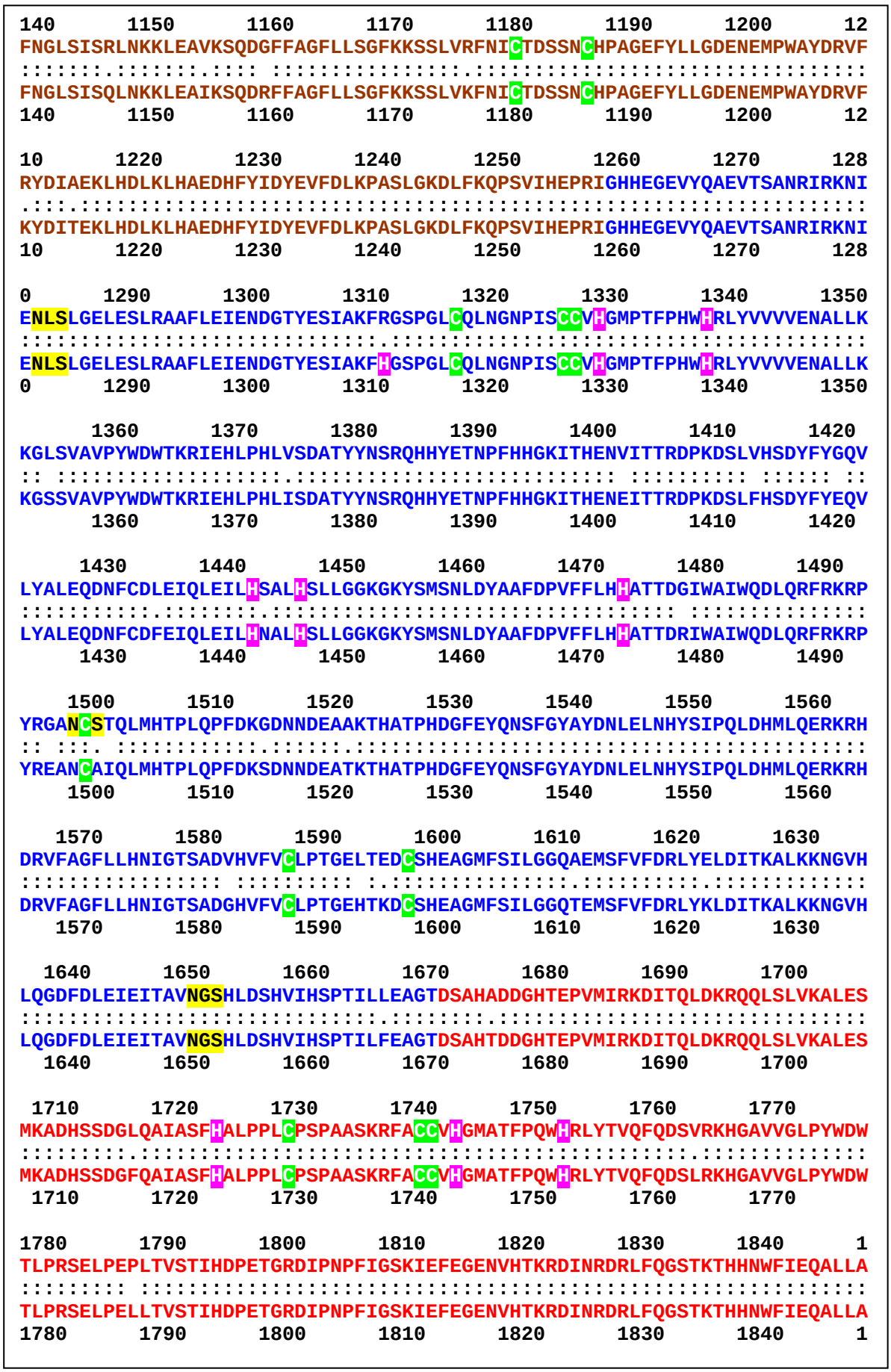
nation (stop). Numerous PCR cycles and extended elongation durations, especially for long templates, as well as inserted cloning steps of each newly generated cDNA in between for stabilization and amplification of the constructs had to be avoided. New protocols were elaborated and found to be as successful as their corresponding longer versions. At the beginning, the fusion of three consecutive cDNAs of KLH1 was accomplished by passing through several steps: a first SOE-PCR for the splicing of the first two cDNAs followed by a subsequent cloning step, a PCR based amplification of this new cDNA from the vector and a second SOE-PCR for the splicing of the third cDNA to the previously generated one, again followed by cloning. The first amelioration was done by mixing first two consecutive cDNAs, followed by 10 cycles to allow hybridization of the complementary sequences and elongation of the newly spliced cDNA. Then the third cDNA was added directly to this mixture, again followed by 10 cycles, and then the outer primer pair was applied for 20 further cycles for amplification of the whole construct. This changed protocol resulted in the cDNAs KLH2-*abc* and KLH2-*efg* (Fig. 76). The previously executed second SOE-PCR as well as the inserted cloning step were therefore avoided and led to a considerable time saving. But, overall, the whole cycling duration remained very long. Additionally, one has to be personally present at exactly the correct times to sequentially add the required components. So, different variations were tested up to the mixing of four consecutive FUs at the same time in one tube, giving them 10 cycles for correct annealing to each other, followed by the addition of the outer primer pair for amplification during 20 further cycles. Thus, this SOE-PCR lasts no longer than a normal PCR having 30 cycles and led to the successful splicing of four consecutive cDNAs: KLH2-*abcd* (Fig. 77). Other investigations within our working group have shown that even the previous amplification of cloned cDNAs from their respective vector can be skipped: the plasmids can be taken directly as templates for SOE-PCR, so that polymerase-related mutations can in theory be avoided. So, bright prospects are indicated regarding the decreasing error rate for the generation of a correct cDNA encompassing the complete subunit of KLH2, due to the improvement of the method of SOE-PCR achieved during this work.

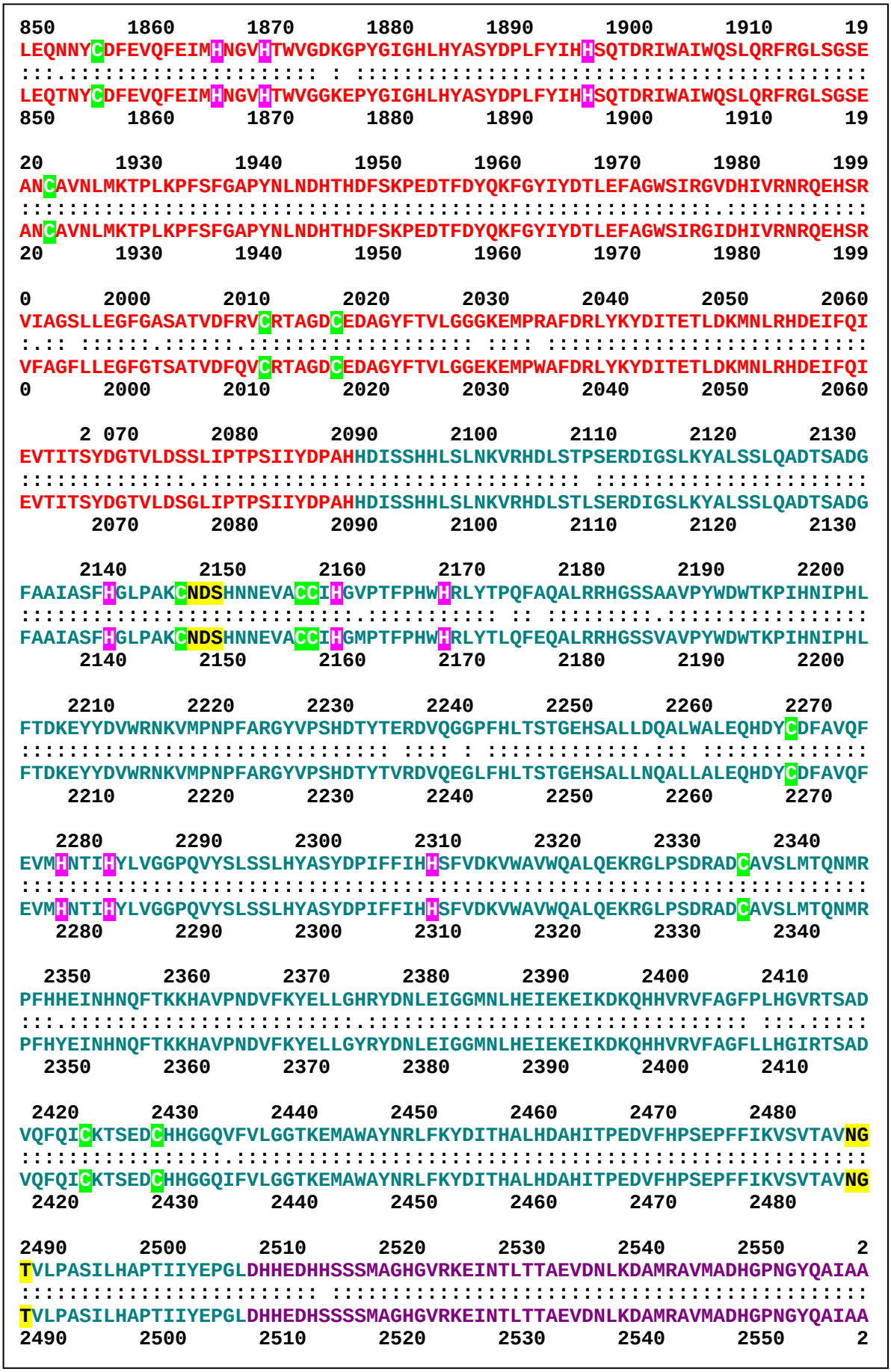
9 Amino acid sequence analyses of the clone KLH1-*abcdefgh* compared to the original sequence

