JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Preparation and functional characterisation of recombinant silicatein-α from the sponge *Suberites domuncula*.

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften
am Fachbereich Biologie
der Johannes Gutenberg-Universität Mainz

Alberto Manfrin

geboren am 15.06.1983
in Sandrigo (Vicenza) – Italien

Mainz, 2014

Dekan:

1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung: 09.12.2014

TABLE OF CONTENTS

1 INTRODUCTION

1.1 SILICON IN BIOLOGICAL SYSTEMS

Silicon is the eighth most common element in the universe by mass, and the second in the Earth's crust. Rarely it occurs as pure free element, but in association with oxygen in the form of silicate minerals, mostly feldspars (containing aluminium) or quartz, composed by pure silicon dioxide (named silica) (Lutgens et al. 2009). Living organisms have evolved ways to incorporate inorganic silica through a process known as biosilicification. Within the plant kingdom, in particular in the monocotyledon family, a great variety of amorphous silica structures has been described in specific parts of the organism (Currie and Perry 2007), and its accumulation in the cell walls seems to be related to a mechanism of defense from herbivores (Keeping and Kvedaras 2008), and as a reinforcement of the tissues (Epstein 1999). Silica has been reported also in the electric organs of the fish *Psammobatis extenta* (Prado Figueroa et al. 2008), but the most impressive examples of biosilicification is found in lower aquatic life forms, including unicellular species like diatoms, and radiolarian, as well as organisms with higher complexity, represented by the sponges (Porifera). The cell walls of diatoms are characterised by intricate siliceous patterns from the nano- to the micro- scale range, reproduced in a specie-specific manner, implying that the process of biosilicification is genetically regulated, and an example of controlled production of nanostructured silica (Kröger et al. 2002). Similarly, within the phylum of the Porifera, several species present skeletal architectures based on siliceous spicules, cylindrical elements that can reach the length of more than 1 meter in the case of *Monorhaphis chuni* (Wang et al. 2011a).

1.1.1 Biomineralisation.

The complex biomineralised structures observed in nature are not the well known crystalline manifestations of the ordered dispositions of atoms in a unit cell typical of abiogenic minerals (Fig. 1.1A), but the result of a more complex cooperation of

organic templates guiding the deposition of the inorganic elements. Biomineralisation can be divided in two subcategories: "biologically induced" and "biologically controlled", depending on the grade of involvement of the mineralizing organism. In the first scenario, cells only provide a surface for nucleation and subsequent mineral growth (Fig. 1.1B), whereas biomineralisation under biological control uses the cellular machinery to direct nucleation, growth, morphology and location of the mineral (Fig. 1.1C-E).

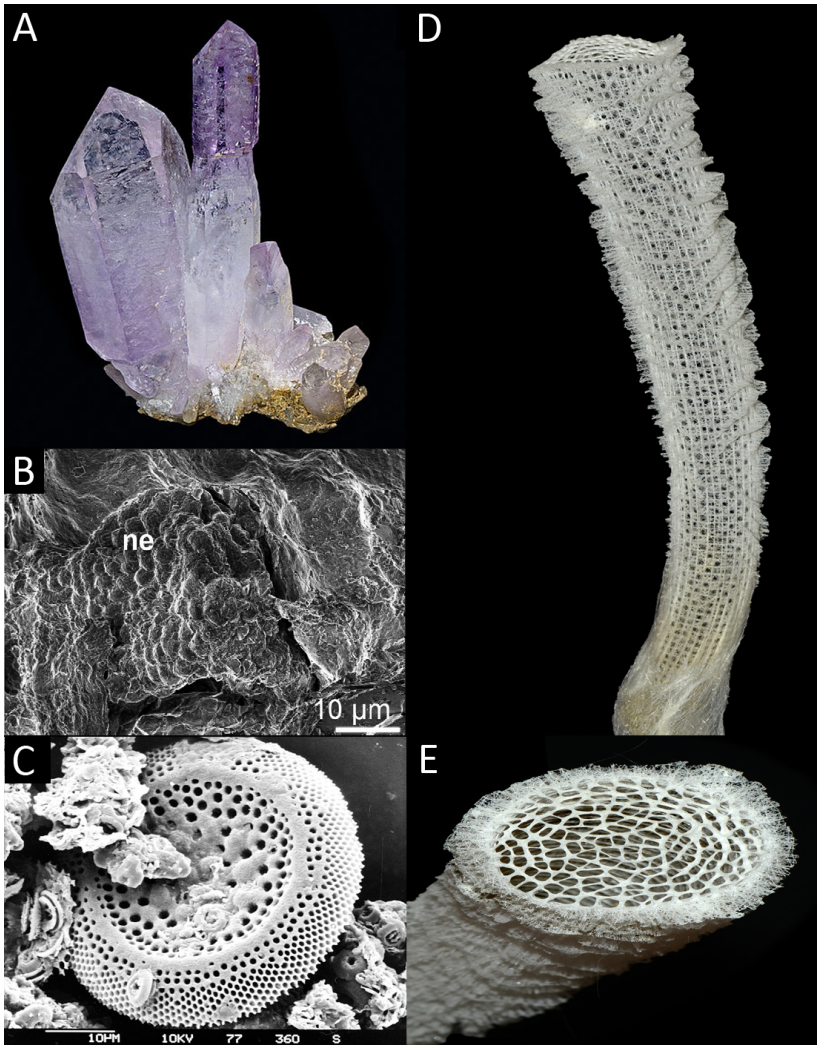


Figure 1.1: A) Example of abiogenic mineral: "Amethyst". Author: Didier Descouens (2013)*. B) Example of biologically induced mineralisation: the outer surface of the varnish rock, composed of coccus-like structures. SEM picture from Wang et al. 2011d. Examples of biologically controlled mineralisation: C) "Microfossils from marine sediments – Diatoms". Author: Hannes Grobe (1980)*; D-E) The skeleton of the glass sponge *Euplectella aspergillum*. Author: Prof. Dr. W.E.G. Müller (2013).

* Graphic material distributed under Creative Commons license.

1.1.2 Biosilica.

Biogenic produced silica, also termed biosilica, derives from the reaction of condensation of the orthosilicic acid (H_4SiO_4) into long polymers. The presence of dissolved silica in the marine environment is associated to riverine influx to the oceans, submarine volcanism, glacial weathering and submarine weathering of basalts (Demaster 1981). Its concentration in seawater is around 3-70 µM, depending on the biological consumption in the local environment; thus, such molecule has to be actively transported into the cells, where is accumulated to concentrations above 100 mM (Schröder et al. 2004). As a comparison, quartz crystals [SiO_2] originate by inorganic mineralisation, through the layered deposition of dissolved H_4SiO_4, usually in hydrothermal environments characterised by high temperatures and pressures. In contrast, biosilicification occurs in physiological conditions, by deposition of silica mediated by an organic matrix that can be composed of proteins, lipids, and proteoglycans (Perry and Keeling-Tucker 2000). In diatoms the process is governed by highly acidic proteins called silaffins, characterised by repetitive elements carrying post-translational modifications, proven to be able to harness silicification *in vitro* (Kröger et al. 2002). In sponges, instead, the breakthrough in the understanding of the basis of their biomineralisation strategies came by the discovery of silicateins (Shimizu et al. 1998)**,** representing a novel paradigm in bioinorganic chemistry (Müller et al. 2013b). In fact, it is the first class of proteins at the present knowledge responsible for the biosilicification in an enzymatic fashion (Krasko et al. 2000). Biosilica has unique mechanical properties, such as strength, stiffness, and toughness because of its hybrid nature of organic/inorganic composites (Mayer 2005). A recent investigation including thermogravimetric analysis, X-ray scattering and solid-state NMR spectroscopy on spicules of *T. aurantia* revealed that biogenic silica is extensively hydrated (10% wt), contains mesostructural domains, and a significant proportion of sodium and potassium ions, thus highlighting the peculiar chemistry of enzymatic produced biosilica (Neilson et al. 2014).

1.2 PORIFERA

1.2.1 Physiology.

Within the Metazoa subkingdom, the most ancient phylum is constituted by the Porifera, evolved 600 millions years ago (Müller 2001). The name "Porifera" derives from the greek πόρος (porus) = pore and φέρω = carry, to describe their unique body structure rich in pores and channels, allowing a constant flowing of the water through them. The majority of the sponge species live in deep sea, although few species are found in freshwater also. Sponges are sessile, heterotrophic, pluricellular organisms that grow attached to rocks or stones, lacking of any nervous, digestive, and circulatory system. These animals are constituted of a gelatinous sack (spongocoel) in most of the cases without body symmetry, contained between two layers of cells: one internal (the choanoderm) and one external (the pinacoderm). The choanoderm presents flagellated cells (choanocites), with a role in the sexual reproduction and the nutrition, which occurs by filtration of microorganisms and solutes present in the water. The filtration is triggered by the continuous movement of the choanocites (from the greek χοάνη = funnel), able to create a current that drives the water intake through the pores, and the outlet towards a central cavity (the osculum). The pinachoderm is formed by pinachocites (from the greek πίναξ = board), a layer of flat cells serving as protection and coating of the organism. The layer between (mesohyl) contains archeoblasts, cells that can differentiate into amebocytes, involved in the processing and distribution of the nutrient, or gametocytes, where the production of the gametes occurs. Most species reproduce sexually, through sperm cells able to fertilised ova, either released or retained in the sponge body. In few species, asexual reproduction occurs by budding, or production of dormant clusters of unspecialised cells (the gemmules) that can form completely new sponges (Bergquist 1998).

1.2.2 Skeletal architectures.

The mesohyl, the intermediate layer of cells in the sponge body, is stiffed by an endoskeleton made of spongin fibers, collagen or mineral spicules. As already