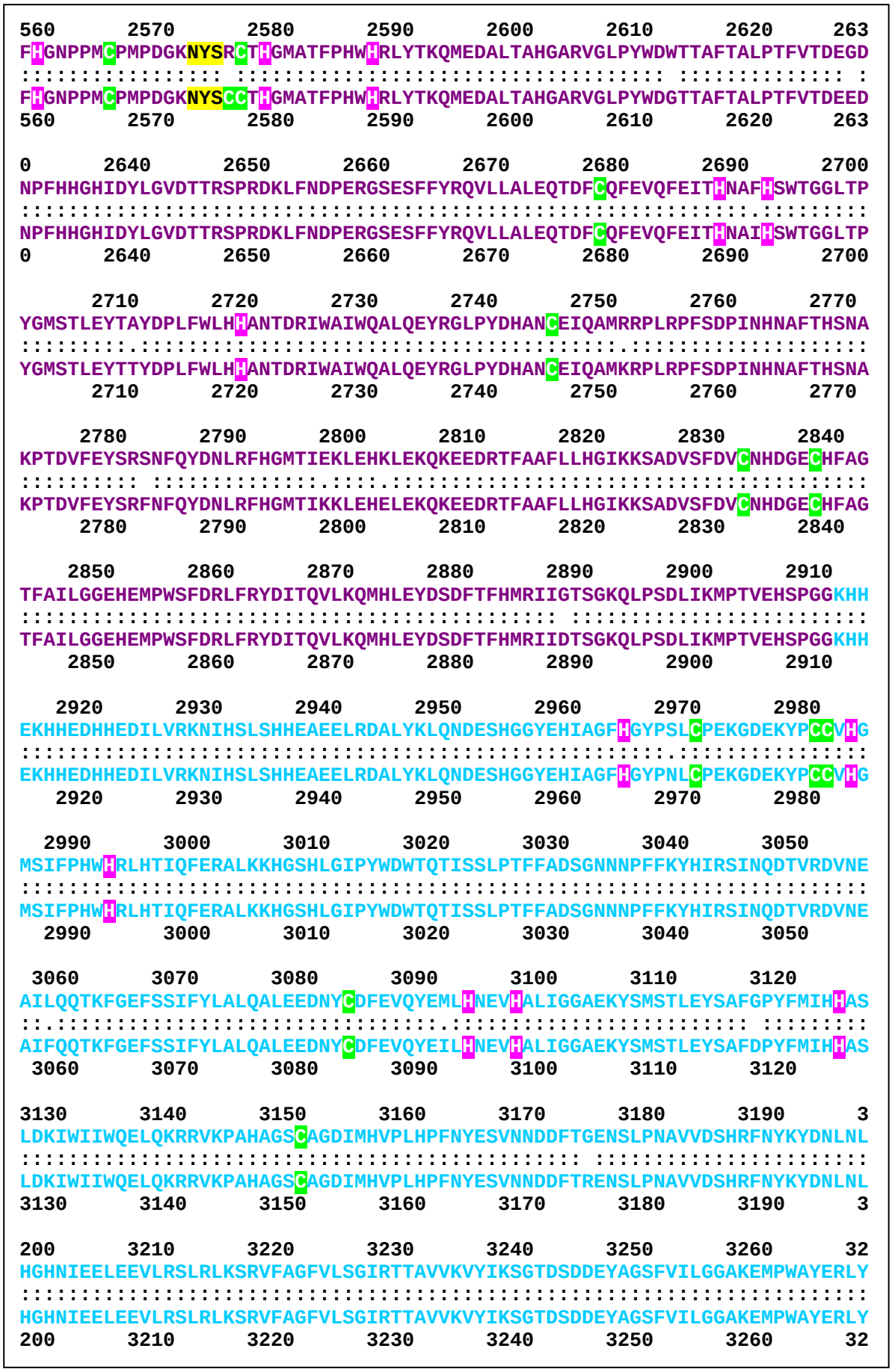
The clone encoding for the complete subunit KLH1-*abcdefgh*, generated during the present work, is so far the most important clone for future expression experiments in eukaryotic cells. Therefore, the analysis of the sequence is important to check the putative error rate and for any introduced mistakes. The sequencing of the clone showed first, that no stop codon had been inserted by polymerase-related nucleotide insertion, deletion or mutation. The alignment of its translated amino acid sequence with the original sequence of KLH1, determined within our working group (unpublished) provided detailed insights (Fig. 84, aligned with: http://xylian.igh.cnrs.fr/bin/align-guess.cgi). The upper line represents the recombinant KLH1 sequence, the line beneath the original KLH1 sequence.

```
        10        20        30        40        50        60        70
MLSVRLLIVVLALANAENLVRKSVEHLTQEETLDLQAALRELQMDSSSIGFQKIAAAHGAPASCVHKDTSI
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
MLSVRLLIVVLALANAENLVRKSVEHLTQEETLDLQAALRELQMDSSSIGFQKIAAAHGAPASCVHKDTSI
        10        20        30        40        50        60        70

       80        90       100       110       120       130       140
ACCIHGMPTFPHWHRAYVVHMERALQTKRRTSGMPYWDWTEPITQLPSLAADPVYTDSQGGKAHTNHWYRG
:::::::::::::::::::::::::::::::::::::::::::::::::::: :::::::::.::::::::
ACCIHGMPTFPHWHRAYVVHMERALQTKRRTSGMPYWDWTEPITQLPSLAADPVYIDSQGGKAHTNYWYRG
       80        90       100       110       120       130       140

    150       160       170       180       190       200       210
NIDFLDKKTNRAVDDRLFGKVKPGQHTHLMESVLGALEPDEFCKFEIQFELAHNAIHYLVGGKHDYSMANL
::::::::::::::::: :::::::::::::: ::: :::::::::::::::::::::::::::::::::
NIDFLDKKTNRAVDDRLFEKVKPGQHTHLMESVLDALEQDEFCKFEIQFELAHNAIHYLVGGKHDYSMANL
    150       160       170       180       190       200       210

   220       230       240       250       260       270       280
EYTAYDPIFFLHHSNVDRIFAIWQRLQELRNKGPKAMDCAQELLHRKMGPFSWEDNDIPLTNEHSTPADLL
:::::::::::::::::::::::::::::::: ::::::::::::.:: ::::::::::::::::::::.
EYTAYDPIFFLHHSNVDRIFAIWQRLQELRNKDPKAMDCAQELLHQKMEPFSWEDNDIPLTNEHSTPADLF
   220       230       240       250       260       270       280

  290       300       310       320       330       340       350
DYRQLHYDYDTLNLSGMTPEELKAYLDERSSRARAFASFRLKGFGGSTNVFVYVRIPDDNDRNDDHWEKAG
:::::::::::::.:::::::.:::::::::::::::::::::::.:::::: :::::::::::: ::::
DYRQLHYDYDTLNLNGMTPEELKTYLDERSSRARAFASFRLKGFGGSANVFVYVCIPDDNDRNDDHCEKAG
  290       300       310       320       330       340       350

 360       370       380       390       400       410       420
DFFVLGGPREMKWQFYRPYLFDLSDTVHKMGMKLDGHYTVKAELFSVNGTALPDDLLPHPVVVHHPEKGFT
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DFFVLGGPSEMKWQFYRPYLFDLSDTVHKMGMKLDGHYTVKAELFSVNGTALPDDLLPHPVVVHHPEKGFT
 360       370       380       390       400       410       420
```

```
430       440       450       460       470       480       490       5
DPPVKHHQSANLLVRKIINDPTREEVLNLREAFRKFQEDRSVDGYQATAEYHGLPARCPRPDAKDRYACCV
:::::::::::::::: ::: ::::::::::::.:::::::::::::::::::::::::::::::::::::
DPPVKHHQSANLLVRKNINDLTREEVLNLREAFHKFQEDRSVDGYQATAEYHGLPARCPRPDAKDRYACCV
430       440       450       460       470       480       490       5

00       510       520       530       540       550       560       57
HGMPIFPHWYRLFVTQVEDALVGRGATIGIPYWDWAEPMTHIPGLAGNKTYVDSRGASHTNPFHSSVIAFE
:::::::::.:::::::::::::::::::::::::::::::.:::::.:::::::::::::::::::::::
HGMPIFPHWHRLFVTQVEDALVGRGATIGIPYWDWTEPMTHIPGLAGNKTYVDSHGASHTNPFHSSVIAFE
00       510       520       530       540       550       560       57

0       580       590       600       610       620       630       640
ENAPHTKRQTDQRLFKPATFGHHTDLLNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYA
::::::::: ::::::::::::::::.:::::::::::::::::::::::::::::::::::::::::::.
ENAPHTKRQIDQRLFKPATFGHHTDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSMSSLHYT
0       580       590       600       610       620       630       640

        650       660       670       680       690       700       710
TFDPLFYFRHSNVDRLWAVWQALQMRRHKPYRAHYAISLEHMHLKPFAFSSPLNNNEKTHANAMPNKIYDY
.::::::::::::::::::::::::::: ::::::::::::::::::::::::::::::::::::::::::
AFDPLFYFHHSNVDRLWAVWQALQMRRHKPYRAHCAISLEHMHLKPFAFSSPLNNNEKTHANAMPNKIYDY
        650       660       670       680       690       700       710


      720       730       740       750       760       770       780
ENVLHYTHEDSTFGGISLENIKKMVHENQQEDGIYAGFLLAGIRTSANVDIFIKTTGSVQHKAETLAVLGG
:::::::.:::::::::::::.::.:::::::: :::::::::::::::::::::::: :::::: :.::::
ENVLHYTYEDLTFGGISLENIEKMIHENQQEDRIYAGFLLAGIRTSANVDIFIKTTDSVQHKAGTFAVLGG
      720       730       740       750       760       770       780

     790       800       810       820       830       840       850
SEEMEWGFDRVLKFDITHALKDLDLTADGDFEVTVDITEVDGTKLASSLIPHASVIREHARVKFDKVPGSR
:.::.:::::::.:::::::.:::::::::::::::::::::::::::::::::::::::::::::::: ::
SKEMKWGFDRVFKFDITHVLKDLDLTADGDFEVTVDITEVDGTKLASSLIPHASVIREHARVKFDKVPRSR
     790       800       810       820       830       840       850

    860       870       880       890       900       910       920
LIRKNVDRLGPGEMNELRKALALLKEDKSAGGFQQLGAFHGEPRWCPSPEASKKFACCVHGMSVFPHWHRL
::::::::::: ::::::::::::::::::::::::::::::.:::::::::::::::::::::::::::::
LIRKNVDRLSPEEMNELRKALALLKEDKSAGGFQQLGAFHGEPKWCPSPEASKKFACCVHGMSVFPHWHRL
    860       870       880       890       900       910       920

   930       940       950       960       970       980       990
LTVQSGNALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNPWFDGHIDTVDKTTTRSVQNK
::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
LTVQSENALRRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNPWFDGHIDTVDKTTTRSVQNK
   930       940       950       960       970       980       990

 1000      1010      1020      1030      1040      1050      1060
LFEQPEFGHNTSIAKQVLLALELDNFCDFEIQYEIAHNYIHALVGGAQPYGMASFRYTASDPLSYLHHSNT
::::::::: :::::::::::: :::::::::::::::::::::::::::::.:::: ::: :::::::::
LFEQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGGAQPYGMASLRYTAFDPLFYLHHSNT
 1000      1010      1020      1030      1040      1050      1060

1070      1080      1090      1100      1110      1120      1130      1
DRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANINTDYVTKEHSVPFNVFDYKTNFNYEYDTLE
:::::::::::::::::::::::::::::::::::::::::.:::::::::::::::::::::::::::::
DRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANINTDHVTKEHSVPFNVFDYKTNFNYEYDTLE
1070      1080      1090      1100      1110      1120      1130      1
```


140 1150 1160 1170 1180 1190 1200 12
FNGLSISRLNKKLEAVKSQDGFFAGFLLSGFKKSSLVRFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVF
FNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSLVKFNICTDSSNCHPAGEFYLLGDENEMPWAYDRVF
140 1150 1160 1170 1180 1190 1200 12
10 1220 1230 1240 1250 1260 1270 128
RYDIAEKLHDLKLHAEDHFYIDYEVFDLKPASLGKDLFKQPSVIHEPRIGHHEGEVYQAEVTSANRIRKNI
KYDITEKLHDLKLHAEDHFYIDYEVFDLKPASLGKDLFKQPSVIHEPRIGHHEGEVYQAEVTSANRIRKNI
10 1220 1230 1240 1250 1260 1270 128
0 1290 1300 1310 1320 1330 1340 1350
ENLSLGELESLRAAFLEIENDGTYESIAKFRGSPGLCQLNGNPISCCVHGMPTFPHWHRLYVVVVENALLK
ENLSLGELESLRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPISCCVHGMPTFPHWHRLYVVVVENALLK
0 1290 1300 1310 1320 1330 1340 1350
1360 1370 1380 1390 1400 1410 1420
KGLSVAVPYWDWTKRIEHLPHLVSDATYYNSRQHHYETNPFHHGKITHENVITTRDPKDSLVHSDYFYGQV
KGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQHHYETNPFHHGKITHENEITTRDPKDSLFHSDYFYEQV
1360 1370 1380 1390 1400 1410 1420
1430 1440 1450 1460 1470 1480 1490
LYALEQDNFCDLEIQLEILHSALHSLLGGKGKYSMSNLDYAAFDPVFFLHHATTDGIWAIWQDLQRFRKRP
LYALEQDNFCDFEIQLEILHNALHSLLGGKGKYSMSNLDYAAFDPVFFLHHATTDRIWAIWQDLQRFRKRP
1430 1440 1450 1460 1470 1480 1490
1500 1510 1520 1530 1540 1550 1560
YRGANCSTQLMHTPLQPFDKGDNNDEAAKTHATPHDGFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRH
YREANCAIQLMHTPLQPFDKSDNNDEATKTHATPHDGFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRH
1500 1510 1520 1530 1540 1550 1560
1570 1580 1590 1600 1610 1620 1630
DRVFAGFLLHNIGTSADVHVFVCLPTGELTEDCSHEAGMFSILGGQAEMSFVFDRLYELDITKALKKNGVH
DRVFAGFLLHNIGTSADGHVFVCLPTGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVH
1570 1580 1590 1600 1610 1620 1630
1640 1650 1660 1670 1680 1690 1700
LQGDFDLEIEITAVNGSHLDSHVIHSPTILLEAGTDSAHADDGHTEPVMIRKDITQLDKRQQLSLVKALES
LQGDFDLEIEITAVNGSHLDSHVIHSPTILFEAGTDSAHTDDGHTEPVMIRKDITQLDKRQQLSLVKALES
1640 1650 1660 1670 1680 1690 1700
1710 1720 1730 1740 1750 1760 1770
MKADHSSDGLQAIASFHALPPLCPSPAASKRFACCVHGMATFPQWHRLYTVQFQDSVRKHGAVVGLPYWDW
MKADHSSDGFQAIASFHALPPLCPSPAASKRFACCVHGMATFPQWHRLYTVQFQDSLRKHGAVVGLPYWDW
1710 1720 1730 1740 1750 1760 1770
1780 1790 1800 1810 1820 1830 1840 1
TLPRSELPEPLTVSTIHDPETGRDIPNPFIGSKIEFEGENVHTKRDINRDRLFQGSTKTHHNWFIEQALLA
TLPRSELPELLTVSTIHDPETGRDIPNPFIGSKIEFEGENVHTKRDINRDRLFQGSTKTHHNWFIEQALLA
1780 1790 1800 1810 1820 1830 1840 1


850 1860 1870 1880 1890 1900 1910 19
LEQNNYCDFEVQFEIMHNGVHTWVGDKGPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSE
LEQTNYCDFEVQFEIMHNGVHTWVGGKEPYGIGHLHYASYDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSE
850 1860 1870 1880 1890 1900 1910 19
20 1930 1940 1950 1960 1970 1980 199
ANCAVNLMKTPLKPFSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGVDHIVRNRQEHSR
ANCAVNLMKTPLKPFSFGAPYNLNDHTHDFSKPEDTFDYQKFGYIYDTLEFAGWSIRGIDHIVRNRQEHSR
20 1930 1940 1950 1960 1970 1980 199
0 2000 2010 2020 2030 2040 2050 2060
VIAGSLLEGFGASATVDFRVCRTAGDCEDAGYFTVLGGGKEMPRAFDRLYKYDITETLDKMNLRHDEIFQI
VFAGFLLEGFGTSATVDFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQI
0 2000 2010 2020 2030 2040 2050 2060
2 070 2080 2090 2100 2110 2120 2130
EVTITSYDGTVLDSSLIPTPSIIYDPAHHDISSHHLSLNKVRHDLSTPSERDIGSLKYALSSLQADTSADG
EVTITSYDGTVLDSGLIPTPSIIYDPAHHDISSHHLSLNKVRHDLSTLSERDIGSLKYALSSLQADTSADG
2070 2080 2090 2100 2110 2120 2130
2140 2150 2160 2170 2180 2190 2200
FAAIASFHGLPAKCNDSHNNEVACCIHGVPTFPHWHRLYTPQFAQALRRHGSSAAVPYWDWTKPIHNIPHL
FAAIASFHGLPAKCNDSHNNEVACCIHGMPTFPHWHRLYTLQFEQALRRHGSSVAVPYWDWTKPIHNIPHL
2140 2150 2160 2170 2180 2190 2200
2210 2220 2230 2240 2250 2260 2270
FTDKEYYDVWRNKVMPNPFARGYVPSHDTYTERDVQGGPFHLTSTGEHSALLDQALWALEQHDYCDFAVQF
FTDKEYYDVWRNKVMPNPFARGYVPSHDTYTVRDVQEGLFHLTSTGEHSALLNQALLALEQHDYCDFAVQF
2210 2220 2230 2240 2250 2260 2270
2280 2290 2300 2310 2320 2330 2340
EVMHNTIHYLVGGPQVYSLSSLHYASYDPIFFIHHSFVDKVWAVWQALQEKRGLPSDRADCAVSLMTQNMR
EVMHNTIHYLVGGPQVYSLSSLHYASYDPIFFIHHSFVDKVWAVWQALQEKRGLPSDRADCAVSLMTQNMR
2280 2290 2300 2310 2320 2330 2340
2350 2360 2370 2380 2390 2400 2410
PFHHEINHNQFTKKHAVPNDVFKYELLGHRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFPLHGVRTSAD
PFHYEINHNQFTKKHAVPNDVFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSAD
2350 2360 2370 2380 2390 2400 2410
2420 2430 2440 2450 2460 2470 2480
VQFQICKTSEDCHHGGQVFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAVNG
VQFQICKTSEDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAVNG
2420 2430 2440 2450 2460 2470 2480
2490 2500 2510 2520 2530 2540 2550 2
TVLPASILHAPTIIYEPGLDHHEDHHSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVMADHGPNGYQAIAA
TVLPASILHAPTIIYEPGLDHHEDHSSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVMADHGPNGYQAIAA
2490 2500 2510 2520 2530 2540 2550 2


560 2570 2580 2590 2600 2610 2620 263
FHGNPPMCPMPDGKNYSRCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDWTTAFTALPTFVTDEGD
FHGNPPMCPMPDGKNYSCCTHGMATFPHWHRLYTKQMEDALTAHGARVGLPYWDGTTAFTALPTFVTDEED
560 2570 2580 2590 2600 2610 2620 263
0 2640 2650 2660 2670 2680 2690 2700
NPFHHGHIDYLGVDTTRSPRDKLFNDPERGSESFFYRQVLLALEQTDFCQFEVQFEITHNAFHSWTGGLTP
NPFHHGHIDYLGVDTTRSPRDKLFNDPERGSESFFYRQVLLALEQTDFCQFEVQFEITHNAIHSWTGGLTP
0 2640 2650 2660 2670 2680 2690 2700
2710 2720 2730 2740 2750 2760 2770
YGMSTLEYTAYDPLFWLHHANTDRIWAIWQALQEYRGLPYDHANCEIQAMRRPLRPFSDPINHNAFTHSNA
YGMSTLEYTTYDPLFWLHHANTDRIWAIWQALQEYRGLPYDHANCEIQAMKRPLRPFSDPINHNAFTHSNA
2710 2720 2730 2740 2750 2760 2770
2780 2790 2800 2810 2820 2830 2840
KPTDVFEYSRSNFQYDNLRFHGMTIEKLEHKLEKQKEEDRTFAAFLLHGIKKSADVSFDVCNHDGECHFAG
KPTDVFEYSRFNFQYDNLRFHGMTIKKLEHELEKQKEEDRTFAAFLLHGIKKSADVSFDVCNHDGECHFAG
2780 2790 2800 2810 2820 2830 2840
2850 2860 2870 2880 2890 2900 2910
TFAILGGEHEMPWSFDRLFRYDITQVLKQMHLEYDSDFTFHMRIIGTSGKQLPSDLIKMPTVEHSPGGKHH
TFAILGGEHEMPWSFDRLFRYDITQVLKQMHLEYDSDFTFHMRIIDTSGKQLPSDLIKMPTVEHSPGGKHH
2850 2860 2870 2880 2890 2900 2910
2920 2930 2940 2950 2960 2970 2980
EKHHEDHHEDILVRKNIHSLSHHEAEELRDALYKLQNDESHGGYEHIAGFHGYPSLCPEKGDEKYPCCVHG
EKHHEDHHEDILVRKNIHSLSHHEAEELRDALYKLQNDESHGGYEHIAGFHGYPNLCPEKGDEKYPCCVHG
2920 2930 2940 2950 2960 2970 2980
2990 3000 3010 3020 3030 3040 3050
MSIFPHWHRLHTIQFERALKKHGSHLGIPYWDWTQTISSLPTFFADSGNNNPFFKYHIRSINQDTVRDVNE
MSIFPHWHRLHTIQFERALKKHGSHLGIPYWDWTQTISSLPTFFADSGNNNPFFKYHIRSINQDTVRDVNE
2990 3000 3010 3020 3030 3040 3050
3060 3070 3080 3090 3100 3110 3120
AILQQTKFGEFSSIFYLALQALEEDNYCDFEVQYEMLHNEVHALIGGAEKYSMSTLEYSAFGPYFMIHHAS
AIFQQTKFGEFSSIFYLALQALEEDNYCDFEVQYEILHNEVHALIGGAEKYSMSTLEYSAFDPYFMIHHAS
3060 3070 3080 3090 3100 3110 3120
3130 3140 3150 3160 3170 3180 3190 3
LDKIWIIWQELQKRRVKPAHAGSCAGDIMHVPLHPFNYESVNNDDFTGENSLPNAVVDSHRFNYKYDNLNL
LDKIWIIWQELQKRRVKPAHAGSCAGDIMHVPLHPFNYESVNNDDFTRENSLPNAVVDSHRFNYKYDNLNL
3130 3140 3150 3160 3170 3180 3190 3
200 3210 3220 3230 3240 3250 3260 32
HGHNIEELEEVLRSLRLKSRVFAGFVLSGIRTTAVVKVYIKSGTDSDDEYAGSFVILGGAKEMPWAYERLY
HGHNIEELEEVLRSLRLKSRVFAGFVLSGIRTTAVVKVYIKSGTDSDDEYAGSFVILGGAKEMPWAYERLY
200 3210 3220 3230 3240 3250 3260 32

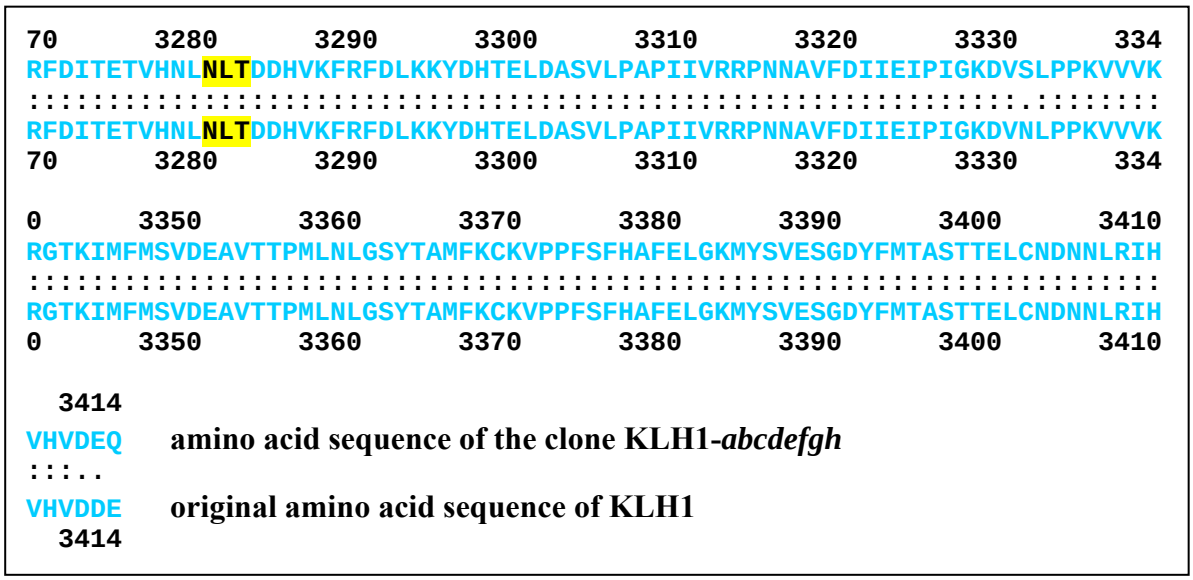


Fig. 84: Absolute amino acid sequence alignment of the clone KLH1-*abcdefgh* (upper line) with the original amino acid sequence of KLH1 (lower line).
(:): Identity of amino acids at this position. (.): Different amino acid but having the same chemical properties. No connection: Introduction of a different amino acid having different chemical properties. Signal peptide: black; FU-1*a*: orange; FU-1*b*: green; FU-1*c*: brown; FU-1*d*: blue; FU-1*e*: red; FU-1*f*: seagreen; FU-1*g*: purple; FU-1*h*: turquoise). Histidines (H) involved in CuA and CuB are marked in pink, Cysteines (C) involved in disulfide bridges or thioether bonds in green, and putative N-glycosylation sites following the sequence motives NXT or NXS are highlighted in yellow.

9.1 Analysis of amino acid exchanges

The sequence alignment shown in Figure 84 reveals that the clone KLH1-*abcdefgh*, generated during the present work, has a sequence identity of 96.3 % compared with the original KLH1 sequence, determined within our working group (unpublished). Identical amino acids at the same position in both sequences are connected by two points (:). For substituted amino acids, it has to be decided whether the newly mutated amino acid has the same chemical properties to the original one (and can therefore potentially perform the same task within the sequence). If this is the case, both amino acids are connected by one point (.). If the substituted amino acid possesses completely different chemical properties to the original one, the amino acids are not connected between both lines. Looking closer to the alignment, one can see that a total of 125 amino acid exchanges have taken place during the multiple cDNA amplification and cloning steps. Thus, the eight FUs have different proportions of exchanges (exchanges by an amino acid having the same chemical properties : exchanges by amino acids not having the same chemical properties): none for the signal peptide, FU-1*a* contains 14 different amino acid exchanges (6 : 8), FU-1*b* has 24 exchanges (17 : 7), in FU-1*c*, 17 exchanges have taken place (9 : 8), FU-1*d* exhibits 20 substitutions (11 : 9), FU-1*e* and FU-1*f* each have 15

changes (9 : 6 and 7 : 8, respectively), in FU-1*g*, a total of 11 exchanges are found (5 : 6) and FU-1*h*, although the longest FU polypeptide, has the fewest changes, namely 9 (6 : 3). Thus, substitutions by amino acids having the same chemical properties prevail in five of eight cases, representing a total of 56 % of all exchanges.

9.2 The copper binding sites

The active sites of the KLH FUs are characterized by six histidines (H), between which two copper atoms are coordinated for the reversible oxygen binding, and which create the copper binding sites CuA and CuB. These motives are found in all FUs, and, as seen in other molluscan hemocyanins, these sequences are highly conserved. Comparing the sequence of the clone KLH1-*abcdefgh* with the original KLH1 sequence (Fig. 84), it emerges that every one of the six histidines is present in each FU (highlighted in pink), except the third histidine in FU-1*b* (starting from the N-terminus: position 508, in CuA) and the first one in FU-1*d* (position 1310, in CuA). The histidine in FU-1*b* is substituted by tyrosine (Y), which can be due to simple mutation of the first nucleotide of the coding triplet (His: CAC; Tyr: TAC). The histidine of FU-1*d* is replaced by arginine (R), which requires two nucleotide exchanges within the same triplet (His: CAT, Arg: CGC). In FU1-*b*, a histidine is present only at two positions in the N-terminal direction. The close location leads to the speculation that this histidine could replace the missing one. Arginine, in contrast, found in FU-1*d*, contains like histidine a nitrogen atom which can perhaps also perform the task of copper coordination. The presence of all copper binding sites within the recombinant clone allows one to draw conclusions about the future biological activity of the recombinant KLH1 subunit. Knowing that oxygen binding in the KLH molecule is subjected to cooperative effects between the FUs (Gielens *et al*., 1997), one can speculate about the correct oxygen binding, when two copper binding sites are not correctly coordinated. The copper atom could, nevertheless, perhaps be coordinated by only two histidines. As a consequence, high resolution analysis of the protein structure of both the fully oxy- and desoxygenated states of KLH are of great interest. Even an only partially oxygenated recombinant molecule might bring new insights regarding oxygenation-depending folding intermediates. Additionally, it would be interesting to determine whether the presence or absence of biological activity plays a role in the immunogenicity of the recombinant KLH1 molecule, which is unlikely.

9.3 Disulfide and thioether bridges

In the original KLH1 sequence, three disulfide bridges, each formed from two cysteines (C) (Fig. 84, highlighted in green) are present in each FU (except FU-1*b* and FU-1*h* which have only two) as well as one thioether bridge, formed between a cysteine (C) and a histidine (H). The sequence alignment of the KLH1-*abcdefgh* clone with the original KLH1 sequence (Fig. 84) shows that all cysteines and histidines required for correct thioether bridge binding are present. However, in FU-1*a* the third disulfide bridge is missing. Both cysteines had been substituted by an arginine (R, position 340) and a tryptophane (W, position 352), respectively. In FU-1*b*, the second disulfide bridge will not be able to form, because the second cysteine is mutated to tyrosine (Y, position 675). Additionally, FU-1*g* will be deficient in its first disulfide bridge due to the second cysteine mutated to arginine (R, position 2575). All changed amino acids do not have the same chemical properties than the original cysteines, but their respective coding triplets differ only in one nucleotide from the codon of cysteine (Cys: TGT and TGC; Arg: CGT and CGC; Tyr: TAT and TAC; Trp: TGG). A mutation was therefore easily introduced by exchanging only one nucleotide at each respective position. The remaining other functional units (FU-1*c*, FU-1*d*, FU-1*e*, FU-1*f* and FU-1*h*) all have the required cysteines for correct disulfide bridge binding. The first and second disulfide bridge bond of each FU are located around the copper binding sites and it is assumed that they play an important role in the correct folding of the protein, as well as for correct positioning and stabilization of the three histidines for the coordinated binding of the copper atoms. This can again contribute to the future biological activity of the recombinantly expressed KLH1 subunit, taking into account the previously mentioned points concerning the missing histidines in FU-1*b* and FU-1*d*. Furthermore, due to two additional cysteines at the end of FU-*h* found in several gastropodan hemocyanins, it is assumed that an additional disulfide bridge can bind at this position. These cysteines can also be found within the sequence of the clone KLH1-*abcdefgh*.

9.4 Potential N-glycosylation sites

Within the original amino acid sequence of KLH1, potential N-glycosylation sites are defined by the sequence motives NXT and NXS. Carbohydrate side chain can bind to these asparagines (N) by N-linkage (Stoeva *et al*., 1995; Cuff *et al*., 1998; Keller *et al*., 1999). Due to the known importance of these carbohydrate side chains for the immunogenic effects of KLH, the analysis of the presence of these N-glycosylation sites within

the KLH1-*abcdefgh* encoding clone will give interesting insights concerning the future potential and likely immuno-stimulating potential. The original KLH1 sequence reveals that a total of eight such N-glycosylation sites can be found (Fig. 84, highlighted in yellow). While FU-1*c* and FU-1*e* have none, the FUs -1*a*, -1*b*, -1*g* and -1*h* each show one glycosylation motif (NGT, NKT, NYS and NLT, respectively). FU-1*d* and FU-1*f* reveal two potential N-glycosylation sites (NLS and NGS for FU-1*d*; NDS and NGT for FU-1*f*). Each of these previously mentioned N-glycosylation sites can also be found within the sequence of the cloned cDNA encoding for the complete subunit. Additionally, two more N-glycosylation sites are now found in the KLH1-*abcdefgh* encoding clone (FU-1*a*: NLS, position 298; and FU-1*c*: NTS, position 1005). Although Stoeva *et al*., (1999) showed that an N-glycosylation site does not guarantee that an effective carbohydrate side chain is formed, the fact that all naturally occurring sites are present in the sequence of the clone KLH1-*abcdefgh* ensures that these sites can potentially be glycosylated during recombinant expression.