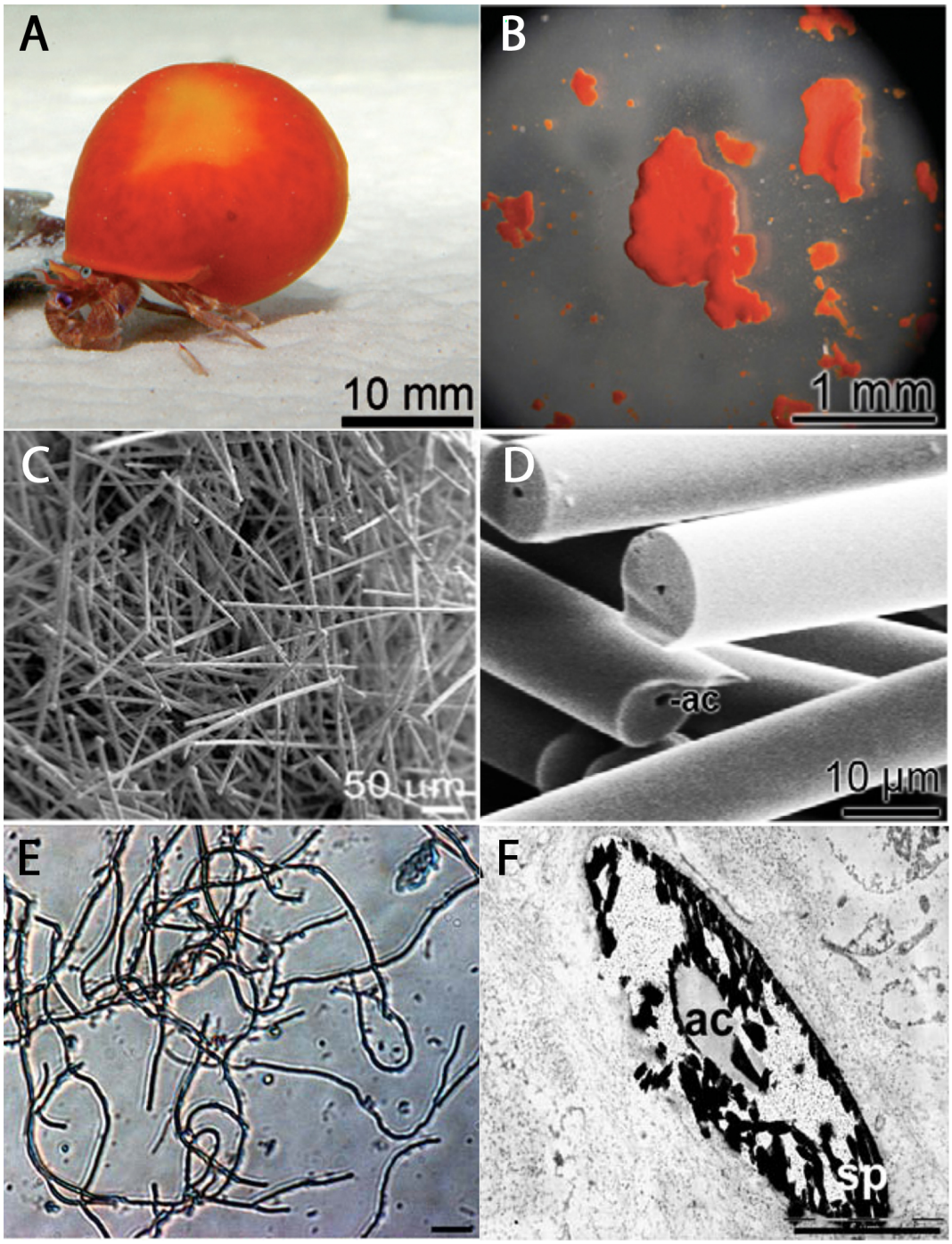


Figure 1.2 | A) *Suberites domuncula*, a member of the class Demospongiae collected in Rovinj, Croatia, and maintained in aquarium at the Institute of Physiological Chemistry, University of Mainz at a temperature of 17°C; the sponge has a globular shape, with an average diameter not longer than 4-5 cm; the color is usually orange-red, although some rare individuals show blue-purple tinge; the sponge adheres on the gastropod shell occupied by a hermit crab (*Paguristes*), in a symbiotic relation offering mobility to *S. domuncula*, and protection from the predators to the crustacean. Picture from Schröder et al. 2008; B) *In vitro* culture of *S. domuncula* cell aggregates: the primmorphs. Picture from Müller et al. 2012; C-D) SEM image of tylostyles isolated from *S. domuncula*. These spicules appear as cylindrical fibres made of silicon dioxide (as confirmed by EDX), surrounding the central cavity (ac); E) Micrograph of the axial filaments extracted from spicules of *S. domuncula* after demineralisation with fluoridric acid; bar = 1 µm; F) TEM image of the section of a spicule, showing the central axial canal (ac) enclosed in a layer of mineral (sp); bar = 1 µm. Pictures C-F from Müller et al. 2005.

mentioned in §1.1, spicules are unique skeletal elements, characterised by a pronounced morphological and dimensional diversity between the different species belonging to this phylum. Such variability is used as a taxonomical tool for identification of the specimens. Two classes of sponges (Demospongiae and Hexactinellidae) produce siliceous spicules, while a third class (Calcarea) is distinguished by having spicules made of calcium carbonate. Generally the spicules are divided in megascleres and microscleres. The first type constitutes the main structure of the sponge skeleton, whereas others are variable in shape and size, and have ancillary functions (Croce et al. 2004).

1.2.3 Spicules formation.

It is generally assumed that the spicules are formed by deposition of silica around an organic filament mainly constituted of proteins and residing in the axial channel (Uriz et al. 2000). The focus of this doctoral project is on the study of the main protein specie forming the filament (Fig 1.2E) and involved in the formation of the spicules (Fig. 1.2C-D). The model organism for our studies is *Suberites domuncula* (Fig. 1.2A), a marine sponge belonging to the class of Demospongiae. The skeleton of this sponge is composed by only two kind of megascleres: monactinal tylostyles and in a less extent diactinal oxeas. These spicules are characterised by an axial canal of 0.3 – 1.6 µm (Fig. 1.2F), surrounded by lamellated layers approximately 0.3-1 µm thick. The dimensions reach up to 450 µm in length and 5-7 µm in diameter (Fig. 1.2C) (Müller et al. 2006). The silicatein-mediated growth is bidirectional, and the increase in length is driven by the elongation of the filament, while the thickening is achieved by apposition of silica, in a serie of several concentric layers. The initial phases of the spicule formation start intracellularly in the sclerocytes, whereas their maturation and growth happens after extrusion into the extra-cellular space (Wang et al. 2011c). Special organelles are responsible for the storage of the silica: the silicasomes (Müller et al. 2008a)**.** A cell culture system (the primmorphs, Fig. 1.2B) permitted these kind of studies for *S. domuncula*, revealing also the interactions of the spicules with galectin and collagen fibers in the extra-cellular space (Eckert et al. 2006), and through immuno-localisation the presence of silicatein either in the axial canal, and

between the silica layers (Müller et al. 2005). The formation of the spicules is definitely a very complex process, where several interactors act in concert at different stages of the growth with regulation functions. Silicase is another enzyme localised through immuno-fluorescence studies on the sections of the tissue-sample of the spicules produced by the primmorphs. It has been shown that it catalyses the dissolution of silica (Schröder et al. 2003), suggesting a mechanism similar to the osteons of the vertebrate bones, result of the opposite activities of synthesising and catabolising enzymes (Garant 2003).

1.3 SILICATEIN

1.3.1 Biological aspects.

It is known for several decades that the axial canal of the siliceous spicules is mainly composed of proteins (Garrone et al. 1981). Several experimental data support this description, in instance the observations at the light microscope on sclerocytes of *Ephydatia muelleri,* which showed the coexistence of silicified and unsilicified protein filaments (Imsiecke et al. 1995), or the SAXS examinations on spicules of several sponges, revealing a high degree of organisation of the protein units (Croce et al. 2004). The first detailed reports describing the nature of the axial filaments as an assembly of a defined family of proteins (the silicateins) were obtained from studies on the Demospongiae *Tethya aurantiam*, by isolation of the protein content upon demineralisation of megascleres, followed by protein characterisation and sequencing. The authors reported the presence of three different isoforms of silicatein (called α, β and γ) in a ratio 12:6:1 (Shimizu et al. 1998). Transcripts of silicatein α and β were identified lately also in *Suberites domuncula*, and upregulation of the genes were observed in the presence of exogenous sodium silicate (Krasko et al. 2000), confirming the involvement of this protein in the spicule formation. This work permitted to predict the aminoacidic sequence of silicateins. Lately, the native silicatein α was extracted and identified also in the spicules of *S. domuncula* (Müller et al. 2005). At the moment in the gene Database (Genebank) are deposited about one hundred sequences of DNA or mRNA related to the various forms of silicatein from different sponges. Moreover, they have been identified by antibody detection in

spicules of several Demospongiae, like *Lubomirskia baicalensis* (Kaluzhnaya et al. 2005), *Geodia cydonium* (Müller et al. 2007b) and *Petrosia ficiformis* (Pozzolini et al. 2004). Members of the class Hexactinellida showed also the presence of silicatein, as reported for *Monorhaphis chuni* (Müller et al. 2008a) and *Crateromorpha meyeri* (Müller et al. 2008c). The full open reading frame (ORF) of the silicatein α gene (Krasko et al. 2000) encodes for a 330 aminoacids long proteins (Fig. 1.3), approximately the same length of the precursor of cathepsins L (a family of proteases similar to silicatein). Therefore, it is generally assumed that also silicatein is produced as an inactive zymogen, and activation occurs upon removal of the N-terminal pro-peptide. The latter is shown to have relevance in the folding and inhibiting properties in cathepsins L (Coulombe et al. 1996). Functional characterisation of the silicatein pro-peptide is under study, although its role in the regulation of localisation, solubility and properties of the enzymatic mineralisation of silicatein is developed in this doctoral work, as well as in recent literature (Müller et al. 2013c; Schröder et al. 2012a).

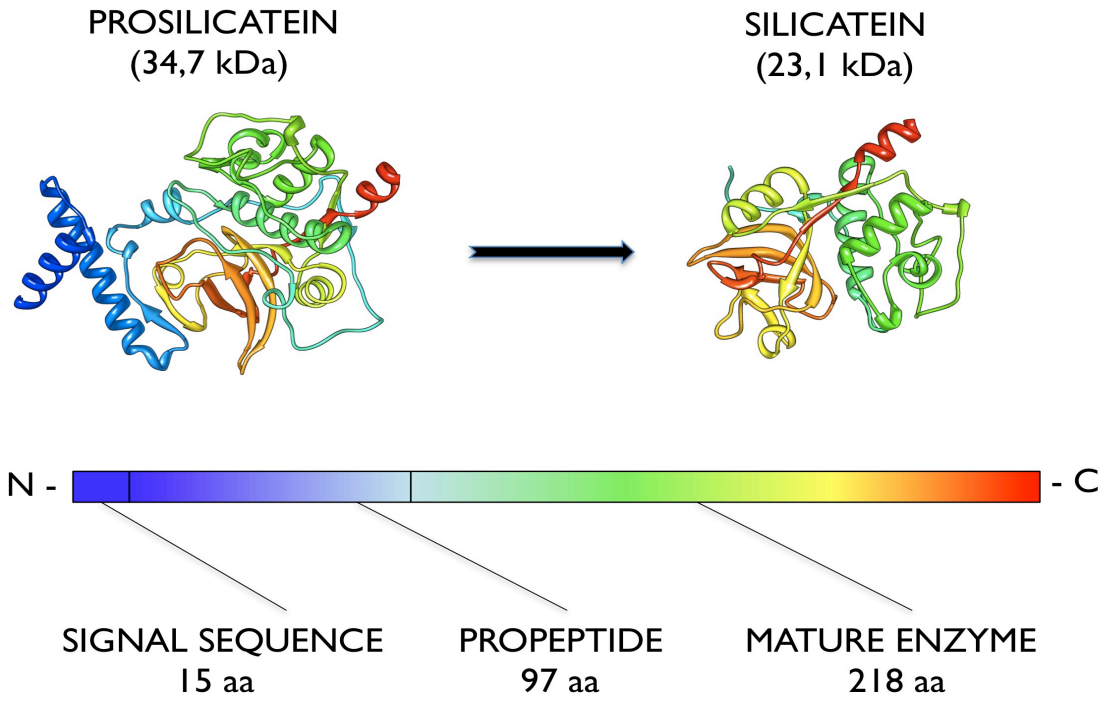


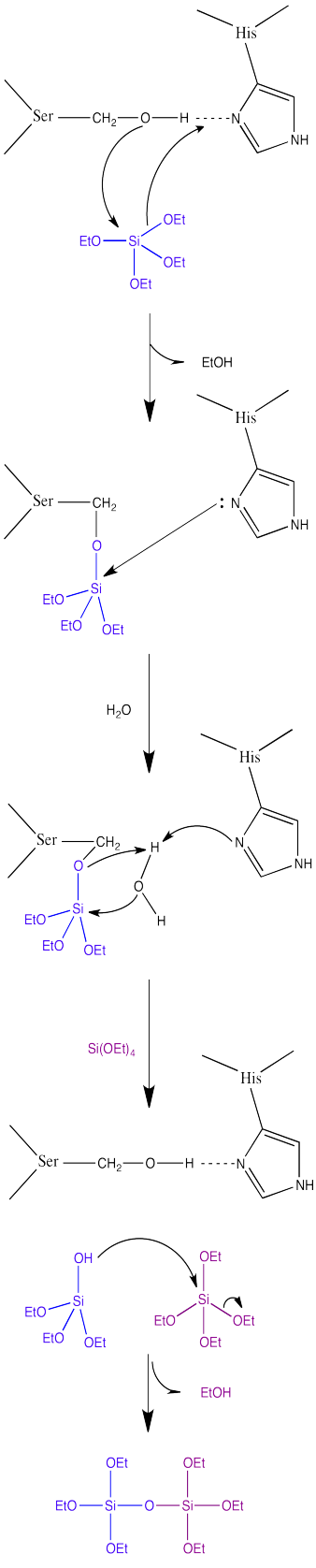
Figure 1.3 | Scheme summarising the main features of silicatein α. The full protein is composed of 330 aminoacids, organised in three domains as follows: from the N-terminal a signal sequence of 15-17 aa, a pro-peptide of 95-97 aa, and the mature enzyme of 218 aa. The calculated molecular weight for the wild type protein is 36,3 kDa, decreasing during maturation after the loss of the signal sequence to 34,7 kDa. Upon cleavage of the pro-peptide, the size of the mature protein is reduced to 23,1 kDa. Original picture produced with softwares Chimera (Salilab) and Adobe Photoshop.