10 Future prospects

In general, regarding the recombinant protein expression in *E. coli*, although not applicable for further recombinant expression of KLH due its large size and the absence of the essential glycosylations, several future experiments can be done to improve this method. The main goal must be the circumvention of inclusion body formation, including avoidance of the denaturation step using 8 M urea, for the recovery of a native protein. A primary change has to be applied to the expression temperature: hydrophobic interactions decrease with lowered temperature and the proteins have more time to fold correctly. Additionally, it was shown that the fusion of thioredoxin to several proteins, previously forming inclusion bodies when expressed in *E. coli*, could then be synthesised as soluble proteins that are biologically active (LaVallie *et al*., 1993, Lu *et al*., 1996). The protein is stabilized and the proportion of soluble protein is therefore enhanced. This could circumvent the use of 8 M urea, which led to the denaturation of the recombinant KLH. For the copper-binding hemocyanins, the expression conditions must be varied with regard to the copper concentration in the incubation medium, to enable the recombinant proteins to incorporate the copper atoms into their active sites. This would be beneficial for correct protein folding and biological activity due to the ability of oxygen binding. Additional investigations could be done to express recombinant proteins with a prokaryotic signal peptide, leading to secreted proteins. The results of the

present work concerning the recombinant expression of SP-KLH1-*a* and SP-KLH2-*a* support this hypothesis. A further tactic for improved folding of recombinantly expressed proteins can be the coexpression of chaperones (DnaK, GroEL, GroES and HtpG). These prokaryotic chaperones help many proteins to fold correctly, including eukaryotic foreign proteins. The coexpression of DnaK increases for example the soluble proportion of recombinantly expressed human growth factor (HGF) in *E. coli* by about 87 % (Blum *et al*., 1992). A final proposal is the application of glycylbetaine and sorbitol in the bacterial growth medium. Glycylbetaine stabilizes newly synthesized proteins, and sorbitol is needed for the uptake of glycylbetaine within the cell. Large amounts of active, soluble recombinant DMAPP/AMP transferase were produced by Blackwell and Horgan (1991) by growing and inducing the *E. coli* cells under osmotic stress in the presence of sorbitol and glycylbetaine. This caused an increase of up to 427-fold in the active yield, and the disappearance of the protein from the pelletable fraction of cell extracts. Correct protein folding assist the understanding of structural features as well as investigations on the biological activity of the recombinant protein.

Especially for the future recombinant production of soluble KLH, the expression in eukaryotic cells is likely to provide promising prospects concerning the presence of essential post-translational modifications, such as disulfide bridges and correct protein folding, with the help of eukaryotic chaperones. The most important factor for the future clinical use of the recombinantly produced KLH, however, will be the correct linkage of the specific carbohydrate side chains at the appropriate N-glycosylation sites. Further expression in insect cells should give new insights concerning this aim, but it is not guaranteed that insect cells will be able to link the correct oligosaccharide chains, that are identical to the ones synthesised in the ER of molluscan cells. KLH is known to have highly branched and very particular carbohydrates side chains (Stoeva *et al*., 1999; Kurokawa *et al*., 2002). The expression of the currently generated KLH1-*abcdefgh* clone as well as future KLH2 encoding clones in a molluscan cell line could represent the most exciting project for correct recombinant expression of this clinically important protein. However, the carbohydrate analyses of native KLH has to advanced further before meaningful conclusions can be drawn concerning the correct glycosylation of recombinantly expressed KLH.

E Summary

The recombinant expression of 19 different substructures of KLH in the prokaryotic system *E. coli* has been successfully achieved: each one of the eight single FUs *a* to *h* of both isoforms, KLH1 and KLH2, two substructures consisting of two consecutive FUs (KLH1-*bc* and KLH1-*gh*) as well as a cDNA encompassing KLH1-*abc*. All recombinant proteins, fused to an N-terminal 6xHis tag, have successfully been detected by immuno precipitation using monoclonal α-His-antibodies and polyclonal α-KLH1- and α-KLH2-antibodies. One exception remained: SP-KLH2-*a*, which was not detected by the α-His-antibodies. This allows speculations as to whether the coexpressed signal peptide can lead, at one hand, to the secretion of the recombinant protein, and on the other to the simultaneous cut-off of the leader peptide, which results in the splitting off of even more N-terminal 6xHis tag, leading to failed recognition by the appropriate antibodies.

The comparison of native KLH with recombinantly expressed prokaryotic (*E. coli*) and eukaryotic (Sf9 insect cells) KLH was done using FU-1*h*. The weak detection by the polyclonal α-KLH1-antibodies of both recombinantly expressed proteins showed that the native protein was the best recognized. For the prokaryotic one, both the denaturation applied for solubilisation of the bacterial inclusion bodies and the inability of bacterial cells to add N-linked glycosylation, are the reason for the poor hybridization. In contrast, KLH1-*h* expressed in eukaryotic insect cells is likely to be glycosylated. The incubation with the α-KLH1-antibodies resulting in the same weak detection, however, revealed that the linked carbohydrate side chains are not those expected.

The establishment of SOE-PCR, together with further improvement, has enabled the generation of a clone encompassing the complete subunit KLH1-*abcdefgh*. The sequence analysis compared to the original KLH1 sequence showed, however, that the resulting recombinant protein is defective in two histidines, required for the copper binding sites in FU-1*b* and FU-1*d* and in three disulfide bridges (FU-1*a*, FU-1*b* and FU-1*g*). This is due to polymerase-related nucleotide exchanges, resulting in a changed amino acid sequence. Nevertheless, all eight potential N-glycosylation sites are present, leading to the speculation that the recombinant protein can in theory be fully glycosylated, which is the most important aspect for the clinical applicability of recombinant KLH as an immunotherapeutic agent. The improvement of this method elaborated during the present work indicates bright prospects for the future generation of a correct cDNA sequence encoding for the complete KLH2 subunit.

F Zusammenfassung

Die rekombinante Expression von 19 verschiedenen KLH-Teilstrukturen ist in dem prokaryotischen System *E. coli* erfolgreich durchgeführt worden: jede einzelne der acht funktionellen Domänen *a* bis *h* beider Isoformen, KLH1 und KLH2, zwei Teilstrukturen bestehend aus zwei funktionellen Domänen (KLH1-*bc* und KLH1-*gh*) sowie der für KLH1-*abc* kodierenden cDNA. Alle rekombinanten Proteine konnten durch spezifische α-KLH1- und α-KLH2-Antikörper, und aufgrund des coexprimierten N-terminalen 6x-His-tags auch durch spezifische α-His-Antikörper, nachgewiesen werden. Eine Ausnahme bildet SP-KLH1-*a*, welches nicht durch den α-His-Antikörper nachgewiesen werden konnte. Die Coexpression des Signalpeptids könnte zum Einen dazu führen, dass das rekombinante Protein sekretiert wird, oder zum Anderen, dass es abgespalten wird, so dass das noch weiter N-terminal gelegene 6x-His-tag mit abgespalten wird, so dass keine Detektion durch den α-His-Antikörper stattfinden kann.

Der Vergleich des nativen KLH mit dem rekombinant exprimierten prokaryotischen (E. coli) und eukaryotischen (Sf9 Insektenzellen) KLH wurde mit FU-1*h* durchgeführt. Beide rekombinanten Proteine werden durch den α-KLH1-Antikörper deutlich schwächer erkannt als die native Form des Proteins. Die Denaturierung zur Lösung der bakteriellen inclusion bodies und die fehlenden N-gebundenen Glykosylierungen könnten beim prokaryotischen Protein die Gründe dafür sein. Das eukaryotisch exprimierte KLH1-*h* ist wahrscheinlich glykosyliert, die schwache Antikörpererkennung lässt allerdings darauf schließen, dass es nicht die richtigen Zuckerketten sind.

Die Etablierung der SOE-PCR ermöglichte die Herstellung eines Klons, der für die gesamte KLH1-Untereinheit kodiert. Der Vergleich mit der Originalsequenz zeigt jedoch, dass dem daraus resultierenden rekombinanten Protein durch polymerasebedingte Nukleotidaustausche zwei Histidine (FU-1*b* und FU-1*d*) und drei Disulfidbrücken (FU-1*a*, FU-1*b* und FU-1*g*) fehlen werden. Alle acht potentiellen N-Glykosylierungsstellen hingegen sind vorhanden, so dass das rekombinante Protein bei korrekter Prozessierung vollständig glykosyliert sein könnte. Dies ist ein wichtiger Aspekt für die spätere klinische Anwendung von rekombinantem KLH als immuntherapeutisches Agens. Die während dieser Arbeit entwickelte Verbesserung der Methode könnte zu einer korrekten, für die gesamte Untereinheit des KLH2 kodierenden cDNA führen.

G Appendix

1 Abbreviations

AP	Alkalic Phosphatase
Bp	basepair
BCIP	5-Brom-4-Chlor-3-Indolylphosphat-p-Toluidinsalt
BSA	bovine serum albumine
cDNA	copy- or complement-DNA
CIA	Chloroform-Isoamylalkohol
Da	Dalton
DEPC	Diethylpyrocarbonat
Dig	Digoxygenin
DMSO	Dimethylsulfoxid
dNTP	dinucleotidetriphosphate
DTT	Dithiothreitol
EDTA	Ethyldiamintetraacetat
FU	Functional unit
dH_2O	Bidistilled water
IPTG	Isopropyl-β-D-Galactosid
kb	Kilobases (1000 nucleotides)
kDa	Kilodalton, 10^3 Dalton
(m)M	(milli) Molar
mRNA	Messenger RNA
NBT	p-Nitrobluetetrazoliumchlorid
OD	Optical censity
PAGE	Polyacrylamid-Gelelectrophorese
PCR	Polymerase chain reaction
RACE	Rapid amplification of cDNA ends
RT	Reverse transkription
SDS	Sodiumdodecylsulfat
Taq	Thermophilus aquaticus
TBE	Tris-boric acid-EDTA
TBS/TBST	Tris-Buffered-Saline / and Tween
Tris	Trishydroxymethylaminomethan
UTR	Untranslatierter Bereich (untranslated region)
UV	Ultraviolett
v/v	volume/volume
w/v	weight/volume

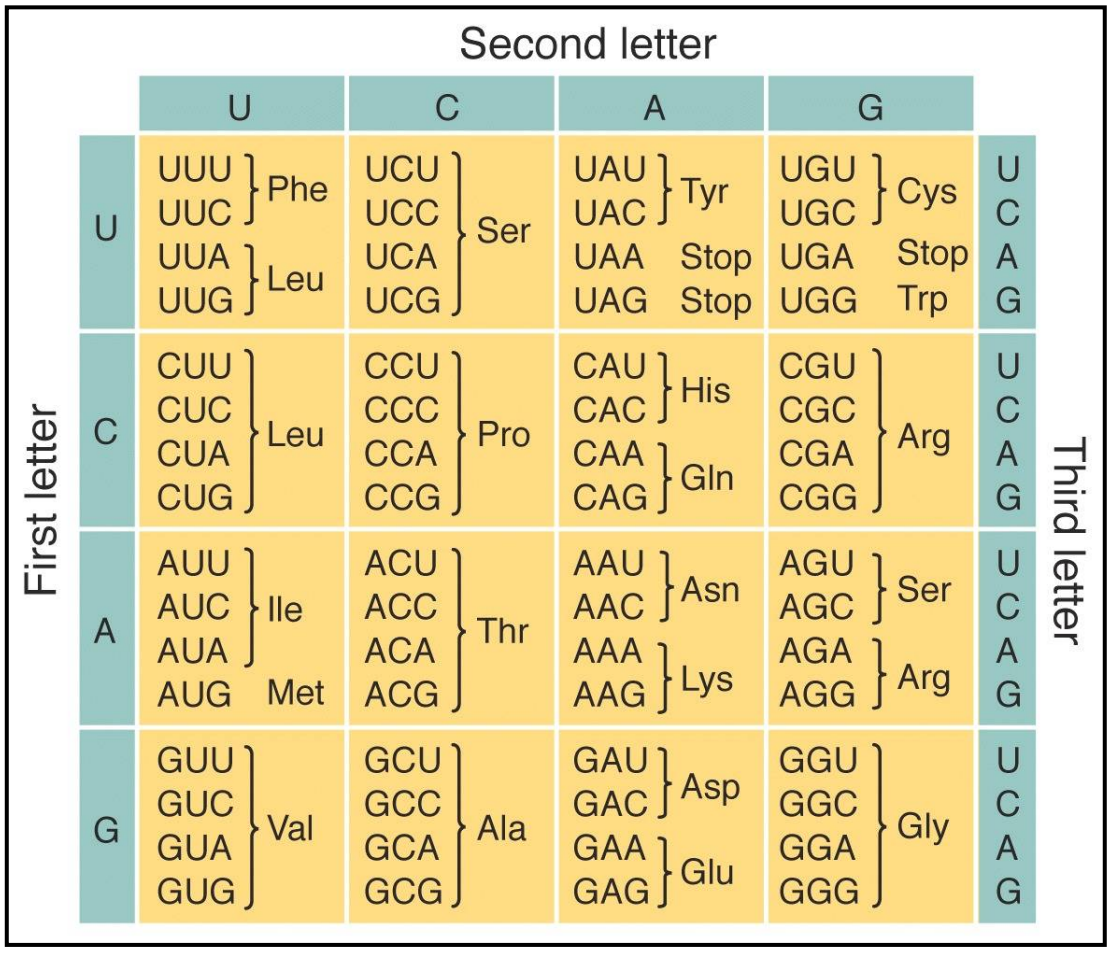
2 Abbreviations of hemocyanins

HpH	*Helix pomatia* hemocyanin
HtH	*Haliotis tuberculata* hemocyanin
KLH	*Megathura crenulata* hemocyanin
OdH	*Octopus dofleini* hemocyanin
RtH	*Rapana thomasiana* hemocyanin

3 Abbreviation-code of amino acids

A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartic acid	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophane
L	Leu	Leucine	Y	Tyr	Tyrosine

4 Standard genetic code of eukaryotic organisms



(www.mun.ca/biology/scarr/MGA2_03-20.html)

5 IUPAC-code for nucleic acids

Code	Beschreibung	Code	Beschreibung
A	Adenine	M	C or A
C	Cytosine	K	T, U, or G
G	Guanine	W	T, U, or A
T	Thymine	S	C or G
U	Uracil	B	C, T, U, or G (not A)
I	Inosine	D	A, T, U, or G (not C)
R	Purine (A or G)	H	A, T, U, or C (not G)
Y	Pyrimidine (C, T, or U)	V	A, C, or G (not T, not U)
N	Any Base (A, C, G, T or U)		

6 Applied primers

Name	Sequence in 5´- 3´ orientation
	GENE-SPECIFIC PRIMERS KLH1
KLH1 5′ UTR+SS US V	GGAGAGGCGCAATGCTGTCGGTCAGGT
KLH1 Anfang a DS seq	TGCTGGATGAGTCCATCTGCAG
KLH1 a/b US	CCAGTTGTTGTTCATCATCCGG
KLH1 b/a DS	ACACTTCTTCACGGGTGAGA
KLH1 b/a DS2	TCAAACGCCTCTCTCAGATTCA
KLH1 b/a DS3	GGCTTGATAACCGTCGACGGAGCG
KLH1 b/c US	CCACATGCTTCTGTCATTCG
KLH1 b/c US2	CTAAACTTGCATCCAGTCTTA
KLH1 b/c US4	GACATCACTGAAGTCGATGG
KLH1 c/b DS	CGAATAAGACGACTCCTTGGCAC
KLH1 c Anfang DS seq	CGAAGTTCATTCATCTCCTCGGGG
KLH1 c US Mitte	AGGAGCATTCAGTGCCATTC
KLH1 c/d US	GTTCAAGCAGCCTTCAGTC
KLH1 d/c DS	CGGTTGGCAGAAGTTACTTCAGC
KLH1 d/c DS lang	GTTGGCAGAAGTTACTTCAGCTTGATATACTTCGCC
KLH1 d/e US1	GGATCTCATCTAGACAGTC
KLH1 d/e US2	GTCATGTCATCCACTCTCCC
KLH1 e/d DS	GCGAATCATCACTGGTTCAGTG
KLH1 e/d DS2	GCTTGTCCAATTGTGTGATAT
KLH1 e/d DS3	GTCCATCATCTGTGTGGGCA