1.3.2 Biochemistry.

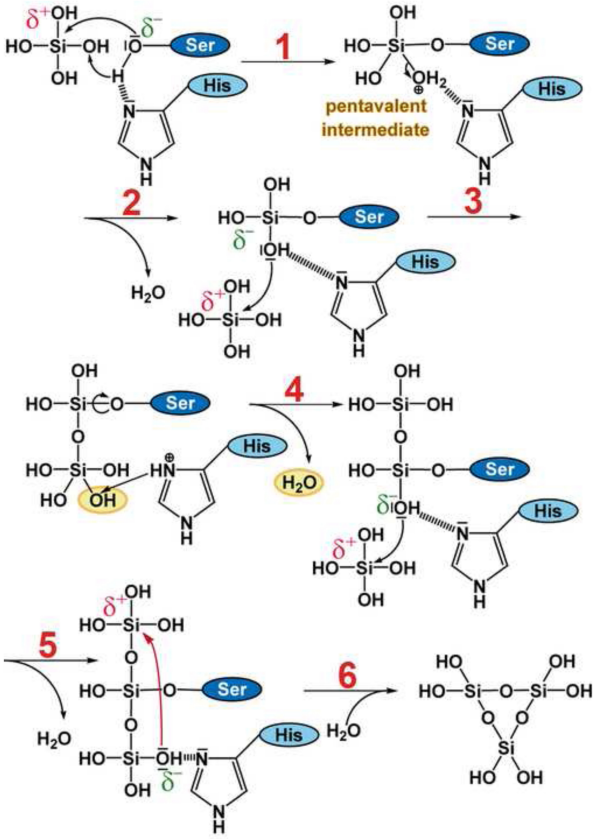
Analyses of sequence similarity revealed that silicatein shares up to 50% of similarity with cathepsins L, becoming a novel member of the papain-like cysteine proteases subfamily. However, the active Cys in cathepsin L is replaced by a Ser residue in silicatein α, whereas the His and the Asn forming the catalytic triad (typical of these proteases) are conserved. In fact, mutants of cathepsin L showed that the replacement of Cys causes the disruption of the protease activity (Coulombe et al. 1996). The calculated pI for silicatein is identical to cathepsin L (pH 5), but is characterised by fewer charged aminoacids and more hydroxyl groups. Hence, a distinctive feature of silicatein is the presence of clusters of hydroxy aminoacids (Ser, Tyr, Thr). Identities between silicatein and cathepsin L regard the six Cys involved in the formation of the three disulfide bridges, which suggests a very similar tertiary structure. Post-translational modifications are also described for *S. domuncula* silicatein α, as demonstrated from the two-dimensional electrophoresis analyses of the spicule extracts, showing at least five phospho-isoforms (Müller et al. 2005). Even a more complex PTM pattern was reported for *Petrosia ficiformis* silicatein β, where methylations, phosphorylations, oxidations, and conversion of cysteines to cisteic acid were revealed by mass spectrometry (Armirotti et al. 2009). Sequence alignments of the latter with ortologous proteins showed that such modifications occur both in conserved and non-conserved aminoacids. This suggests that some PTMs might have important roles in the protein function, thus retaining a certain grade of specie-specificity.

1.3.3 Enzymatic mechanism.

The enzymatic nature of silicatein as a silica esterase, as well as silica polymerase, has been demonstrated (Müller et al. 2008b; Wolf et al. 2010) and its key parameters (i.e. Michaelis-Menten constant) determined (Müller et al. 2008b). A system to reproduce *in vitro* the biosilicifying activity was established by the use of recombinant silicatein, on an artificial substrate: the tetraethoxysilane (TEOS). The silica products can be quantified by the colorimetric molybdic blue assay (Iler 1979), or characterised by observations of the precipitated silica through several microscopic techniques,



A



B

Figure 1.4 | Proposed catalytic mechanisms of silicatein.

A) Reaction with **TEOS**. Hydrogen bonds between the imidazole nitrogen of the conserved histidine and the hydroxyl of the active-site serine increases the nucleophilicity of the serine oxygen, potentiating its attack on the silicon atom of the substrate. Nucleophilic attack on the silicon displaces ethanol, forming a covalent bond protein-O-Si intermediate. The addition of water completes hydrolysis of the first alkoxide bond. Condensation initiated by nucleophilic attack of the released Si-O^- on the silicon of the second substrate molecule then forms the disiloxane product (Cha et al. 1999).

B) Reaction with natural substrate **H_4SiO_4**. (1) Nucleophilic attack of the Ser oxygen atom at the silicon atom of the substrate and transfer of a proton from the imidazole nitrogen of the His residue to an OH ligand of the silicic acid molecule; (2) Release of a water molecule from the formed pentavalent intermediate; (3) Nucleophilic attack of the oxygen atom of one of the OH ligands of the covalently bound silicic acid molecule at the silicon atom of a second orthosilicic acid molecule, facilitated by hydrogen bridge formation of the bound silicic acid to the imidazole nitrogen. (4) Loss of water after proton transfer from the imidazole group. (5) Nucleophilic attack of the oxygen atom of a further OH ligand of the first silicic acid molecule, after rotation of the Si-O-C bond between this molecule and the enzyme Ser residue, at the silicon atom of a third orthosilicic acid species, again facilitated by hydrogen bridge formation to the imidazole nitrogen. (6) Cyclisation of the resulting trisilicic acid by nucleophilic attack of the (negatively charged) oxygen of an OH ligand of the lastly incorporated silicic acid species at the silicon of the second silicic acid species and release of the formed reactive trisiloxane ring after hydrolysis of the Si-O-C bond (Schröder et al. 2012c).

especially transmission electron microscopy (TEM), or scanning electron microscopy (SEM). Studies with mutants of silicatein α on the catalytic site confirmed the importance of the Ser and His residues for the bioactivity of the protein (Zhou et al. 1999). The mechanism proposed for the silica polymerisation in vitro mediated by silicatein follows the principles of catalysis of the well known Ser-His and Cys-His proteases (Cha et al. 1999). Similar conclusions came from the crystallographic investigation on the structure of a mutant of cathepsins L able to form biosilica, carrying features of silicatein (cathsilicatein) (Fairhead et al. 2008). In brief, a unit of TEOS is polymerised into siloxane oligomers first by hydrolisis of the alkoxyilic part, followed by condensation with a second unit (Fig. 1.4a). Because TEOS are synthetic substrates, a model of the reaction occurring *in vivo* was proposed considering the polycondensation of orthosilicic acid units, instead of TEOS and illustrated in Fig. 1.4b (Schröder et al. 2012c).

1.3.4 Self-assembly.

It was possible to observe that silicatein has a high tendency to aggregate, and that this auto-assembly occurs in a fractal way (Murr and Morse 2005). By computational analysis, and in accordance to the proportions of the two isoforms of silicatein identified in *S. domuncula* (α and β), a basic oligomerisation model was proposed, depicting the axial filament as the result of the self-assembly of pentamers composed by 4 monomers of silicatein α and one of silicatein β (Müller et al. 2007a). Such model (Fig. 1.5) has been corroborated by recent observations with electron microscopy techniques of 10 nm thick nanofibrils, composing immature spicules of *S. domuncula* primmorphs (Müller et al. 2013a). Moreover, SAXS analyses coupled with EDX element mapping on the spicules of *Monorhaphis chuni*, revealed a highly ordered internal organisation of the silicatein oligomers, assembling in a perfect 3D lattice, with ability of forming mesoporous silica around it (Zlotnikov et al. 2014). Studies are in progress to correlate the ordered pattern of the axial filament proteins and the interaction dynamics with silintaphins, recently identified by pull-down assays, and proved to be able to enhance the silicatein activity (Wang et al. 2011c; Wiens et al. 2011).

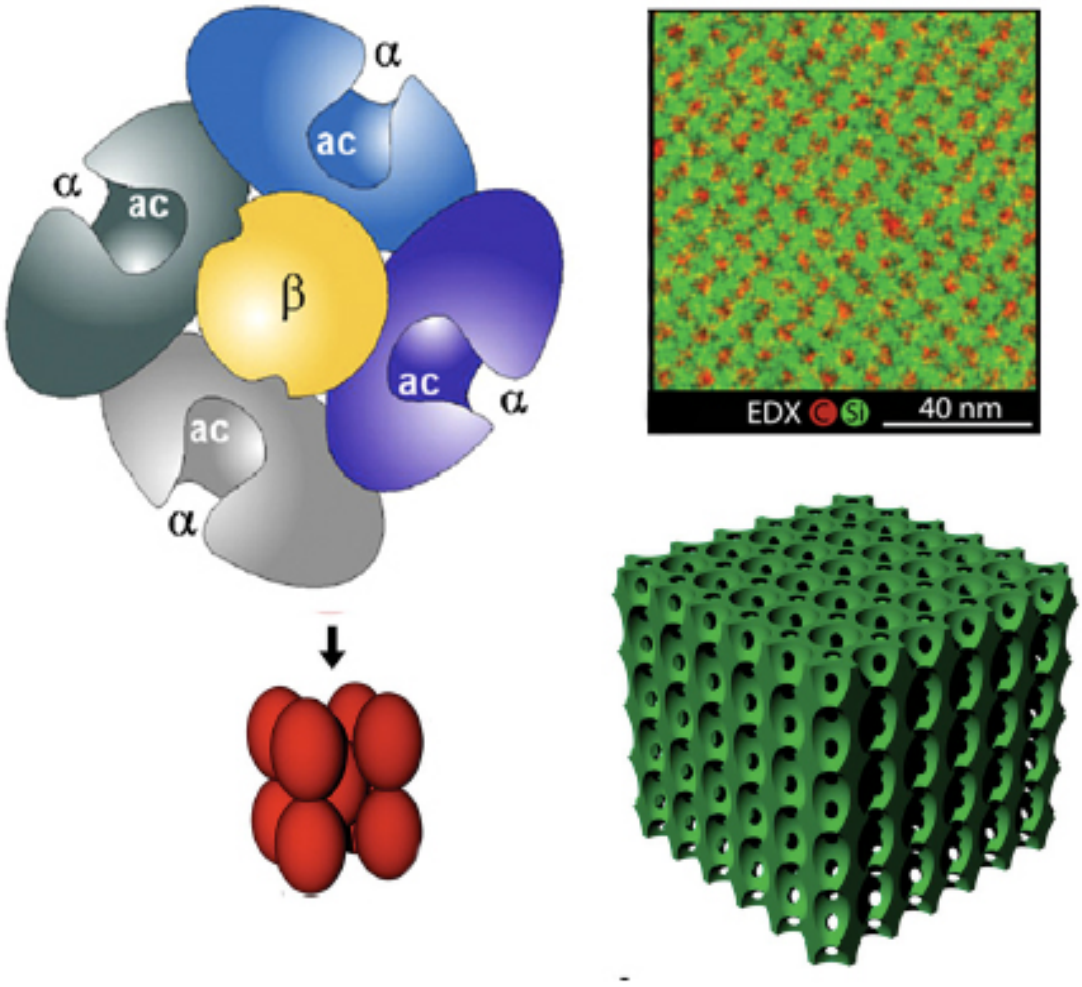


Figure 1.5 | Oligomerisation of silicatein α. A model is proposed for a supramolecular architecture of silicatein proteins. Four monomers of the α form assemble with one monomer of the β form, creating the basic arrangement for the constitution of the protein filament (Müller et al. 2007a). High angle annular dark-field scanning transmission electron microscopy combined with two-dimensional energy-dispersive X-ray spectroscopy analyses in the longitudinal section of *Monorhaphis chuni* spicules agrees with the calculated average diameters of the protein monomers (5 nm). The proteinaceous matrix (red) arrange in a perfect 3D lattice and form a mesoporous silica material (green) with body-centered tetragonal pore system. Picture modified from Zlotnikov et al. 2014.