KLH1 e/d DS4	CCAGTGACAGTTGTTGA
Ende 1e US für fehlendes C	CGACATGACGAAATCTTCCAGATTGAAGTAACCATTACATCCTACG
Ende 1e DS für fehlendes C	CGTAGGATGTAATGGTTACTTCAATCTGGAAGATTTCGTCATGTCG
KLH1 e/f US	GGCCTTATTCCCACACCGTC
KLH1 f/e DS	CATGACGAACCTTGTTGAGCG
KLH1 f/g US	GCCGTCAACGGAACAGTTC
KLH1 g/f DS	CCTTTCTGACACCATGTCC
KLH1 g/f DS2	GATGGGCAGTGGTAAGTGTG
KLH1 g/h US	CCTGATCAAGATGCCAACGG
KLH1 h/g DS	CGGCTTCATGGTGTGACAG
KLH1 h Ende DS	CTATTCGTCATCCACATGGACATGTATCCT
KLH1 3′ UTR DS	CATTACACGAAGATTCCATTTAAC
KLH1 3′ UTR DS1	CAGGAGCCCATTCCTTCTCATC
KLH1 3′ UTR DS2	AGAGTACTATATTCCAAACG
	GENE-SPECIFIC PRIMERS KLH2
KLH2 Sp/a for	ATGTGGACCATCTTGGCTCTCCT
KLH2 a Anfang US	GTAGACACAGTGGTGAGAAAGAATG
KLH2 a/b US	CGACAACCCACCTTAGTGC
KLH2 b /a DS	CCATAGCTTTCCTCAAGCTG
KLH2 b/c US	GGATCTCTCCTATCTGCAGATCTC
KLH2 b/c rev	CCACCAGTAAAGCCATGTTTCCTCAG
KLH2 c/d US	GGCGgAAGTCTTAGTGgAAGTA
KLH2 d/c DS	GGCTAGCGCTTGTTACTGCC
KLH2 d/e US	GGATGTGTTGATCAAGATGCGATCC
KLH2 d rev	CATAATATATATTCTTACGTTCACCAGGAACAA
KLH2 e/f US	GCTGGAGCTGAGTTGGACAGC
KLH2 e/f US2	CCATCAAGGTTCACATAAAGG
KLH2 e rev	CCTTCTGTTCTCCTTGAAGAAGGAAAGC
KLH2 f/e DS	CGGTGAGTTCAGATAGTTCC
KLH2 f rev1	CCTCATCATCATGACAAGAAGC
KLH2 f/g US	GAGGTCCATGGTGTAAACAAGACTGCTC
KLH2 f rev2	CTTCACCAGCTGAGTAGATTATAGTAGGTGC
KLH2 g/h US2	GTTCCACTTTGAGTTGAAG
KLH2 h/g DS	CCATCGTGACGTTCACTG
KLH2 h/g DS2	TCGTGACCTGAGGTGTCACCGCC
KLH2 h rev	TTACTCGTCTTCAATGTGGATATGTATTCTAACATTTCC

KLH2 h Ende kurz	TTACTCGTCTTCAATGTGGATATG
KLH2 3'UTR DS	GTATTAGAGACATAGTGTCGC
	ATTACHMENT PRIMERS KLH1
N-fus attB 49 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTGTCGGTCAGGTTGCTTATAGTC GTGTTG
N-fus attB 60 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGAACCTGGTCAGAAAGAGTGTGGA GCATCTG
49 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACACTTCTTCACGGGTGAGATCATTG ATGA
N-fus attB 13 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCAGTTGTTGTTCATCATCCGGAGAAG GGTTTCA
N-fus attB 1 b/c for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCCACATGCTTCTGTCATTCGT
attB 1 c/b Rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACGAATAAGACGACTCCTTGGCAC
13 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGTTGGCAGAAGTTACTTCAGCTT GATATAC
N-fus attB 14 for cpl. richtig	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCAAGCAGCCTTCAGTCATTCAGAA CCA
attB KLH1 e/d rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAGCGAATCATCACTGGTTCAGT
N-fus attB 44 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGATCTCATCTAGACAGTCATGTCATC CACTCTC
14 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACATGACGAACCTTGTTGAGCGACA GGTGGTGC
N-fus attB 17 cpl. rich-tig	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGCCTTATTCCCACACCGTCAATCATC TA
44 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACCCTTTCTGACACCATGTCCAGCC ATAGAAG
N-fus attB 46 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCCGTCAACGGAACAGTTCTTCCGG CTTCAATCCT
17 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACGGCTTCATGGTGAGACAGAGAGT GAATG
N-fus attB 20 for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCTGATCAAGATGCCAACGGAAGAACA CAGCCCAG
46 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACAGGAGCCCATTCCTTCTCATCAAG AGTACTATAT
20 attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATTACACGAAGATTCCATTTAACTC GAAGTTA
	ATTACHMENT PRIMERS KLH2
N-fus attB 2 SP-a for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGTGGACCATCTTGGCTCTCCT
2 b/a attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTCCATAGCTTTCCTCAA

N-fus attB 2 a/b for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCGACAACCCACCTTAGTGCA
2 b/c revA attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATCCACCAGTAAAGCCATG
N-fus attB 2c for compl.	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGATGCCAAAGACTTTGGCCATAGCAG
2c attB rev compl.	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTGCAGCTGGCTCGAAAATCACCG
N-fus attB 2 c/d for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGCGGAAGTCTTAGTGGAAGT
2d revA attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGTTCACCAGGAACAAAAAGCAGACTCGG
N-fus attB 2 d/e for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGATCGCATCTTGATCAACACAT
2 f1 revA attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATCCCTCATCATCATGACAAGAA
N-fus attB 2 e/f US2 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATCAAGGTTCACATAAAGGA
2 f2 revA attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAACCTTCACCAGCTGAGTGAGTT
N-fus attB 2 f/g for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAGGTCCATGGTGTAAACAAGAC
2 h/g DS2 attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATCGTGACCTGAGGTGTCACCGCC
N-fus attB 1g/h US2 for	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTCCACTTTGAGTTGAAGA
2 3UTR DS attB rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATATTAGAGACATAGTGTCGC
	VECTOR-SPECIFIC PRIMERS
M13 forward	GTAAAACGACGGCCAG
M13 reverse	CAGGAAACAGCTATGAC
pDONR201 for	TCGCGTTAACGCTAGCATGGATCTC
pDONR201 rev	GTAACATCAGAGATTTTGAGACAC
T7	TAATACGACTCACTATAGGG
pDEST17 rev seq	GGCTTTGTTAGCAGCCTCGA
pDEST 17 rev2 seq	CCCGTTTAGAGGCCCCAAGGGGTTA

7 Nucleotide sequence of the clone KLH1-*abcdefgh*

ATGCTGTCGGTCAGGTTGCTTATAGTCGTGTTGGCTTTGGCCAATGCAGAGAACCTGGTCAGAAAGAGTGT
GGAGCATCTGACTCAGGAAGAGACCCTGGACCTGCAGGCTGCCCTGCGTGAGCTGCAGATGGACTCATCCA
GCATTGGTTTCCAGAAAATAGCTGCTGCTCATGGCGCACCCGCATCTTGTGTGCATAAGGACACTTCTATA
GCTTGCTGTATTCACGGCATGCCTACTTTCCCGCACTGGCACAGGGCGTACGTAGTTCACATGGAGCGAGC
TCTGCAGACGAAGAGGCGTACCTCTGGGATGCCATACTGGGACTGGACAGAACCCATCACCCAGTTGCCCT
CTCTCGCTGCGGACCCTGTTTACACCGATTCCCAGGGCGGAAAGGCTCACACCAACCACTGGTACAGAGGG
AACATTGACTTTCTAGACAAGAAAACCAATCGTGCGGTCGATGACCGTCTTTTTGGGAAAGTGAAGCCAGG
CCAGCACACCCATCTTATGGAGAGTGTTTTGGGCGCACTGGAGCCGGATGAATTCTGTAAATTCGAAATCC
AGTTCGAGTTAGCACATAATGCGATTCACTACCTGGTCGGTGGAAAACATGATTACTCTATGGCAAACCTC
GAGTACACCGCCTACGATCCGATCTTCTTTCTTCACCACTCCAACGTTGACAGAATTTTCGCCATCTGGCA
GAGGCTCCAAGAGCTCCGAAACAAAGGCCCCAAAGCTATGGATTGTGCCCAAGAATTGTTACATCGGAAAA
TGGGACCTTTCAGTTGGGAGGACAATGACATACCACTAACAAATGAACATTCCACTCCAGCCGATCTCCTT
GACTATCGTCAACTTCACTACGACTATGACACGTTGAATCTTAGTGGTATGACTCCAGAGGAACTGAAAGC
CTATTTGGATGAGCGGTCTTCAAGGGCTAGAGCCTTCGCAAGCTTCCGTCTGAAAGGATTTGGAGGTTCAA
CCAATGTCTTTGTGTACGTCCGCATCCCCGACGACAACGATCGTAATGATGACCACTGGGAGAAGGCCGGT
GATTTCTTCGTTCTGGGCGGACCAAGGGAGATGAAATGGCAGTTCTACAGGCCATACTTGTTTGACTTGTC
CGACACTGTTCACAAGATGGGAATGAAGTTGGATGGACACTATACCGTCAAGGCAGAGCTTTTCAGTGTCA
ATGGTACAGCTCTTCCCGATGATTTGTTACCTCATCCAGTTGTCGTTCATCATCCGGAGAAGGGTTTCACT
GATCCTCCGGTGAAGCATCACCAAAGCGCCAATCTcCTCgTTAGAAAGATCATCAATGATCC-
CACCCGTGAAGAAGTGTTGAATCTGAGAGAGGCGTTTCGTAAATTCCAAGAAGATCGCTCCGTCGACGGTT
ATCAAGCCACAGCTGAGTATCATGGTCTTCCGGCCAGATGCCCCCGCCCTGATGCCAAGGACAGATATGCT
TGTTGTGTCCATGGAATGCCAATCTTCCCTCACTGGTACAGGCTTTTTGTCACACAAGTTGAAGATGCTTT
AGTAGGCCGTGGAGCTACCATTGGTATCCCATACTGGGACTGGGCTGAACCCATGACACACATTCCAGGTC
TGGCAGGAAACAGAACTTATGTGGATTCTCGTGGTGCATCCCACACAAATCCTTTTCATAGTTCAGTGATT
GCATTTGAAGAAAATGCTCCCCACACCAAAAGACAAACAGATCAAAGACTCTTTAAACCCGCTACCTTTGG
ACACCACACAGACCTGCTCAACCAGATTTTGTATGCCTTCGAACAAGAAGATTACTGTGACTTTGAAGTCC
AATTTGAGATTACCCATAACACAATTCACGCTTGGACAGGAGGAAGCGAACATTTCTCAATGTCGTCCCTA
CATTACGCAACTTTCGATCCTTTGTTTTACTTTCGCCATTCTAACGTTGATCGTCTTTGGGCCGTTTGGCA
AGCCTTACAGATGAGACGGCATAAACCCTACAGGGCCCACTACGCCATATCTCTGGAACATATGCATCTGA
AACCATTCGCCTTTTCATCTCCCCTTAACAATAACGAAAAGACTCATGCTAATGCCATGCCAAACAAGATC
TACGACTATGAAAATGTCCTCCATTACACACACGAAGATTCAACATTTGGAGGCATCTCTCTGGAAAACAT
AAAAAAGATGGTCCACGAAAACCAGCAAGAAGACGGAATATATGCCGGTTTTCTCCTGGCTGGCATACGTA
CTTCAGCAAATGTTGATATCTTCATTAAAACTACCGGTTCCGTGCAACATAAGGCTGAAACACTTGCAGTG
CTCGGTGGAAGCGAGGAAATGGAGTGGGGATTTGATCGCGTTCTCAAGTTTGACATCACGCACGCTTTGAA
AGATCTCGATCTCACTGCAGATGGCGATTTCGAAGTTACCGTCGACATCACTGAAGTCGATGGAACTAAAC
TTGCGTCCAGTCTTATTCCACATGCTTCTGTCATTCGTGAGCATGCACGTGTTAAATTTGACAAAGTGCCA
GGGAGTCGTCTTATTCGGAAAAATGTAGACCGTTTGGGCCCCGGGGAGATGAATGAACTTCGTAAAGCCCT
AGCCCTACTGAAAGAGGACAAAAGTGCCGGTGGATTCCAGCAGCTTGGTGCATTCCATGGGGAGCCAAGAT
GGTGTCCTAGTCCCGAAGCATCTAAAAAATTTGCCTGCTGTGTTCACGGCATGTCTGTGTTCCCTCACTGG
CATCGACTGTTGACGGTTCAGAGTGGAAATGCTTTGAGACGACATGGCTACGATGGAGCTCTGCCGTACTG
GGATTGGACCTCTCCTCTTAATCACCTTCCCGAACTGGCAGATCATGAGAAGTACGTCGACCCTGAAGATG
GGGTAGAGAAGCATAACCCTTGGTTCGATGGCCATATAGATACAGTCGACAAAACAACAACAAGAAGTGTT
CAGAATAAACTCTTCGAACAGCCTGAGTTTGGTCATAATACAAGCATTGCCAAACAAGTACTGCTAGCGTT
GGAACTGGACAATTTCTGTGACTTTGAAATCCAATATGAGATTGCCCATAACTACATCCATGCACTTGTAG
GAGGCGCTCAGCCTTATGGTATGGCATCGTTTCGCTACACTGCTTCTGATCCACTATCCTACTTGCATCAC
TCCAATACAGATCGTATATGGGCAATATGGCAGGCTTTACAGAAGTACAGAGGAAAACCGTACAACGTTGC
TAACTGTGCTGTTACATCGATGAGAGAACCTTTGCAACCATTTGGCCTCTCTGCCAATATCAACACAGACT
ATGTAACCAAGGAGCATTCAGTGCCATTCAACGTCTTTGACTACAAGACCAATTTCAATTATGAATATGAC
ACTTTGGAATTTAACGGTCTCTCAATCTCTCGGTTGAATAAAAAGCTCGAAGCGGTAAAGAGCCAAGACGG
GTTCTTTGCAGGCTTCCTGTTATCTGGTTTCAAGAAATCATCTCTTGTTAGATTCAATATTTGCACCGATA
GCAGCAACTGTCACCCCGCTGGAGAGTTTTACCTTCTGGGTGATGAAAACGAGATGCCATGGGCATACGAT
AGAGTCTTCAGATACGACATAGCCGAAAAACTCCACGATCTAAAGCTGCATGCAGAAGACCACTTCTACAT
TGACTATGAAGTATTTGACCTTAAACCAGCAAGCCTGGGAAAAGATTTGTTCAAGCAGCCTTCAGTCATTC
ATGAACCAAGAATAGGTCACCaTGAAGGCGAAGTATATCAAGCTGAAGTAACTTCTGC-
CAACCGTATTCGAAAAAACATTGAAAATCTGAGCCTTGGTGAACTCGAAAGTCTGAGAGCTGCCTTCCTGG
AAATTGAAAACGATGGAACTTACGAATCAATAGCTAAATTCCGTGGTAGCCCTGGTTTGTGCCAGTTAAAT
GGTAACCCCATCTCTTGTTGTGTCCATGGCATGCCAACTTTCCCTCACTGGCACAGACTGTACGTGGTTGT
CGTTGAGAATGCCCTCCTGAAAAAAGGATTATCTGTAGCTGTTCCCTATTGGGACTGGACAAAACGAATCG
AGCATTTACCTCACCTGGTTTCAGACGCCACTTACTACAATTCCAGGCAACATCACTATGAGACAAACCCA
TTCCATCATGGCAAAATCACACACGAGAATGTAATCACTACTAGGGATCCCAAGGACAGCCTCGTCCATTC