1.3.5 Applications of silicatein.

The peculiar properties of biosilica (formed enzymatically by silicatein catalysis) have met the interest of many research groups, and enabled the start of several projects focused on the development of innovative technologies in the field of nano-biotechnology, biomedicine, drug delivery, dentistry, and electronics (Hoffmann et al. 2006b). Recombinant His-tagged silicatein α was used to prepare functionalised surfaces by immobilisation on gold (111) (Tahir et al. 2004), and metal oxides (Tahir et al. 2006). The protein showed the ability to retain enzymatic activity after

immobilisation through Ni^{2+} complexation, representing in this way an attractive tool for coating heat- or pH- sensitive materials. Applications in the production of semiconductor to be used in optoelectronics, solar cells, and photocatalysis are under study, by the capability of silicatein of elaborating biochemical reactions on different metal water stable substrates, like titanium (IV) bis (ammonium-lactato) dihydroxide, anionic hexafluorozirconate (Tahir et al. 2005), and gallium oxide (Kisailus et al. 2006). The manufacture of microelectronics is also experimenting the potential advantages of a biological tool like silicatein. Formation of silicified scaffolds can be achieved by application of soft lithographical procedures, where several techniques are available for the controlled deposition of the protein, like the microcontact printing (Wiens et al. 2012). Biosilicification has become also an interesting approach for the design of novel composites for surgery and dentistry. The discovery that biosilica exhibits osteogenic and osteoblast-stimulating activity (Schröder et al. 2012b) makes this material a potential new tool for the cure of the osteoporosis (Wang et al. 2011a). Moreover, biocompatible silica coated TiO_2 dental implants are being developed, involving the application of the silicatein protein (Schröder et al. 2009), as well as novel 3D printed scaffolds with improved mechanical stability (Müller et al. 2014a). Sponge spicules have been shown to have light transmission properties (Cattaneo-Vietti et al. 1996), which can be connected to a simple nervous system, responsible for a regulation of the circadian clock of the animal (Müller et al. 2013d). These fibres are characterised also by interesting mechanical properties, including a higher flexibility in respect of a pure inorganic glass fibre (Wang et al. 2011b). This is due to the nature of a hybrid material consisting of both silica glass, and an organic protein phase. Therefore, the advent of a new generation of bio-inspired optic fibres might follow the natural patterns of biomineralisation.

1.4 HETEROLOGOUS EXPRESSION OF PROTEINS

1.4.1 The recombinant DNA technology.

Like other biological macromolecules such as polysaccharides and nucleic acids, proteins are an essential part of an organism and participate in virtually every cellular

process. Many of them are enzymes, and catalyse biochemical reactions; others have structural or mechanical functions, and play important roles in cell signalling, cell cycle, immune response, and cell adhesion. Their primary sequence is determined at genetic level, and their biosynthesis is regulated by the precise mechanism of translation of the information carried by the mRNA transcripts, codified from the relative gene (DNA) (Voet and Voet 2004).

It tracks back to November 1973 the first report stating that an individual gene can be cloned in autonomously replicating circular bacterial genetic elements (plasmids) and introduced into bacteria (Cohen et al. 1973). Few months later, the demonstration that interspecies recombinant DNA can be propagated and enable the production of a biologically functional protein in a foreign host (Chang and Cohen 1974) set the base for a true revolution in biotechnology. Since then, the development of protocols, engineered hosts and tailored vectors, has shaped a fast-growing technology to support biological and biomedical sciences, as well as the birth of the biopharmaceutical industry. Nowadays a number of proteins (medicines, bioinsecticides, diagnostic kits, and enzymes) are produced at commercial scale through the recombinant DNA technology. And as the field of "protein production" expands, also more widespread is an appreciation of the difficult and strategic choices inherent the process (Graslund et al. 2008). Some experimental approaches, such as protein crystallography, often require milligram quantities of soluble, folded expressed protein (Lieberman et al. 2013). However, proteins sometimes express poorly or fold improperly, constituting a hurdle for any downstream application (Braun and LaBaer 2003). This issue is addressed with time-consuming trial-and-error attempts, and success might depend on the use of several versions of the target protein, designed by genetic engineering strategies. Also, the careful choice of the expression system can have a dramatic impact: each one has its strengths and weaknesses concerning proper folding, post-translational modifications, cost, yields and ease of use (Braun and LaBaer 2003). At the present time, the evolution of this technique made possible to express recombinant proteins in cell cultures of dozens of bacteria strains, yeasts, molds, mammals, plants or insects, as well as transgenic plants and animals. (Demain and Vaishnav 2009).

1.4.2 Recombinant silicateins.

Functional studies and applications (outlined in §1.3.5) based on proteins involved in biomineralisation in sponges have used multiple recombinant DNA methodologies in order to produce the protein samples, mostly because the native counterpart is usually difficult to obtain. In the case of silicatein, only limited amounts of native protein can be recovered from the spicules, and due to the harsh extraction conditions (fluorhydric acid), used for the dissolution of the mineral layers, protein denaturation constitutes an event which is hard to avoid (Shimizu et al. 1998). Although milder procedures were developed (Müller et al. 2008a), bacterial overexpression of His-tagged recombinant proteins afforded significantly higher yields and allowed more straightforward procedures for their purification.

Production of recombinant *S. domuncula* silicatein α (precursor and mature form) was successfully obtained in *E.coli* using the vector pQE30 (Qiagen), engineered with an enterokinase cleavage site between the target protein and the hexahistidine tag (Müller et al. 2003). Variants of such protein, including hexaglutamate-tagged (Natalio et al. 2010), hexacysteine-tagged (Wiens et al. 2012) and also the production of wild-type silicatein β (Wiens et al. 2011) was achieved in bacteria (*E.coli*). However, it has always been observed that the overexpressed protein has the propensity to form aggregates in inclusion bodies, and has limited solubility after refolding. The use of strategies based on the fusion of *T. aurantia* silicatein α to large tags for bacterial expression, such as maltose binding protein (Zhou et al. 1999) or outer membrane protein A (Curnow et al. 2005), also required procedures of reconstitution from the denatured state after separation from the partner tag. A recent study on surface functionalisation claims the production of soluble *S. domuncula* silicatein α in *E. coli*, although issues regarding yields, protein stability, and grade of purity are not extensively discussed (Ki et al. 2013). In conclusion, due to its specific physical and chemical properties, the production of recombinant silicatein α remains a technical impediment, which deserves further development to meet the high demands of multiple downstream applications, both in terms of quantity and quality of the protein samples. Herewith, we present new strategies adopted in order to address the solubility issues, including the construction of a fusion protein Trigger Factor –

silicatein (described in chapter §3), and the implementation of eukaryotic hosts (insect cells and yeast cells) able to express silicatein (described in chapter §4).

1.5 AIMS OF THIS WORK

Widespread studies on the role of organic agents guiding biomineralisation processes and their exploitation in biomedicine, electronics and nanotechnology (outlined in §1.3.5) are in progress in our research group.

Sponges, the most ancient animals in the evolutive timescale, have unique and fascinating biological ways to perform biomineralisation. The importance of this biological process has been acknowledged because it lays the groundwork for the development of multiple biologically-inspired engineering approaches. Although extensive studies have addressed biosilicification mechanisms and the role of matrix-organised silicateins (and other proteins) to mediate the formation of mineralised fine structures, some open questions still remain. In addition, silicatein-produced biosilica, proposed to be also a new biomaterial with invaluable potential in regenerative medicine and tissue engineering (Wang et al. 2014), triggered the interest to explore alternative methods (complementary to the one already existing) to produce *in vitro* the key component for its biosynthesis in sponges: the silicateins. This doctoral work is a contribution that describes the results obtained by an integrative approach involving multiple techniques in molecular biology and protein chemistry, adopted in order to shed some light onto the biological and biochemical features related to this incredibly unique protein.

The focus of this project embraces two main topics:

- The interpretation of the events occurring during the process of maturation from the long to the short form of silicatein-α (Fig. 1.3).
- The technical evaluation on the use of large fusion tags (§3) or synthetic DNA (§4) for the expression of silicatein-α in eukaryotic systems (insect cells and yeasts), as a starting step for the production pipeline of a suitable protein for further structural studies and biotechnological applications.

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Consumables.

Roth *(Karlsruhe,DE)*

- Electroporation cuvettes
- Glass beakers (25, 50, 100, 250, 500, 1000, 2000 and 5000 ml)
- Inoculation loop
- Magnetic stir bars
- Measuring cylinders (10, 100, 1000 and 2000 ml)
- Micropipette tips
- Pasteur pipettes
- Petri dishes
- pH strips
- Plastic conical centrifuge tubes (15 and 50 l)
- Plastic disposable pipettes (1, 5, 10 and 25 ml)
- Quartz and plastic cuvettes
- Reaction tubes (0,5, 1 and 2 ml)
- Sealing film (Parafilm M™)
- Sterile scalpels
- Temperature indicators
- Tissue culture flasks

Millipore *(Schwalbach,DE).*

- Protein centrifugal-filters devices (3 - 5 - 10 kDa cut-off, PES or cellulose membrane; Amicon™ 2, 15, 50 ml)
- Sterile vacuum bottle-top filters
- Syringe filters (0,22 µm – 0,45 µm – 0,8 µm – 5 µm cut-off)

Thermo Fisher Scientific *(Schwerte,DE).*

- Snakeskin™ Dialysis tubing, 10 MWCO

Nunc *(Wiesbaden,DE).*

- Cryotubes vials (1,8 ml)
- Cryotubes rack
- Sterile Erlenmeyer flasks with screw-cap (100, 250, 500 ml)

Gilson *(Villiers Le Bel,FR).*

- Micropipetter (1, 10, 20, 100, 200, 500, 1000, 5000 µl)

2.1.2 Reagents.

Applichem *(Darmstadt,DE).*

- Acetic acid
- β – mercaptoethanol
- Ethidium bromide

Bio-Rad *(München,DE).*

- Bio-Safe™ Coomassie Staining
- Precision Plus Protein™ Dual Color Standard

Fermentas *(St. Leon-Rot,DE)*

- 6X DNA loading dye
- GeneRuler™ 1 kb DNA Ladder

Life Technologies *(Carlsbad,USA)*

- Blocking Reagent
- NuPage® MOPS SDS Running Buffer
- NuPage® MES SDS Running Buffer