AGACTACTTTTACGGGCAGGTCCTTTACGCCTTGGAGCAGGATAACTTCTGTGATCTCGAGATTCAGTTGG
AGATATTACACAGTGCATTGCATTCCTTACTTGGTGGCAAAGGTAAATATTCCATGTCAAACCTTGATTAC
GCTGCTTTTGATCCTGTGTTCTTCCTTCATCACGCAACGACTGACGGAATCTGGGCAATCTGGCAAGACCT
TCAGAGGTTCCGAAAACGGCCATACCGAGGAGCGAATTGCTCTACCCAGTTGATGCACACGCCACTCCAGC
CGTTTGATAAGGGCGACAACAATGACGAGGCAGCGAAAACGCATGCCACTCCACATGATGGTTTTGAATAT
CAAAACAGCTTTGGTTATGCTTACGATAATCTGGAACTGAATCACTACTCGATTCCTCAGCTTGACCACAT
GCTACAGGAAAGAAAAAGGCATGACAGAGTATTCGCTGGCTTCCTCCTTCACAATATTGGAACATCTGCCG
ATGTCCATGTATTTGTATGTCTCCCAACTGGGGAACTCACGGAGGACTGCAGTCATGAGGCTGGTATGTTC
TCCATCTTAGGCGGTCAAGCGGAGATGTCCTTTGTATTTGACAGACTTTACGAACTTGACATAACTAAAGC
CTTGAAAAAGAACGGTGTGCACCTGCAAGGGGATTTCGATCTGGAAATTGAGATTACGGCTGTGAATGGAT
CTCATCTAGACAGTCATGTCATCCACTCTCCCACTATACTGCTTGAGGCCGGAACAGATTCTGCCCACGCA
GATGATGGACACACTGAACCAGTGATGATTCGCAAAGACATCACACAATTGGACAAGCGTCAACAACTGTC
ACTGGTGAAAGCCCTCGAGTCCATGAAAGCCGACCATTCATCTGATGGGCTCCAGGCAATCGCTTCCTTCC
ATGCTCTTCCTCCTCTTTGTCCATCACCAGCCGCTTCAAAGAGGTTTGCGTGCTGCGTCCATGGCATGGCA
ACGTTCCCACAATGGCACCGTCTGTACACAGTCCAATTCCAAGATTCTGTCAGAAAACACGGTGCAGTCGT
TGGACTTCCGTACTGGGACTGGACCCTACCTCGTTCTGAATTACCAGAGCCCCTGACCGTCTCAACTATTC
ATGACCCGGAGACAGGCAGAGATATACCAAATCCATTTATTGGTTCCAAAATAGAGTTTGAAGGAGAAAAC
GTACATACTAAAAGAGATATCAATAGGGATCGTCTTTTCCAGGGATCAACAAAAACACATCATAACTGGTT
TATTGAGCAAGCACTGCTTGCTCTTGAACAAAACAACTACTGCGACTTCGAGGTTCAGTTTGAAATTATGC
ATAATGGTGTTCATACCTGGGTTGGAGACAAGGGGCCCTATGGAATTGGCCATCTGCATTATGCTTCCTAT
GATCCACTTTTCTACATCCATCACTCCCAAACTGATCGTATTTGGGCTATATGGCAATCGTTGCAGCGTTT
CAGAGGACTTTCTGGATCTGAGGCTAACTGTGCTGTAAATCTCATGAAAACTCCTCTGAAGCCTTTCAGCT
TTGGAGCACCATATAATCTTAATGATCACACGCATGATTTCTCAAAGCCTGAAGATACATTCGACTACCAA
AAGTTTGGATACATATATGACACTCTGGAATTTGCAGGGTGGTCAATTCGTGGCGTTGACCATATTGTCCG
TAACAGGCAGGAACATTCAAGGGTCATTGCCGGATCCTTGCTTGAAGGATTTGGCGCCTCTGCCACTGTCG
ATTTCCGGGTCTGTCGCACAGCGGGAGACTGTGAAGATGCAGGGTACTTCACCGTGTTGGGAGGTGGAAAA
GAAATGCCTCGGGCCTTTGATCGGCTTTACAAGTACGACATAACAGAAACCTTAGACAAGATGAACCTTCG
ACATGACGAAATCTTCCAGATTGAAGTAACCATTACATCCTACGATGGAACTGTACTCGATAGTAGCCTTA
TTCCCACACCGTCAATCATCTATGATCCTGCTCATCATGATATTAGTTCGCACCACCTGTCGCTCAACAAG
GTTCGTCATGATCTGAGTACACCGAGTGAGCGAGATATTGGAAGCCTTAAATATGCTTTGAGCAGCTTGCA
GGCAGATACCTCAGCAGATGGTTTTGCTGCCATTGCATCCTTCCATGGTCTGCCTGCCAAATGTAACGACA
GCCACAATAACGAGGTGGCATGCTGTATCCATGGAGTGCCTACATTCCCCCACTGGCACAGACTCTACACC
CCCCAATTCGCGCAAGCTCTAAGAAGACATGGCTCCAGTGCAGCAGTACCCTATTGGGACTGGACAAAGCC
AATACATAATATTCCACATCTGTTCACAGACAAAGAATACTACGATGTCTGGAGAAATAAGGTAATGCCAA
ATCCATTTGCCCGAGGGTATGTCCCCTCACACGATACATACACGGAAAGAGACGTCCAAGGAGGCCCGTTC
CACCTGACATCAACGGGTGAACACTCAGCGCTTCTGGATCAAGCTCTTTGGGCGCTGGAACAGCACGACTA
CTGCGATTTTGCAGTCCAGTTTGAAGTCATGCACAACACAATCCATTACCTAGTGGGAGGACCTCAAGTCT
ATTCTTTGTCATCCCTTCATTATGCTTCATATGATCCGATCTTCTTCATACACCACTCCTTTGTAGACAAG
GTTTGGGCTGTCTGGCAGGCTCTTCAAGAAAAGAGGGGCCTTCCATCAGACCGTGCTGACTGCGCTGTTAG
TCTCATGACTCAGAACATGAGGCCTTTCCATCACGAAATTAACCATAACCAGTTCACCAAGAAACACGCAG
TTCCAAATGATGTTTTCAAGTACGAACTCCTGGGTCACAGATACGACAATCTGGAAATCGGTGGCATGAAT
TTGCATGAAATTGAAAAGGAAATCAAAGACAAACAGCACCATGTGAGAGTGTTTGCAGGGTTCCCCCTTCA
CGGAGTTAGAACCTCAGCTGATGTCCAATTCCAGATTTGTAAAACATCAGAAGATTGTCACCATGGAGGCC
AAGTCTTCGTTCTTGGGGGGACTAAAGAGATGGCCTGGGCTTATAACCGTTTATTCAAGTACGATATTACC
CATGCTCTTCATGACGCACACATCACCCCAGAAGACGTATTCCATCCCTCTGAACCATTCTTCATCAAGGT
GTCGGTGACAGCCGTCAACGGAACAGTTCTTCCGGCTTCAATCCTGCATGCACCAACCATTATCTATGAAC
CTGGCCTCGACCATCACGAAGaTCaTCATTCTTCTTCTATGGCTGGACATGGTGTCAGAAAG-
GAAATCAACACACTTACCACTG-
CAGAGGTGGACAATCTCAAAGATGCCATGAGAGCCGTCATGGCAGACCACGGTCCAAATGGATACCAGGCT
ATAGCAGCGTTCCATGGAAACCCACCAATGTGCCCTATGCCAGATGGAAAGAATTACTCGCGTTGTACACA
CGGCATGGCTACTTTCCCCCACTGGCACAGACTGTACACAAAACAGATGGAAGATGCCTTGACCGCCCATG
GTGCCAGAGTCGGCCTTCCTTACTGGGACTGGACAACTGCCTTTACAGCTTTGCCAACTTTTGTCACAGAT
GAAGGGGACAATCCTTTCCATCATGGTCACATAGACTATTTGGGAGTGGATACAACTCGGTCGCCCCGAGA
CAAGTTGTTCAATGATCCAGAGCGAGGATCAGAATCGTTCTTCTACAGGCAGGTTCTCTTGGCTTTGGAGC
AGACAGATTTTTGTCAGTTTGAAGTCCAGTTCGAGATTACCCACAATGCCTTCCATTCCTGGACTGGTGGA
CTGACCCCCTACGGAATGTCCACTCTGGAGTACACGGCCTATGATCCTCTTTTCTGGCTTCATCATGCCAA
TACTGACCGAATATGGGCTATTTGGCAAGCTCTTCAAGAATACAGAGGTCTCCCATACGACCATGCCAACT
GCGAAATCCAGGCAATGAGGAGGCCCCTACGGCCATTCTCTGATCCTATTAATCACAATGCGTTCACTCAT
TCTAACGCAAAACCAACAGACGTTTTCGAGTACAGTAGATCCAACTTCCAGTATGATAACTTACGGTTTCA
TGGTATGACAATAGAAAAGTTGGAGCATAAGCTTGAGAAACAGAAGGAAGAAGACAGGACATTCGCTGCCT
TCCTGCTCCACGGCATTAAAAAAAGTGCTGATGTAAGCTTTGACGTGTGCAACCATGATGGAGAATGCCAC
TTTGCAGGTACCTTCGCTATTCTTGGAGGCGAGCACGAGATGCCTTGGTCCTTTGACAGACTTTTCCGATA
CGACATCACCCAAGTTCTTAAACAAATGCACCTGGAGTATGACTCTGACTTCACGTTCCATATGAGAATTA

TTGGCACTTCGGGTAAACAACTGCCATCAGACCTGATCAAGATGCCAACGGTTGAACACAGCCCAGGAGGA
AAACATCATGAGAAGCATCATGAAGATCACCATGAAGACATACTTGTTCGTAAAAACATTCACTCTCTGTC
ACACCATGAAGCCGAAGAATTGAGAGATGCATTGTACAAGCTTCAAAACGATGAAAGTCACGGTGGATATG
AACATATTGCTGGATTCCATGGTTATCCTAGCCTCTGTCCTGAAAAAGGGGATGAAAAGTATCCTTGCTGT
GTTCATGGTATGTCTATATTCCCCCACTGGCACAGGCTCCACACGATACAGTTTGAAAGAGCTCTGAAGAA
ACATGGTTCCCATCTCGGTATTCCTTACTGGGATTGGACTCAGACGATATCGTCTCTTCCTACCTTCTTTG
CAGATTCTGGTAATAATAATCCATTCTTCAAATACCATATCAGAAGCATAAATCAAGACACTGTTCGAGAC
GTTAATGAAGCGATACTCCAACAAACGAAGTTTGGAGAATTTTCGTCCATATTTTACCTGGCCTTGCAGGC
TCTAGAAGAGGACAATTATTGCGATTTTGAGGTGCAGTACGAGATGCTTCACAATGAGGTGCATGCCTTGA
TTGGTGGAGCCGAGAAGTATTCCATGTCCACGCTGGAATATTCAGCCTTCGGTCCATACTTCATGATCCAC
CACGCGTCTTTGGACAAGATCTGGATAATCTGGCAGGAGCTTCAGAAACGCCGGGTTAAGCCTGCACATGC
TGGTTCGTGTGCCGGAGACATTATGCACGTTCCACTTCATCCTTTCAATTATGAAAGCGTCAACAACGATG
ACTTTACCGGAGAGAATTCTCTCCCCAACGCGGTCGTTGACAGTCATAGATTCAACTATAAATATGACAAT
CTTAaTCTTCACGGCCaCaATATAGAGGAGCTTGAGGAGGtTCTCCGCAGCCTGAGACTTA-
AATCTCGTGTTTTCGCAGGATTTGTTCTGTCCGGAATCCGCACAACTGCAGTTGTGAAAGTGTATATCAAA
AGCGGAACCGATTCCGACGACGAATATGCCGGTTCGTTTGTCATCCTTGGAGGAGCCAAAGAGATGCCTTG
GGCATATGAGAGGTTGTATAGATTTGACATTACAGAAACTGTACACAATCTTAACCTTACGGATGATCATG
TTAAATTCAGGTTTGACTTAAAGAAATATGATCATACTGAATTGGATGCATCAGTACTGCCAGCTCCTATA
ATCGTTCGACGTCCAAATAATGCGGTTTTTGATATCATTGAGATTCCAATAGGGAAAGACGTGAGTCTTCC
ACCGAAGGTGGTCGTCAAGAGAGGCACTAAAATCATGTTCATGTCTGTTGATGAAGCCGTTACCACTCCCA
TGCTGAACCTGGGTAGTTATACAGCAATGTTCAAATGTAAAGTGCCACCGTTCAGCTTCCATGCTTTTGAG
CTTGGGAAAATGTATTCTGTCGAATCTGGTGACTATTTCATGACGGCATCAACAACTGAATTATGCAACGA
TAATAACTTAAGGATACATGTCCATGTGGATGAGCAA*CATCACCATCACCATCAC***TAG**

Black: signal peptide, orange: FU-1*a*, green: FU-1*b*, brown: FU-1*c*, blue: FU-1*d*, red: FU-1*e*, blue-green: FU-1*f*, purple: FU-1*g*, turquoise: FU-1*h*, *black italic*: 6xHis tag, **black bold**: stop codon.