Roche *(Mannheim,DE).*

- Complete™ protease inhibitor cocktail tablets
- Complete™ Lysis-Y
- Phenylmethanesulfonyl fluoride (PMSF)
- NBT / BCIP

Roth *(Karlsruhe,DE).*

- 4 X Roti®-Load 1
- Agarose
- Carbenicillin
- E-64
- EDTA
- Ethanol
- Glycerol
- Guanidinium chloride
- Hydrochloride acid (HCl)
- Imidazole
- Iso-propanol
- Isopropylthiogalactopiranoside (IPTG)
- L-aminoacids
- Maleic acid
- Phosphoric acid
- Potassium dihydrogen phosphate (KH_2PO_4)
- Potassium hydrogen phosphate (KH_2PO_4)
- Sodium chloride (NaCl)
- Sodium hydroxide (NaOH)
- Tris base
- Tween-20
- Urea
- Yeast extract

Serva *(Heidelberg, DE)*

- Coomassie Blue G-250

Sigma-Aldrich *(Taufkirchen, DE)*

- Ampicillin
- Ammonium peroxosulphate (APS)
- BSA standards
- Dithiotreytol (DTT)
- Fetal Bovine Serum (FBS)
- Gentamycin
- Kanamycin
- Sodium dodecyl sulphate
- Soybean peptone
- TEMED
- Tetracyclin
- X-gal

2.1.3 Kits.

- Biozym Quantitas DNA marker 2 kb (Biozym Scientific, H. Oldendorf, DE)
- EasyComp™ Transformation kit (Life Technologies, Carlsbad, USA)
- HiFi PCR kit (KAPA Biosystems, Wilmington, USA)
- High Fidelity PCR kit (Fermentas, St. Leon-Rot, DE)
- High Pure PCR Template Preparation kit (Roche, Mannheim, DE)
- Nucleospin® Gel and PCR Clean-Up (Macherey-Nagel, Düren, DE)
- Pichia EasyComp™ kit (Life Technologies, Carlsbad, USA)
- QiaPrep® Spin MiniPrep (Qiagen, Hilden, DE)

2.1.4 Instruments.

- Agarose gel electrophoresis device (Wide Mini-Sub® Cell GT, Bio-Rad, München, DE)
- Automated Protein Purification system (Profinia, Bio-Rad, München, DE)
- Conical tubes centrifuge (5804 R, Eppendorf, Wesseling-Berzdorf, DE)
- Dot-Blot device (Bio-Dot, Bio-Rad, München, DE)
- Electronic scale (PB 300, Mettler-Toledo, Giessen, DE)
- Incubator with agitation system (BE 400, Memmert, Schwabach, DE)
- Laminar flow hood (Slee, BioNova, Roma, IT)
- PCR thermocycler (iCycler, Bio-Rad, München, DE)
- pHmeter (Seven Compact, Mettler Toledo, Giessen, DE)
- Polyacrylamide gel electrophoresis casting and electrophoresis (MiniProtean II, Bio-Rad, München, DE)
- Power-supply (Power Pac 2000, Bio-Rad, München, DE)
- Printer for agarose gels (P93, Mitsubishi Electrics, Ratingen, DE)
- Scanner for documentation of gels (Odyssey, Li-Cor Biosciences, Bad Homburg, DE)
- SDS – PAGE device (NuPAGE® SDS PAGE Gel system Model, Life Technologies, Carlsbad, USA)
- UV-vis microvolume spectrophotometer (Nanodrop2000, Peqlab)
- Steam autoclave (DX-23, Systec, Wettenberg, DE)
- Thermal block (Wealtec HB-1, Peqlab, Erlangen, DE)
- UV-vis spectrophotometer (Smartspec, Bio-Rad, München, DE)
- UV-transilluminator (E-Box 1000, Peqlab, Erlangen, DE)
- Vortexer (Reax 2000, Heidolph, Lodz, PL)
- Western blot device (iBlot, Life Technologies, Carlsbad, USA)

2.1.5 Buffers.

- TAE = 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, pH 8
- TE = 10 mM Tris-HCl, 10 mM EDTA, pH 8
- TBS = 50 mM Tris-HCl, 150 mM NaCl, pH 7,4
- TBST = 50 mM Tris-HCl, 150 mM NaCl, 0,1 % v/v Tween 20, pH 7,4
- AP buffer = 100 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, 0,05% v/v Tween 20, pH 9,5
- Running buffer = 25 mM Tris, 192 mM glycine, 1% w/v SDS

2.1.6 Culture Media.

- LB medium (Roth, Karlsruhe, DE)
- Terrific Broth (Roth, Karlsruhe, DE)
- SOC medium (Roth, Karlsruhe, DE)
- Sf900™ II SFM (Life Technologies, Carlsbad, USA)
- Express Five® SFM (Life Technologies, Carlsbad, USA)

2.1.7 Microorganisms.

- One Shot® TOP10 Chemically Competent *E. coli* (Life Technologies, Carlsbad, USA)
- One Shot® BL21(DE3) pLysS (Life Technologies, Carlsbad, USA)
- Sf9 cells (Life Technologies, Carlsbad, USA)
- High Five™ cells (Life Technologies, Carlsbad, USA)
- MAX Efficiency® DH10Bac™ Competent Cells (Life Technologies, Carlsbad, USA)
- *Pichia pastoris* (strains GS115 and KM71, Life Technologies, Carlsbad, USA)

2.1.8 Enzymes.

- EcoRI (New England Biolabs, Frankfurt am M., DE)
- Factor Xa (Merck, Darmstadt, DE)
- Lyticase (Sigma-Aldrich, Taufkirchen, DE)
- NcoI (New England Biolabs, Frankfurt am M., DE)
- NdeI (Promega, Mannheim, DE)
- NotI (New England Biolabs, Frankfurt am M., DE)
- Proteinase K (Promega, Mannheim, DE)
- SacI (New England Biolabs, Frankfurt am M., DE)
- StuI (New England Biolabs, Frankfurt am M., DE)
- T4 DNA ligase (New England Biolabs, Frankfurt am M., DE)
- Thrombin, Bovine Plasma (Sigma-Aldrich, Taufkirchen, DE)
- XhoI (New England Biolabs, Frankfurt am M., DE)

2.1.9 DNA plasmids.

- pCold™ TF DNA Vector (Takara Bio, St. Germain en Laye, FR)
- pPic9 (Life Technologies, Carlsbad, USA)
- pBac (customised from pFastBac1, Life Technologies, Carlsbad, USA)

2.1.10 Antibodies.

- Primary antibody “Antisilicatein-α #387” (raised in rabbit and provided by W.E.G. Müller group)
- Secondary antibody “Anti-Rabbit IgG – Alkaline Phosphatase conjugated” (Sigma-Aldrich, Taufkirchen, DE)
- Primary antibody “Anti His pAB” (raised in mouse, Roth, Karlsruhe, DE)
- Secondary antibody “Anti-Mouse IgG – Alkalic phosphatase conjugated” (Roth, Karlsruhe, DE)

2.2 GENERAL METHODS

2.2.1 Use of the water.

Water filtered by the MilliQ® technology is used for the most sensitive techniques, like molecular cloning and PCR, while in the other cases (e.g. preparation of buffers, culture media, etc.) standard deionised or distilled water is tolerated.

2.2.2 Sterilisation of the material.

Glassware used for bacteria cultivation is subjected to sterilisation at 180°C in a dry cabinet. Plastic tubes, liquids and other supplies are sterilised by steam autoclave 121°C/15'. Colour shifts of autoclave tape indicate exposure to sterilizing temperatures and is the standard marker used for monitoring the process.

2.2.3 Preparation of the solutions.

Buffers and other solutions used in this work are prepared by dissolution of a precise amount of one or more solutes in the appropriate volume of solvent, following the rules of stoichiometry. Concentrations are expressed in molarity, or other quantitative notations. In certain cases, concentrated stocks are purchased and diluted to the working volume. When required, the pH of the solution is adjusted adding acid or basic solutions (commonly used acids: hydrochloric acid, phosphoric acid, acetic acid; commonly used bases: sodium hydroxide, potassium hydroxide). Electronic pH meters (equipped with a glass electrode) are used to measure the pH of solutions of at least 10 ml. For smaller volumes, pH is estimated by spotting a drop of solution on a pH indicator paper, and the value determined visually by the resulting colour change.

2.2.4 Filtration of the solutions.

To remove impurities, aggregates, or to achieve complete sterility of the solution, several filtration techniques are employed. For volumes until 100 ml the solutions are filtered through syringe filter units with variable membrane type and cutoff. Larger volumes are preferably filtered with vacuum filtration bottle-top devices. Filter-sterilisation is performed using sterile devices, and membranes with a minimum porosity of 0,22 µm.

2.2.5 Weights.

Measurement of weights is performed on electronic scales, with different grade of range, according to the quantity to be measured or the accuracy required.

2.3 MICROBIOLOGY METHODS

2.3.1 Preparation of the culture media.

Culture medium is a nutrient system for the cultivation of bacteria or other cells, which consists of a complex mixture of organic and inorganic materials. Depending on the organism to grow, appropriate nutrients are dissolved in dH_2O and sterilised by autoclave 121°C/15'. Culture medium may be liquid (broth) or solid (when agar is added). When thermolabile molecules (e.g. antibiotics) are required, filter-sterilised stocks have to be prepared separately and added to the autoclaved base medium. The compositions of the nutrient broths are described for every experiment.

2.3.2 Preparation of the agar plates.

Growth of different microorganisms (i.e. bacteria and fungi) is made on agar plates. Agar is a complex mixture of polysaccharides obtained from marine red algae, that melts at 100°C, and solidifies at 44°C, creating a solid and stiff gel. Nutrient broth is mixed with 1 % w/v agar, sterilised by autoclave at 121°C/15', and poured into Petri (25 ml for a dish ⌀ = 9 cm) under a sterile cabinet. After solidification, agar plates

are sealed with plastic paraffin tape (Parafilm), and stored at 4°C. Recipes of the different agar plates are described for every experiment.

2.3.3 Cultivation on agar plates.

Seeding of bacteria and fungi on agar plates can be performed from another agar plate, or from a suspension culture. In the first case a pre-existing colony is picked with a sterile loop and streaked onto the new plate. In the latter, an aliquot of suspension is distributed evenly on the surface of the agar plate with a sterile Drigalski spatula.

2.3.4 Cultivation in suspension.

Small suspension cultures can be started by inoculating a colony grown on an agar plate into sterile liquid broth, contained in a conical centrifuge tube, or an Erlenmeyer flask. Scale-up of the culture volumes are usually achieved by diluting suspensions of cells in mid-logaritmic phase in a larger volume of culture medium. Suspensions are incubated at the desired temperature and agitation speed in orbital shakers.

2.4 MOLECULAR BIOLOGY METHODS

2.4.1 Agarose gel electrophoresis.

DNA elements are size separated and analysed by agarose gel electrophoresis (Serwer and Pichler 1978). Depending on the size of the gel to be casted, 50-100 ml of 1% v/v solution of agarose in TAE buffer is boiled under agitation, then cooled down for 10'-15', and poured into the chamber where one or more combs are placed on the sample loading area. After polymerisation of the agarose, the combs are gently removed, and the tray containing the gel is moved into the electrophoresis chamber and soaked in TAE buffer. DNA samples to be tested are mixed with DNA loading dye and loaded into the wells according to their volume.

Electric field of 80 V from a power supply is applied until the visible dye migrates to the end of the gel. DNA fragments are visualised on a UV-transilluminator, after the gel is stained in a 0,02% v/v EtBr bath for 15' (Waring 1965). Photographs are recorded through a camera directly connected to the scanner, or directly printed on a suitable device for documentation.