8 Internet addresses

Method / program	Internet address
Align	http://www2.igh.cnrs.fr/bin/align-guess.cgi
Blast	http://www.blast.genome.ad.jp
Chromas	http://www.technelysium.com.au/chromas.html
Expasy-tools	http://www.expasy.ch/tools/
Lalign	http://www2.igh.cnrs.fr/bin/lalign-guess.cgi
Oligonucletid properties	http://www.basic.nwu.edu/biotools/oligocalc.html
Primer Selection	http://alces.med.umn.edu/rawprimer.html
Translate	http://www.expasy.ch/tools/dna.html
Institute for Zoology, Dpt.2	http://www.uni-mainz.de/FB/Biologie/Zoologie/abt2

During this work, these internet addresses were available

H Literature

Alves-Brito, C.F., Simpson, A.J., Bahia-Oliverira, L.M., Rabello, A.L., Rocha, R.S., Lambertucci, J.R., Gazzinelli, G., Katz, N., Correa-Oliveira, R., 1992. Analysis of anti-keyhole limpet haemocyanin antibody in Brazilians supports its use for the diagnosis of acute schistosomiasis mansoni. *Trans R Soc Trop Med Hyg*. Jan-Feb; 86 (1): 53-6.

Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V., Pierson, L.S. 3rd, 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J*. Feb; 5 (2): 433-40.

Baker. R.K., Lively, M.O., 1987. Purification and characterization of hen oviduct microsomal signal peptidase. *Biochemistry*. Dec 29; 26 (26): 8561-7.

Bergman, S., 2004. Die Hämocyanine zweier lebender Fossile: *Nautilius pompilius* und *Nucula nucleus*. Doctoral thesis, pp. 184.

Bernard, P., Couturier, M., 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol*. Aug 5; 226 (3): 735-45.

Bernard, P., Kezdy, K.E., Van Melderen, L., Steyaert, J., Wyns, L., Pato, M.L., Higgins, P.N., Couturier, M., 1993. The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. *J Mol Biol*. Dec 5; 234 (3): 534-41.

Black, M.T., 1993. Evidence that the catalytic activity of prokaryote leader peptidase depends upon the operation of a serine-lysine catalytic dyad. *J Bacteriol.* Aug; 175 (16): 4957-61.

Blackwell, J.R., Horgan, R., 1991. A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* Dec 16; 295 (1-3): 10-2.

Blum, W.F., Ranke, M.B., Savage, M.O., Hall, K., 1992. Insulin-like growth factors and their binding proteins in patients with growth hormone receptor deficiency: suggestions for new diagnostic criteria. The Kabi Pharmacia Study Group on Insulin-like Growth Factor I Treatment in Growth Hormone Insensitivity Syndromes. *Acta Paediatr Suppl*. Sep; 383: 125-6.

Bushman, W., Thompson, J.F., Vargas, L., Landy, A., 1985. Control of directionality in lambda site specific recombination. *Science.* Nov 22; 230 (4728): 906-11.

Cohen, S.N., Chang, A.C., Hsu, L., 1972. Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. USA* 69(8), 2110-2114.

Cuff, M.E., Miller, K.I., van Holde, K.E., Hendrickson, W.A., 1998. Crystal structure of a functional unit from *Octopus* hemocyanin. *J Mol Biol*. May 15; 278 (4): 855-70.

Curtis, J. E., Hersh, E. M., Harris, J. E., Mc Bride, C., Friereich, E. J., 1970. The human primary immune response to keyhole limpet hemocyanin: interrelationships of delayed hypersensitivity, antibody response and in vitro blast transformation". *Clin. Exp. Immunol.*; Vol. 6, 473-491.

Dissous, C., Grzych, J.M., Capron, A., 1986. *Schistosoma mansoni* shares a protective oligosaccharide epitope with freshwater and marine snails. *Nature.* Oct 2-8; 323 (6087): 443-5.

Flamm, J., Bucher, A., Holtl, W., Albrecht, W., 1990. Recurrent superficial transitional cell carcinoma of the bladder: adjuvant topical chemotherapy versus immunotherapy. A prospective randomized trial. *J Urol.* Aug; 144 (2 Pt 1): 260-3.

Foster, D., Davie, E.W., 1984. Characterization of a cDNA coding for human protein C. *Proc Natl Acad Sci USA*. Aug; 81 (15): 4766-70.

Gebauer, W., Harris, J.R., Heid, H., Süling, M., Hillenbrand, R., Söhngen, S., Wegener-Strake, A., Markl, J., 1994. Quaternary structure, subunits and domain patterns of two dicrete forms of keyhole limpet hemocyanin: KLH1 and KLH2. *Zoology* 98, 51-68.

Gebauer, W., Stoeva, S., Voelter, W., Danese, E., Salvato, B., Beltramini, M., Markl, J., 1999a. Hemocyanin subunit organization of the gastropod *Rapana thomasiana*. *Arch. Biochem. Biophys.* 372, 128-134.

Gebauer, W., Harris, J.R., 1999b. Controlled cleavage of KLH1 and KLH2 by the V8 protease from *Staphylococcus aureus* reassociation, electrophoretic and transmission electron microscopy study of peptide fragments. *Eur J Biochem*. May; 262 (1): 166-75.

Gielens, C., De Sadeleer, J., Preaux, G., Lontie, R., 1987. Identification, separation and characterization of the hemocyanin components of *Helix aspersa*. *Comp. Biochem. Physiol*. 88B (1), 181-186.

Gielens, C., Vanhoergaerden, R., Myulle, K., Preaux, G., 1990. Microheterogenity of functional unit d of the b_c-hemocyanin of *Helix pomatia*. In: Preaux, G., Lontie, R. (Eds.). Invertebrate Dioxygen Carriers. Leuven University Press, Leuven, 147-152.

Gielens, C., De Geest, N., Xin, X.Q., Devreese, B., Van Beeumen, J., Preaux, G., 1997. Evidence for a cysteine-histidine thioether bridge in functional units of molluscan haemocyanins and location of the disulfide bridges in functional units d and g of the betaC-haemocyanin of *Helix pomatia*. Eur *J Biochem.* Sep 15; 248 (3): 879-88.

Greenburg, G., Shelness, G.S., Blobel, G., 1989. A subunit of mammalian signal peptidase is homologous to yeast SEC11 protein. *J Biol Chem.* Sep 25; 264 (27): 15762-5.

Grodberg, J., Dunn, J.J., 1988. ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol.* Mar; 170(3) v1245-53.

Grzych, J.M., Dissous, C., Capron, M., Torres, S., Lambert, P.H., Capron, A., 1987. *Schistosoma mansoni* shares a protective carbohydrate epitope with keyhole limpet hemocyanin. *J Exp Med.* Mar 1; 165 (3): 865-78.

Harris, J.R., Markl, J., 1999. Keyhole limpet hemocyanin (KLH): A biomedical review. *Micron* 30, 597-623

Hartley, J.L., Temple, G.F., Brasch, M.A., 2000. DNA cloning using in vitro site-specific recombination. *Genome Res*. Nov; 10 (11): 1788-95.

Herscowitz, H.B., Harold, W.W., Stavitsky, A.B., 1972. Immunochemical and immunogenic properties of a purified keyhold limpet haemocyanin. Immunology. Jan; 22 (1): 51-61.

Herskovits, T.T. Hamilton, M.G., 1991. Higher order assemblies of molluscan hemocyanins. Comp Biochem Physiol B. 99 (1): 19-34.

Hochuli, E., 1989. Genetically designed affinity chromatography using a novel metal chelate absorbent. *Biologically active molecules*, 217 - 239.

Idakieva, K., Severov, S., Svendsen, I., Genov, N., Stoeva, S., Beltramini, M., Tognon, G., Di Muro, P., Salvato B., 1993. Structural properties of *Rapana thomasiana* Grosse hemocyanin: isolation, characterization and N-terminal amino acid sequence of two different dissociation products. *Comp. Biochem. Physiol.* 106B, 53-59.

Jurincic, C.D., Engelmann, U., Gasch, J., Klippel, K.F., 1988. Immunotherapy in bladder cancer with keyhole-limpet hemocyanin: a randomized study. *J Urol*. Apr; 139 (4): 723-6.

Jurincic, C.D., Pixberg, H.U., Gasser, A., Klippel, K.F., 1990. Prostate-specific antigen in prostatic carcinoma. *Urol Int.*, 45 (3): 153-9.

Jurincic-Winkler, C.D., von der Kammer, H., Beuth, J., Scheit, K.H., Klippel, K.F., 1996. Antibody response to keyhole limpet hemocyanin (KLH) treatment in patients with superficial bladder carcinoma. *Anticancer Res*. Jul-Aug; 16 (4A): 2105-10.

Jurincic-Winkler, C.D., Metz, K.A., Beuth, J., Klippel, K.F., 2000. Keyhole limpet hemocyanin for carcinoma in situ of the bladder: a long-term follow-up study. *Eur Urol.* 37 Suppl 3: 45-9.

Kamerling, J. P., Vliegenthardt, J. F. G., 1997. Hemocyanins. *Glycoproteins* (Montreuil, J., Vliegenthart, J. F. G. & Schachter, H., eds.), 123-142. Elsevier Science, Amsterdam.

Kantelhardt, S.R., Wuhrer, M., Tennis, R.D., Doenhoff, M.J., Bickle, Q., Geyer, R., 2002. Fuc(alpha1-->3)GalNAc-: the major antigenic motif of *Schistosoma mansoni* glycolipids implicated in infection sera and keyhole-limpet haemocyanin cross-reactivity. *Biochem J* Aug 15; 366 (Pt 1): 217-23.

Keller, H., Lieb, B., Altenhein, B., Gebauer, D., Richter, S., Stricker, S., Markl, J., 1999. Abalone (*Haliotis tuberculata*) hemocyanin type 1 (HtH1). Organization of the approximately 400 kDa subunit, and amino acid sequence of its functional units f, g and h. *Eur. J. Biochem.* 264, 27-38.

Kertbundit, S., De Greve, H., Deboeck, F., Van Montagu, M., Hernalsteens, J.P., 1991. In vivo random beta-glucuronidase gene fusions in *Arabidopsis thaliana. Proc Natl Acad Sci USA*. Jun 15; 88 (12): 5212-6.

Klippel, K.F., 1991. Weniger Blasenkrebs-Rezidive durch Hämocyanin. *Immuntherapie des Krebses.* 43. Jahrgang, Nr. 7: 29 - 31.

Ko, A.I., Harn, D.A., 1987. Characterization of protective and non-protective surface membrane carbohydrate epitopes of *Schistosoma mansoni. Mem Inst Oswaldo Cruz*; 82 Suppl 4: 115-9.

Krieg, J., Hartmann, S., Vicentini, A., Glasner, W., Hess, D., Hofsteenge, J., 1998. Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. *Mol Biol Cell*. Feb; 9 (2): 301-9.

Kurokawa, T., Wuhrer, M., Lochnit, G., Geyer, H., Markl, J., Geyer, R., 2002. Hemocyanin from the keyhole limpet *Megathura crenulata* (KLH) carries a novel type of N-glycans with Gal(beta1-6)Man-motifs. *Eur J Biochem*. Nov; 269 (22): 5459-73.

Laemmli, U.K, 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. Aug 15; 227 (5259): 680-5.

Lambert, O., Boisset, N., Penczek, P., Lamy, J., Taveau, J.C., Frank, J., Lamy, J.N., 1994. Quaternary structure of *Octopus vulgaris* hemocyanin. Three-dimensional reconstruction from frozen-hydrated specimens and intramolecular location of functional units Ove and Ovb. *J Mol Biol.* Apr 22; 238 (1): 75-87.

Lambertucci, J.R., 1993. Acute schistosomiasis: clinical, diagnostic and therapeutic features. *Rev Inst Med Trop Sao Paulo*. Sep-Oct; 35 (5): 399-404.

Lamy J, Gielens C, Lambert O, Taveau JC, Motta G, Loncke P, De Geest N, Preaux G, Lamy J. 1993. Further approaches to the quaternary structure of octopus hemocyanin: a model based on immunoelectron microscopy and image processing. *Arch Biochem Biophys*. Aug 15; 305 (1): 17-29.

Landy, A., 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem*. 58: 913-49.

Lang, W.H., 1988. cDNA cloning of the *Octopus dofleini* hemocyanin: sequence of the carboxyl-terminal domain. *Biochemistry.* Sep 20; 27 (19): 7276-82.

Lang, W.H., van Holde, K.E., 1991. Cloning and sequencing of *Octopus dofleini* hemocyanin cDNA: derived sequences of functional units Ode and Odf. *Proc Natl Acad Sci USA.* Jan 1; 88 (1): 244-8.

Laurell, C.B., 1965. Antigen-antibody crossed electrophoresis. *Anal. Biochem.*10, 358-361.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., McCoy, J.M., 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* (N Y). Feb; 11 (2): 187-93.

Lee, J.H., Heffernan, L., Wilcox, G., 1980. Isolation of ara-lac gene fusions in *Salmonella typhimurium* LT2 by using transducing bacteriophage Mu d (Apr lac). *J Bacteriol.* Sep; 143 (3): 1325-31.

Lee, J.H., Burke, K., Wilcox, G., 1986. Mutations resulting in promoter-like sequences which enhance the expression of araC in *Salmonella typhimurium*. *Gene*. 46 (1): 113-21.

Lieb, B., Altenhein, B., Lehnert, R., Gebauer, W., Markl J., 1999. Subunit organization of the abalone *Haliotis tuberculata* hemocyanin type 2 (HtH2), and the cDNA sequence encoding its functional units d, e, f, g and h. *Eur. J. Biochem.* 265, 134-144.

Lieb, B., Altenhein, B., Markl, J., 2000. The sequence of a gastropod hemocyanin (HtH1 from *Haliotis tuberculata*). *J. Biol. Chem.* 275, 5675-5681.

Lieb, B., Altenhein, B., Markl, J.,Vincent, A., van Olden, E., van Holde, K.E., Miller, K.I., 2001. Structures of two molluscan hemocyanin genes: significance for gene evolution. *Proc. Natl. Acad. Sci. USA* 98, 4546-4551.

Lieb, B., Boisguerin, V., Gebauer, W., Markl, J., 2004. cDNA sequence, protein structure, and evolution of the single hemocyanin from *Aplysia californica*, an ophistobranch gastropod. *J. Mol. Evol.* 59, 1-10.

Linn, J.F., Black, P., Derksen, K., Rubben, H., Thuroff, J.W., 2000. Keyhole limpet haemocyanin in experimental bladder cancer: literature review and own results. *Eur Urol.* 37 Suppl 3: 34-40.