2.4.2 Restriction digestion of DNA.

Vectors and inserts used for molecular cloning are first digested with restriction endonucleases, a particular class of enzymes able to cut DNA at specific recognition regions known as restriction sites (Roberts 1976). The DNA is incubated from 1 to 16 hours, with up to 10 U of restriction enzyme, in the appropriate buffer, in a volume of 20-30 µl, at room temperature or at 37°C. It is possible to perform a double digestion adding simultaneously two different restriction enzymes, when the buffering system is compatible. Details are provided in this dissertation for every experiment.

2.4.3 DNA ligation reaction.

The ligation is the joining of DNA operated by the enzyme DNA ligase (Lehman 1974), which catalyses the formation of a phosphodiester bond between 3'-hydroxyl groups of an acceptor and the 5'-phosphate end of a donor nucleotide, upon annealing of complementary sticky or blunt ends previously generated by the restriction endonucleases. Generally the reaction is performed in volumes of 20 µl mixing the DNA fragments with compatible ends, the enzyme DNA ligase and its related activity buffer. Temperature and incubation can be modulated according to the performance of the reaction.

2.4.4 DNA purification.

Purification of DNA from an enzymatic reaction is performed using the kit Nucleospin® Gel and PCR Clean-Up (Macherey-Nagel). The sample is mixed with

Binding Buffer, and adsorbed to a silica membrane (Marko et al. 1982). Contaminations are removed by washing with an ethanolic solution, and finally the pure DNA is eluted in 15-30 µl of buffer TE. The entire procedure is executed following the instructions and the reagents provided by the manufacturer. In case the DNA is embedded in agarose gel (cut-out band), the matrix needs to be dissolved first, by incubating the excised sample at 50°C for 5-10 minutes until complete melting.

2.4.5 Small scale plasmid extraction (Mini-Prep).

Plasmids can be isolated and purified using the kit Qiaprep® Spin (Qiagen). The procedure is based on alkaline lysis of bacterial cells followed by adsorption of the DNA onto silica membranes in the presence of high salt. Pure plasmidic DNA is eluted in 25-50 µl of water or buffer TE. The entire procedure is performed following the instructions and the buffers provided by the manufacturer.

2.4.6 DNA quantitation.

The concentration of DNA is estimated by measuring the absorbance of the light at 260 nm on a UV-vis microvolume spectrophotometer. Droplets of 0,5 – 2 µl of blanking solution, and, afterwards, of sample, are loaded on the clean lower pedestal (sensor), and the instrument controlled by the software Nanodrop2000 (Thermo Fisher Scientific Inc.). Concentration is calculated, assuming the extinction coefficient for DNA at 50 ng per unit of absorbance at 260 nm. Presence of proteins or phenolic compounds that lower the purity of the DNA solution is also estimated measuring the ratio A_{260}/A_{280}. Values higher than 1,8 are indicative of contamination.

2.4.7 Polymerase Chain Reaction.

The *in vitro* amplification of a desired sequence of nucleic acids relies on a well known technique: the Polymerase Chain Reaction, or PCR (Mullis et al. 1986). It reproduces a specific passage occurring in the cell during the mitosis phase: the

DNA replication. New DNA filaments are synthetised artificially reproducing the sequence of a template, starting from a mixture of free deoxynucleotide triphosphates (dNTPs), that are assembled in the correct order following the rules of the base complementarity between pirimidines and purines (G:::C and A:::T) of the double strand DNA (dsDNA). The reaction is catalysed by the DNA polymerase (the most common is Taq DNA polymerase), an enzyme with a temperature optimum at around 70°C. Two synthetic primers (Forward and Reverse) complementary to the 3′ ends of each of the sense and anti-sense (Fwd and Rev, respectively) strand of the DNA target provide the initiation sites for the elongation of the strands. Practically, a basic PCR reaction is performed in a reaction volume of 10-200 µl, in small reaction tubes of 200-500 µl volume. The DNA template containing the fragment to be amplified is mixed with the primers Fwd and Rev, the dNTPs, the DNA polymerase, the buffer solution, and a source of divalent cations ($MgCl_2$). Typically, the reaction consists of a series of 20-40 cycles, each of one consisting of 3 discrete temperature steps. The fast changes of temperature are operated by thermocyclers, special incubators equipped with a thermal block designed for the PCR reaction tubes. Each cycle consists of: 1) a denaturation step (94-98°C for 20-30 s), which causes the DNA melting (separation of the two strands) by disruption of the hydrogen bonds between the complementary bases; 2) an annealing step (50-65°C for 20-40 s), where the primers match with high affinity the template sequence, and the polymerase binds the primer-template hybrid; 3) an elongation step, occurring at the optimal temperature of the specific polymerase used and time period depending on the length of the fragment to be formed. Starting from the 3′ side of the primers, the enzyme condenses the 5′-phosphate group of the dNTPs complementary to the template 5′ to 3′ direction, with the 3′-hydroxyl group at the end of the nascent DNA strand. For DNA polymerases that require heat activation, an initialisation step (94-96°C for 1-9 minutes) might be required. In addition, after the last cycle a final elongation step (70-74°C for 5-15 minutes) is performed to ensure full extension of the single-stranded DNA available. PCR is used in this work for various purposes; therefore, details about reagents, and cycles sequence are mentioned at the corresponding paragraphs.

2.4.8 Bacterial transformation.

Transformation is the process in which free DNA is transferred directly into a competent recipient cell, either for the amplification of plasmidic DNA, or for the heterologous expression of proteins in bacteria. Competent cells (stored at -80°C) are melted on ice and incubated for 30'/0°C with a suitable amount of vector according to its concentration in solution, and the intake capacity of the specific bacteria strain. The mixture is then exposed to heat shock 42°C/45", and quickly cooled on ice for 3'. The bacteria are then resuspended in 250 µl of SOC medium, and the growth initiated by incubation 37°C/45' under agitation. Bacteria are then distributed on an agar plate surface (LB medium), containing the correct antibiotic for the selection of the transformants, which are actively expressing the resistance agent.

2.5 PROTEIN CHEMISTRY METHODS

2.5.1 Quantification by UV.

The concentration of protein solutions can be estimated by measuring their absorbance at 280 nm on a UV-vis spectrophotometer. The absorption of radiation depends on the presence of aromatic aminoacids, like Tyr and Trp, and in a smaller extent by the amount of Phe and S-S bonds. Hence, the concentration of a protein can be derived from the Lambert-Beer law:

$$A_{280} = \varepsilon \, l \, c \quad \rightarrow \quad c = A_{280} / l \, \varepsilon$$

where

- A_{280} = absorbance at 280 nm
- ε = extinction coefficient (M^{-1} cm^{-1})
- l = path length (cm)
- c = concentration (M)

The extinction coefficient of a protein can be predicted using several algorithms, when the aminoacidic composition is known. In this work, the bio-informatic service ProtParam (http://web.expasy.org/protparam) is used, upon submission of

the protein sequence in FASTA format. The absorbance at 280 nm is measured in quartz cuvettes, as they are transparent to this wavelenght. Dilution of the protein sample is necessary when the absorbance value is higher than 1.

2.5.2 Quantification by the Bradford method.

An additional way to assess the concentration of pure proteins in solution is the Bradford method (Bradford 1976), which is a colorimetric assay based on the absorbance shift of the dye Coomassie Brilliant Blue G-250, once it interacts with the basic residues of a protein. The stable protein-dye complex has an absorption spectrum with maximum at 595 nm (A_{595}), and the intensity of the optical absorbance measured spectrophotometrically is proportional to the quantity of the protein present in solution. A calibration curve must be generated by preparing solutions of a protein standard (BSA) in the same buffer solution as the sample to be measured, at concentrations of 0 to 1 mg/ml. Generally, the reaction is performed by incubating 25 µl of the sample for 10'/r.t. with 1 ml of Coomassie based dye, in plastic disposable cuvettes. The absorbance of the standard protein at 595 nm (A_{595}) is plotted versus the corresponding concentration. The plot is a straight line expressed by the following equation and its slope (k) and constant term (q) can be found by interpolation. The concentration of the unknown protein can be derived from the line by interpolation if its A_{595} is measured.

$$A_{595} = kc + q \rightarrow c = (A_{595} - q) / k$$

where

- A_{595} = absorbance at 595 nm
- k and q = constants derived from interpolation of the standard
- c = concentration of the unknown protein

2.5.3 SDS-PAGE.

SDS polyacrylamide gel electrophoresis (SDS-PAGE) performs the separation of fully denaturated and dissociated polypeptide chains (derived from proteins) on a gel made of polyacrylamide and it is the most common technique for the estimation of their molecular weight (Laemmli 1970). Denaturation of the proteins is induced by addition of the amphiphilic detergent sodium dodecyl sulphate (SDS), which promotes the disruption of the hydrophobic interactions and consequently their unfolding into an unorganised form and a negative surface charge. If the protein is oligomeric, its subunits are separated.

Prior to loading onto the gel, it is common to add to the protein samples concentrated denaturating solutions available in the market (known as Laemmli buffer, or gel loading buffer), which contain: SDS, glycerol, reducing agents (like DTT or β-mercaptoethanol, for the reduction of disulfide bonds) and a tracking dye to follow visually the course of the migration into the gel. An additional heating step (95°C / 5') is applied to achieve full denaturation.

The polyacrylamide gel functions as a molecular sieve, where the denaturated polypeptides, driven by the electric field applied, have a mobility inversely proportional to the logarithm of their molecular weight (Shapiro and Maizel 1969). Charge and conformation of the proteins are features that do not affect their mobility once the SDS denaturation is complete.

Table 2.1 | Polyacrylamide gel casting: reaction mixture for the preparation of 12% SDS-PAGE

	Resolving Gel (5 ml)	**Stacking Gel (5 ml)**
1,5 M Tris pH 8,8	1,25 ml	
0,5 M Tris pH 6,8		625 µl
10% SDS	100 µl	50 µl
40% acrylamide / bis-acrylamide (ratio 37 / 1)	1,5 ml	375 µl
10% APS	25 µl	12,5 µl
TEMED	2,5 µl	1,25 µl
dH_2O	2,1 ml	1,4 ml

The polyacrylamide gel casting should be done at the appropriate concentration for optimal resolution in the range of the expected sizes of the desired proteins. The polyacrylamide matrix is formed from a water solution containing Tris/HCl, (that keeps the pH at alkaline values), by the crosslinking of bis-acrylamide molecules to the acrylamide chains. The reaction is initiated by the addition of a catalyser (APS) and a stabiliser (TEMED) (Allison et al. 1974). The usual proportion between acrylamide and bis-acrylamide is around 35 to 1. To improve the resolution, especially for the detection of the smaller peptides, pre-casted gradient gels are also used (e.g. NuPage® 4%-12% gels, Life Technologies).

In Tab. 2.1, the recipe and reagents for the production of a standard 12% polyacrylamide (90 mm x 65 mm x 1 mm) is given, suited for the gel electrophoresis with the system "mini-Protean" (BioRad). The casting of the gel is performed as follows: two glasses are placed vertically in the proper support, separated by a spacer, and the fresh prepared resolving gel carefully but quickly distributed between the glasses. Before solidification occurs, a layer of iso-propanol is added, to flat the surface. After complete removal of the alcohol, the stacking gel solution is added on the top of the resolving gel, and the comb for the creation of the wells, dipped. Upon solidification of the stacking gel, the comb is removed, the gel is immerged in the electrophoresis chamber containing the running buffer (700-800 ml), and the protein samples are loaded according to the volume of the wells (usually maximum 20 µl). A protein ladder, composed by a mixture of proteins of known concentration is loaded in one of the gel lanes, as a "ruler" for the determination of the unknown species sizes. According to the type of comb used to generate the wells, it is possible to create a variable number of electrophoresis lanes, with different loading volumes.

Visualisation of the protein separation by electrophoresis is performed by soaking the gel (staining) into a solution containing Coomassie Brilliant Blue R-250 (Bennett and Scott 1971). The dye binds aminoacids with basic properties, resulting in a clear intense blue band on the gel. Excess of staining solution is removed by several washings in water (destaining). Once the background is clear, and the protein bands visible, documentation of the gels is collected by recording high-resolution pictures through a suited scanner.

2.5.4 Immuno-Blotting.

Proteins separated by electrophoresis can be detected from a protein mixture by antibody recognition. The technique is generally named Western blot (Towbin et al. 1979). After separation by electrophoresis the proteins are transferred from the polyacrylamide gel into a nitrocellulose membrane, using the device iBlot (Life Technologies), following the instructions of the manufacturer. The membrane is then immerged in 25 ml of blocking solution to prevent non-specific binding, and subsequently in the same volume of primary antibody (incubation overnight at 4°C). The primary antibodies used in this work are: I) polyclonal anti-silicatein-α produced in rabbit by our group (used at a dilution of 1:3000 in TBST buffer); II) monoclonal mouse anti-Histidine tag C-terminal (Roth, used at a dilution of 1:2500 in TBST buffer). After removal of the solution, the membrane is washed 3 times for 5' with TBST, before the addition of the secondary antibody (I) anti-rabbit or (II) anti-mouse - IgG-alkaline phosphatase conjugated and incubation at room temperature for one hour under mild agitation. Additional washings of a duration of 5' each are performed, first with TBST, then TBS, and finally alkaline phosphatase buffer. Detection of the protein bands is achieved exposing the membrane to a solution containing a cocktail of NBT and BCIP. The first acts as an oxidant, while the second constitutes the substrate of the alkaline phosphatase. The enzyme catalyses the hydrolisis of BCIP, with formation of an intermediate that upon oxidation by NBT precipitates as an insoluble dark blue diformazan salt. Finally, the reaction is stopped in water, to prevent darkening of the background.

A similar approach is the dot-blot, which differs from the western blot by the fact that the proteins are not separated by electrophoresis but directly blotted into the membrane. The detection of a specific protein in this case relies on the specificity of the antibody, characteristics which have to be confirmed in advance. The device used in this work is the Bio-Dot (BioRad), which allows the blotting of 96 samples at once. The procedure of blocking, binding, and staining is the same as described for Western Blot.

2.5.5 Protein concentration.

The concentration of proteins in solution is increased by usage of protein centrifugal-filter devices, which selectively retain (during a centrifugation cycle) the macromolecules with a size higher than the filter cut-off. The same filters can function also as buffer exchangers, by repeated cycles of dilution/concentration into the new solvation system.

2.5.6 Protein dialysis.

Desalting or buffer exchange is routinely performed by dialysis. Protein solutions are loaded in suited membranes with the appropriate cut-off, hermetically closed and immerged into a vessel containing a volume several times higher than the protein one. Under gentle agitation, small solutes are exchanged by diffusion between the inside and outside of the dialysis membrane.

2.6 PROTEIN PURIFICATION TECHNIQUES

2.6.1 Immobilisation on Metal Affinity Chromatography (IMAC / Ni-NTA).

The Ni-NTA purification system is optimal for the purification of polyhistidine tagged (His-tag) recombinant proteins expressed in bacteria, insect and mammalian cells (Uhlen 2008). The resin is constituted of nitrilotriacetic acid (NTA), a tetradentate chelating ligand, in a highly cross-linked 6% agarose matrix. The His-tag shows a high affinity to Ni^{2+} ions, which are bound to NTA through four coordinations sites (Hochuli et al. 1987).

After a clear lysate is filtered 0,45 µm and loaded into a pre-equilibrated Ni-NTA resin, His-tagged proteins selectively bind to the resin. Upon several washing steps, unbinding of the desired proteins is obtained by flowing of elution buffers containing high amounts of imidazole.

In automated systems, like Profinia (Bio-Rad), the sample is loaded on a pre-equilibrated BioRad Ni-NTA column (bed volume 5 ml) at a flow rate of 1 ml/min. Two washing steps are applied by flowing 6 column volumes (CVs) of buffer (50

mM KH_2PO_4, 300 mM KCl) containing increasing amount of imidazole (10 mM imidazole for the first washing and 20 mM imidazole for the second). Elution of the pure protein is eventually obtained in 4 CVs of buffer containing 250 mM imidazole. The system is equipped with a detector, which records the absorbance of the UV light at 280 nm.

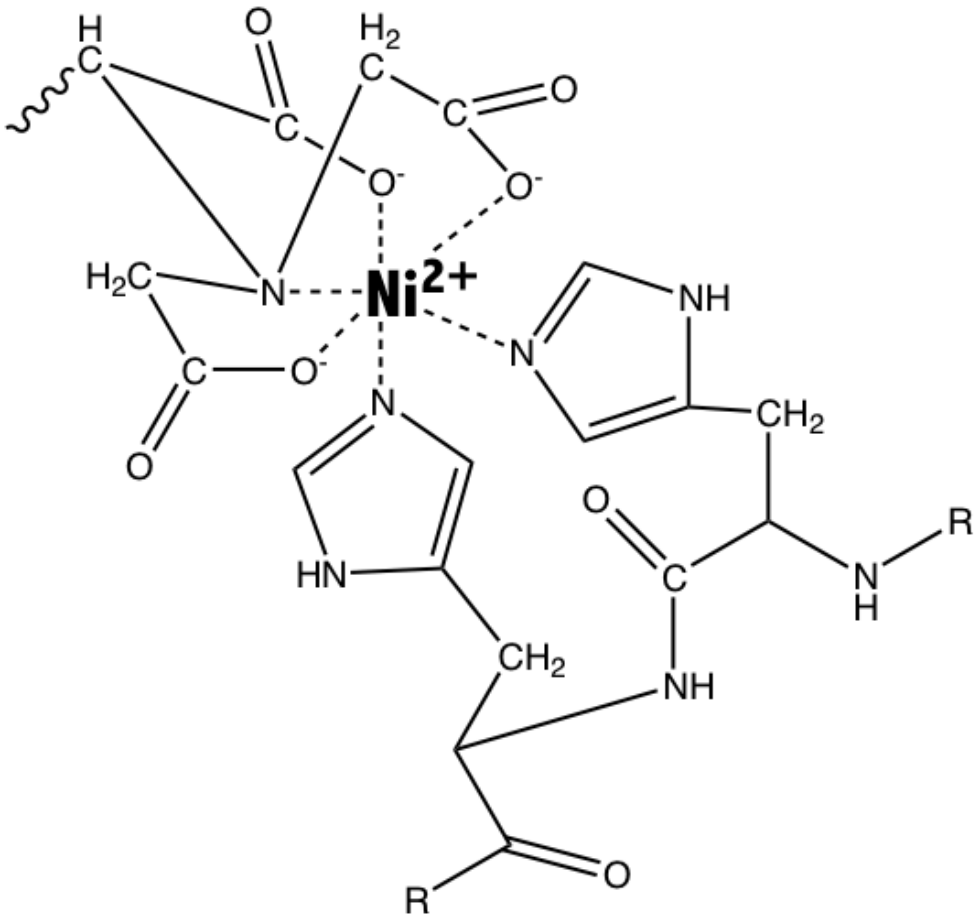


Figure 2.1 Molecular representation of the chemical interaction between His-tagged proteins and nickel^{2+} ions, coordinated by the nitrilotriacetic moieties of the stationary phase. Illustation created with ChemDraw Ultra (CambridgeSoft).

2.6.2 Gel filtration.

In order to achieve increased purity of the recombinant proteins, an additional purification step involving size exclusion chromatography is introduced. In this system, separation of proteins occurs based on their size, by retention on the pores of the stationary phase (Striegel 2008). A protein sample is usually concentrated to 1 mg/ml, before injection into a pre-equilibrated chromatographic column. Proteins are selectively eluted and fractionated into 1,2 ml aliquots, which are consequently analysed by SDS-PAGE. The system is equipped with a pre-packed column Sephacryl S200 HiPrep 16/60 High Resolution (GE Healthcare LifeSciences), characterised by a solid phase of allyl dextran cross-linked by N, N'-methylenebisacrylamide, permitting a fractionation of globular proteins between 5 and 250 kDa.

2.6.3 Ion exchange chromatography.

Ion exchange chromatography is based on adsorption and reversible binding of charged molecules to oppositely charged groups attached to an insoluble matrix. The column used is an anion exchanger, the Hi Trap Q HP 5 ml (GE Healthcare), with a stationary phase made of cross-linked spherical agarose particles, carrying charged groups ($-CH_2N(CH_3)_3$). Step-wise washing solutions with increasing ionic strength (concentration of the salt) are performed and the selective release from the solid phase is assessed by SDS-PAGE. Fractions containing more than 200 mM NaCl are desalted by dialysis to avoid precipitation into the polyacrylamide gel.

3 CLONING, EXPRESSION AND CHARACTERISATION OF FUSION PROTEINS TRIGGER FACTOR / SILICATEIN

3.1 INTRODUCTION

Recombinant techniques where genes of two or more proteins/polypeptide chains are spliced together to generate a new fused protein (chimera) have become a standard tool for biochemistry and biotechnology. Especially in the field of the production of recombinant proteins, the use of the fusion proteins has been described as a powerful strategy in those cases where the target protein shows scarce solubility, or low expression level in common expression systems (Young et al. 2012). However, so far it is not possible to predict the effect of different fusion partners. Often, the screening of a battery of constructs is required to determine the most suitable tag (Lebendiker and Danieli 2014). Various fusion partners are described as solubility and folding enhancers, including glutathione S-transferase (GST), maltose-binding protein (MBO) (Kapust and Waugh 1999), thioredoxin A (TrxA) (LaVallie et al. 2003), small ubiquitin related modifier (SUMO) (Malakhov et al. 2004), N-utilisation substance A (NusA) (De Marco et al. 2004), protein disulfide

isomerase I (DsbA) (Collins-Racie et al. 1995) and short hyper-acidic tags (Zou et al. 2008).

In this chapter we present the expression of silicatein as a chimeric protein fused to a molecular chaperone present in all species of eubacteria, the *E. coli* trigger factor (TF). Nascent polypeptides emerging from the ribosome during the translation phase assume their native conformation with the help of a set of ribosome-associated proteins. TF belongs to this category (Hoffmann et al. 2006a), as it binds the ribosomal 50S subunit. Crystal structure data reveal its multidomain organisation, comprising: an N-terminal domain (domain I or "head"), located adjacent to the ribosomal translation exit tunnel, and thought to stabilise the unfolded polypeptide before the initiation of the following folding steps; a central one (domain II or "body") with peptidyl-prolyl *cis-trans* isomerase (PPIase) activity; a C-terminal loop (domain III or "tail") determining its chaperone activity (Ludlam et al. 2004). In contrast to many other intracellular chaperones, TF is ATP independent (Hartl et al. 2011), and most of its folding-promoter properties depends on its capability of suppressing competing reactions that lead to degradation (Hoffmann et al. 2006a) or aggregation (Huang et al. 2000).

The following are discussed in the next paragraphs :

1) The cloning, overexpression and purification of recombinant fusion proteins TF-pre-pro-silicatein (TF-P1) and TF-pro-silicatein (TF-P2)(Fig. 3.4), obtained by heterologous expression in *E. coli*.
2) The optimisation of the reaction parameters for the proteolytic cleavage reaction and protein solubility considerations.
3) The characterisation of the enzymatic activity, related to the ability of silicatein to produce biosilica *in vitro*.