Lu, Z., DiBlasio-Smith, E.A., Grant, K.L., Warne, N.W., LaVallie, E.R., Collins-Racie, L.A., Follettie, M.T., Williamson, M.J., McCoy, J.M., 1996. Histidine patch thioredoxins. Mutant forms of thioredoxin with metal chelating affinity that provide for convenient purifications of thioredoxin fusion proteins. *J Biol Chem*. Mar 1; 271 (9): 5059-65.

Markl, J., Nour El Din, M., Winter-Simanowski, S., Simanowksi, U. A., 1991a. Specific IgG activity of sera from egyptian Schistosomiasis patients to Keyhole Limpet hemocyanin (KLH). *Naturwissenschaften*; Vol. 78, 30-31

Markl, J., Savel-Niemann, A., Wegener-Strake, A., Süling, M., Schneider, A., Gebauer, W., Harris, J. R., 1991b. The role of two distinct subunit types of Keyhole Limpet hemocyanin (KLH). *Naturwissenschaften*; Vol. 78, 512-514.

Markl, J. Decker, H., 1992. Molecular structure of the arthropod hemocyanins. *Adv. In Comp. & Envir. Physiol.* Vol. 13, pp 325 - 376.

McFadden, D.W., Riggs, D.R., Jackson, B.J., Vona-Davis, L., 2003. Keyhole limpet hemocyanin, a novel immune stimulant with promising anticancer activity in Barrett's esophageal adenocarcinoma. *Am J Surg*. Nov; 186 (5): 552-5.

Meissner, U., Dube, P., Harris, J.R., Stark, H., Markl, J., 2000. Structure of a molluscan hemocyanin didecamer (HtH1 from *Haliotis tuberculata*) at 12 A resolution by cryoelectron microscopy. *J Mol Biol.* Apr 21; 298 (1): 21-34.

Miki, T., Park, J.A., Nagao, K., Murayama, N., Horiuchi, T., 1992. Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. Mutants of DNA gyrase subunit A suppress letD (ccdB) product growth inhibition. *J Mol Biol.* May 5; 225 (1): 39-52.

Miller, K.I., Schabtach, E., van Holde, K.E., 1990. Arrangement of subunits and domains within the *Octopus dofleini* hemocyanin molecule. *Proc Natl Acad Sci USA*. Feb; 87 (4):1496-500.

Miller, K.I., Cuff, M.E., Lang, W.F., Varga-Weisz, P., Field, K.G., van Holde, K.E. 1998. Sequence of the *Octopus dofleini* hemocyanin subunit: structural and evolutionary implications. *J Mol Biol.* May 15; 278 (4): 827-42.

Miyada, C.G., Stoltzfus, L., Wilcox, G., 1984. Regulation of the araC gene of *Escherichia coli*: catabolite repression, autoregulation, and effect on araBAD expression. *Proc Natl Acad Sci USA.* Jul; 81 (13): 4120-4.

Olsson, C. A., Chute, R., Rao, C. N, 1974. Immunologic reduction of bladder cancer recurrence rate. *J. Urol.,* Vol. 111, 173-176

Ptashne, M., 1992. A genetic switch: Phage (Lambda) and higher organisms. (Cambridge, MA: *Cell Press*).

Perbandt, M., Guthohrlein, E.W., Rypniewski, W., Idakieva, K., Stoeva, S., Voelter, W., Genov, N., Betzel, C. 2003. The structure of a functional unit from the wall of a gastropod hemocyanin offers a possible mechanism for cooperativity. *Biochemistry.* Jun 3; 42 (21): 6341-6.

Perlman, D., Halvorson, H.O., 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *J Mol Biol*. Jun 25; 167 (2): 391-409.

Porath, J., Carlsson, J., Olsson, I., Belfrage, G., 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*. Dec 18; 258 (5536): 598-9.

Riggs, D.R., Jackson, B., Vona-Davis, L., McFadden, D., 2002. In vitro anticancer effects of a novel immunostimulant: keyhole limpet hemocyanin. *J Surg Res. Dec*; 108 (2): 279-84.

Riggs, D.R., Jackson. B.J., Vona-Davis, L., Nigam, A., McFadden, D.W., 2005. In vitro effects of keyhole limpet hemocyanin in breast and pancreatic cancer in regards to cell growth, cytokine production, and apoptosis. Am J Surg. Jun; 189 (6): 680-4.

Sander, C., Schneider, R., 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*. 9 (1): 56-68.

San Millan, J.L., Boyd, D., Dalbey, R., Wickner, W., Beckwith, J., 1989. Use of phoA fusions to study the topology of the *Escherichia coli* inner membrane protein leader peptidase. *J Bacteriol.* Oct; 171 (10): 5536-41.

Sargent, E.R., Williams, R.D., 1992. Immunotherapeutic alternatives in superficial bladder cancer. Interferon, interleukin-2, and keyhole-limpet hemocyanin. *Urol Clin North Am.* Aug; 19 (3): 581-9.

Savel-Niemann, A., Wegener-Strake, A., Markl, J., 1990. Keyhole Limpet hemocyanin: On the structure of a widely used immunological tool. *Invertebrate dioxygen carriers* (G. Preaux, R. Lontie; eds), 351-356, Univ. of Leuven Press ; Louvain, Belgium.

Senozan, N. M., Landrum, J., Bonaventura, J., Bonaventura, C., 1981. Hemocyanin of the giant keyhole limpet, *Megathura crenulata*. In: Lamy, J., Lamy, J. (Eds.). *Invertebrate Oxygen-binding Proteins*. Dekker, New York, 703-717.

Senozan, N.M., Briggs, M., 1989. Hemocyanin levels in the giant keyhole limpet, *Megathura crenulata*, from the coast of California. *Comp. Biochem. Physiol.* 94 A, 195-199.

Siezen, R.J., van Bruggen, E.F., 1974. Structure and properties of hemocyanins. XII. Electron microscopy of dissociation products of *Helix pomatia* alpha-hemocyanin: quaternary structure. *J Mol Biol.* Nov 25; 90 (1):77-89.

Sjöström, M., Wold, S., Wieslander, A., Rilfors, L., 1987. Signal peptide amino acid sequences in *E-scherichia coli* contain information related to final protein localization. A multivariate data analysis. *EMBO J*. Mar; 6 (3): 823-31.

Söhngen, S.M., Stahlmann, A., Harris, J.R., Muller, S.A., Engel, A., Markl, J., 1997. Mass determination, subunit organization and control of oligomerization states of keyhole limpet hemocyanin (KLH). *Eur. J. Biochem.* 248, 602-614.

Somasundar, P., Riggs, D.R., Jackson, B.J., McFadden, D.W., 2005. Inhibition of melanoma growth by hemocyanin occurs via early apoptotic pathways. *Am J Surg.* Nov; 190 (5): 713-6.

Stoeva, S., Rachev, R., Severov, S., Voelter, W., Genov, N., 1995. Carbohydrate content and monosaccharide composition of *Rapana thomasiana* grosse (Gastropoda) hemocyanin and its structural subunits. Comparison with gastropodan hemocyanins. *Comp Biochem Physiol B Biochem Mol Biol.* Apr; 110 (4): 761-5.

Stoeva, S., Schutz, J., Gebauer, W., Hundsdorfer, T., Manz, C., Markl, J., Voelter, W., 1999. Primary structure and unusual carbohydrate moiety of functional unit 2-c of keyhole limpet hemocyanin (KLH). *Biochim Biophys Acta.* Nov 16; 1435 (1-2): 94-109.

Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*. May 5; 189 (1): 113-30.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., Dubendorff, J.W., 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*. 185: 60-89.

Sulkowski, E., 1985. Purification of proteins by IMAC. *Trends Biotechnol*. 3, 1 - 7.

Swanson, M. A., Schwartz, R. S., 1967. Immunosuppressive therapy. The relation between clinical response and immunologic competence. *New. Eng J. Med*. Vol. 277, 163-170

Swerdlow, R.D., Ebert, R.F., Lee, P., Bonaventura, C., Miller, K.I., 1996. Keyhole limpet hemocyanin: structural and functional characterization of two different subunits and multimers. *Comp. Biochem. Physiol. B* 113, 537-548.

Terwilliger, N.B. 1998. Functional adaption of oxygen-transport proteins. *J. Exp. Biol*. Apr; 201 (Pt 8):1085-98.

Topham, R., Tesh, S., Westcott, A., Cole, G., Mercatante, D., Kaufman, G., Bonaventura, C., 1999. Disulfide bond reduction: A powerful, chemical probe for the study of structure-function relationships in the hemocyanins. *Arch. Biochem. Biophys.* 369, 261-266.

Tzianabos, A.O., 2000. Polysaccharide immunomodulators as therapeutic agents: structural aspects and biologic function. *Clin Microbiol Rev*. Oct; 13 (4): 523-33.

Van Breemen, J.F., Ploegman, J.H., Van Bruggen, E.F., 1975. Structure of *Helix pomatia* oxy-beta-hemocyanin and deoxy-beta-hemocyanin tubular polymers. *Eur J Biochem*. Oct; 100 (1): 61-5.

Van Holde, K.E. and Miller, K.I., 1982. Hemocyanins. *Q Rev Biophys.* Feb; 15(1):1-129.

Van Holde, K.E., Miller, K.I., ang, W.H., 1992. Molluscan hemocyanins: Structure and function. *Adv. Comp. Envir. Physiol.* (Ch.P. Mangum, ed.); Vol 13, pp 258 - 293.

Van Holde, K.E. and Miller, K.I., 1995. Hemocyanins. *Adv Protein Chem*. 47:1-81.

Van Kuik, J. A., Kamerling, J. P., Vliegenthardt, J. F. G., 1990. Carbohydrate analysis of hemocyanins. Preaux, G., Lontie, R. (Eds.). *Invertebrate Oxygen Carriers*, Leuven University Press, Leuven, 157-163

Vliegenthart, J.F., Casset, F., 1998. Novel forms of protein glycosylation. *Curr Opin Struct Biol*. Oct; 8 (5): 565-71.

Vona-Davis, L., Vincent, T., Zulfiqar, S., Jackson, B., Riggs, D., McFadden, D.W., 2004. Proteomic analysis of SEG-1 human Barrett's-associated esophageal adenocarcinoma cells treated with key-hole limpet hemocyanin. *J Gastrointest Surg*. Dec; 8 (8): 1018-23.

Weeke, B., 1973. Crossed immunoelectrophoresis. *Scand. J. Immunol.* 2, 47-56.

Weigle. W.O., 1964. Immunochemical properties of hemocyanin. *Immunochemistry*. Dec; 12: 295-302.

Weisberg, R.A., Landy, A., 1983. Site-specific recombination in Phage Lambda. *Lambda II*, R.W. Hendrix, J.W. Roberts, F.W. Stahl and R.A. Weisberg, eds (Cold Spring Harbor, NY: Cold Spring Harbor Press), pp. 211 - 250.

Wickner, W., Driessen, A.J., Hartl, F.U., 1991. The enzymology of protein translocation across the *Escherichia coli* plasma membrane. *Annu Rev Biochem*. 60: 101-24.

Wolfe, P.B., Wickner, W., Goodman, J.M., 1983. Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. *J Biol Chem*. Oct 10; 258 (19): 12073-80.

Wuhrer M., Dennis, R.D., Doenhoff, M.J., Bickle, Q., Lochnit, G., Geyer, R., 1999. Immunochemical characterisation of *Schistosoma mansoni* glycolipid antigens. *Mol Biochem Parasitol*. Oct 15; 103 (2): 155-69.

Wuhrer, M., Robijn, M.L., Koeleman, C.A., Balog, C.I., Geyer, R., Deelder, A.M., Hokke, C.H., 2004. A novel Gal(beta1-4)Gal(beta1-4)Fuc(alpha1-6)-core modification attached to the proximal N-acetylglucosamine of keyhole limpet haemocyanin (KLH) N-glycans. *Biochem J. Mar* 1; 378 (Pt 2): 625-32.

YaDeau, J.T., Klein, C., Blobel, G., 1991. Yeast signal peptidase contains a glycoprotein and the Sec11 gene product. *Proc Natl Acad Sci USA.* Jan 15; 88 (2): 517-21.

Danksagung - Acknowledgement - Remerciements

[redacted] danke ich für die Bereitstellung des interessanten Themas sowie für die fachkundige Betreuung meiner Arbeit.

Vielen Dank an [redacted] für die Zweitkorrektur dieser Arbeit.

Ein ganz besonders herzlicher Dank geht an [redacted]: ich kenne niemanden, der so wie Du, egal wie viel Arbeit er hat, immer für seine Leute da ist, sich alle fachlichen und/oder persönlichen Probleme anhört und sich auch noch immer eine Lösung dazu ausdenkt kann! DANKE!

I would like to thank [redacted] for the lot of work he had with the corrections of this thesis. Thank you very much for the very competent hints and the worded encouragements concerning the English language.

Unseren Computerexperten danke ich für die vielen geduldigen Lösungsversuche bei allen schwierigen oder auch noch so banalen Computerproblemen!

Einen ganz lieben Dank möchte ich meinen ehemaligen und aktuellen Kollegen aus den diversen Kandidatenräumen aussprechen: insbesondere [redacted] für die diversen, unglaublich lustigen Cocktailabende, Grillfeten, Weihnachtsfeiern und Tagungen. Vor allem aber für euer offenes Ohr bei allen beruflichen und zwischenmenschlichen Fragen. Ihr wart ein ganz tolles Team!

Der gesamten Arbeitsgruppe danke ich für das angenehme Klima und die Hilfsbereitschaft, ich werde euch alle vermissen!

Ich danke auch [redacted] ganz besonders für die zahlreiche Hilfe, das Korrekturlesen und dafür, dass sie es mir überhaupt ermöglicht hat, soweit zu kommen!

Meine geliebte [redacted], du bist eine unerschöpfliche Quelle und eine unermüdliche Hilfe in allen naturwissenschaftlichen Fragen! Vor allem danke ich Dir aber dafür, dass Du mir in jeder erdenklichen privaten Lage immer mit Rat und Tat zur Seite gestanden hast und an mich geglaubt hast! Danke!

Je souhaite aussi à remercier toute [redacted] pour son soutien moral pendant les moments difficiles. Et surtout [redacted]pour sa bonne humeur, sa gentillesse et son amour immense.

All meinen Freunden und besonders [redacted]danke ich dafür, dass sie immer für mich da waren und mich so gut es ging unterstützt und aufgebaut haben, auch wenn gar nicht alle so genau wissen, was ich da tatsächlich mit den Schnecken gemacht habe. Oder was denn ein Blot ist...

[redacted] danke ich für das Durchstehen der vielen Höhen und Tiefen, der schwierigen Momente sowie der schönen Augenblicke.
Ich liebe Dich!

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich meine Dissertation selbstständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Ich habe keinen anderen Promotionsversuch unternommen.

Mainz, den 2006

Valesca Boisguérin