3.2 DESCRIPTION OF THE EXPRESSION SYSTEM

3.2.1 The vector pCold and the cold shock technology.

The expression vector employed for the creation of the fusion proteins TF-Ps is "pCold™TF DNA" (Accession Number AB213654, Fig. 3.1), property of the company Takara Bio Inc. (South Korea). Several successful examples render the aforementioned plasmid a powerful tool for a high yield production of proteins with improved solubility, purity and stability (Qing et al. 2004).

The 5769 bps plasmid is defined as a "cold shock expression vector", for the presence of the *cspA* promoter (cold shock protein A) and related elements. This feature permits to up-regulate the production of the target proteins at lowered incubation temperatures (15°C), while repressing the expression of other cellular proteins. Protein expression is induced by concentrations of isopropylthiogalactopiranoside (IPTG) up to 1 mM, which binds to the *lac* repressor

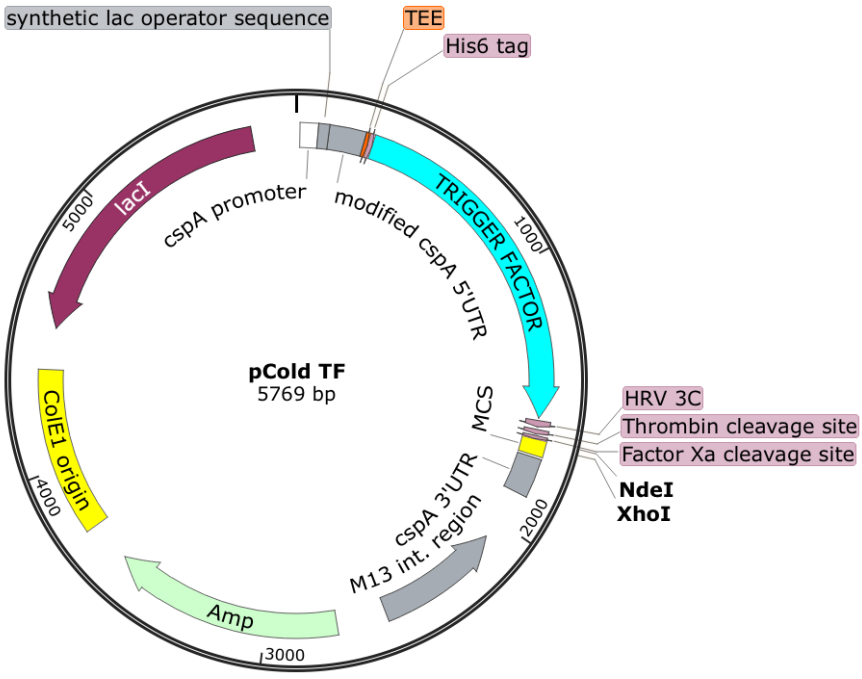


Figure 3.1 | Map of pCold vector. Size: 5769 bp. ColE1 ori: origin of replication of the bacterial plasmid ColE1 (low copy number: 20 copies/cell); Amp: ampicillin resistance gene; lacI: lac operator; M13IG: intergenic region of M13 phage; MCS: multiple cloning site; TEE: translation enhancing element; His6tag: hexahistidine tag. The region coding for the TF tag is located upstream the three cleavage sites (blue segment). The expression cassette is composed by other non coding-elements (in grey). Illustration created with Snapgene free online software (http://www.snapgene.com).

in an allosteric manner, thereby allowing the transcription of the gene under its control (Marbach and Bettenbrock 2012). The structure of the vector is the following: under the control of the promoter (*cspA*) is located a *lac* operator, permitting a strictly regulated induction. At the downstream, subsequently, it is followed by a 5'-untraslated region (5'-UTR), a translation enhancing element (TEE), a hexahistidine tag (6xHis), the TF coding sequence, and the multi cloning site (MCS)(Fig. 3.1). Upon molecular cloning of the foreign gene (in frame with the tag), the target protein is expressed fused at the C-terminal of the TF. The tag can be separated from the target protein, thanks to the presence of three cleavage sites (specific aminoacidic sequences recognised either by thrombin, factor Xa, or HRV 3C protease, Fig. 3.2).

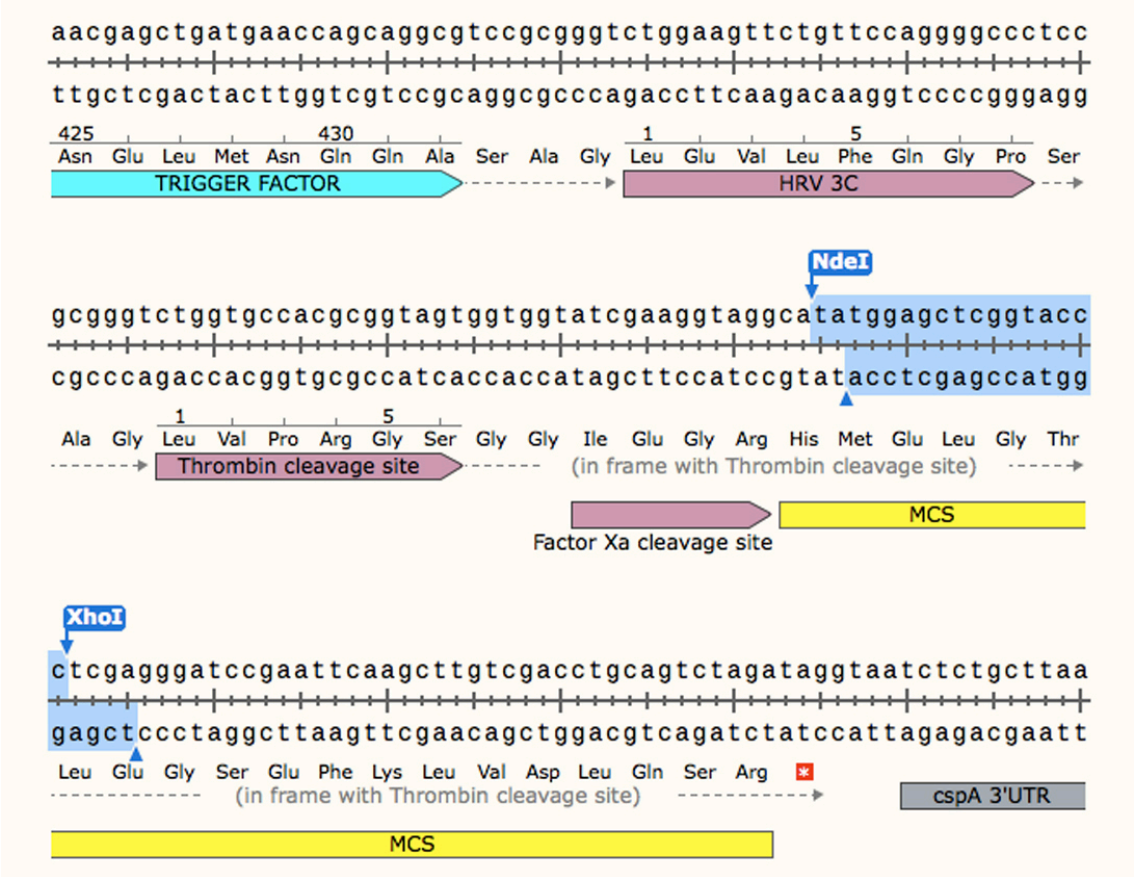


Figure 3.2 | DNA sequence and translation of the plasmid region inserted between the C-terminal of the TF tag and the multicloning site (MCS). The position of the restriction sites NdeI and XhoI and relative sticky ends are highlighted in blue. Illustration created with Snapgene free online software (http://www.snapgene.com).

3.3 RESULTS

3.3.1 Amplification of the cDNA inserts.

Two different DNA fragments (named **P1** and **P2**, corresponding to the full ORF of the silicatein gene and a variant carrying a deletion at the N-terminal, respectively) are amplified by means of PCR (§2.4.7) using as template the cDNA sequence encoding for the precursor of *S. domuncula* silicatein α (accession number AJ272013). The reaction is performed in volumes of 20 µl using the High Fidelity PCR kit (Fermentas), and the following set of cycles: 1 cycle 95°C / 120" + 35 cycles (95°C / 30" + 56°C / 30" + 72 °C / 120") + 1 cycle 72°C / 5'.

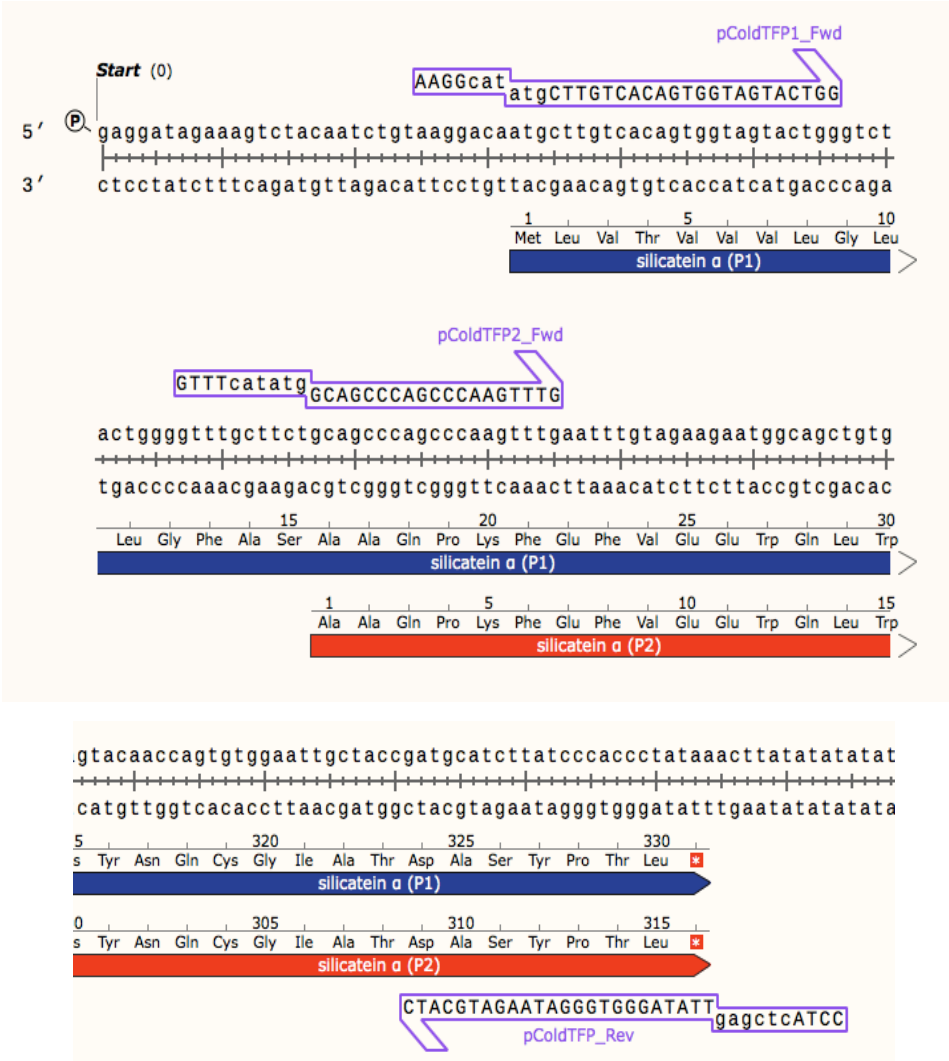


Figure 3.3 | Regions of the silicatein gene (accession number AJ272013) used for the initiation of the polymerisation reaction (PCR) for the production of the fragments P1 (in blue) and P2 (in red) to be cloned in the pCold expression vector. Primers pColdTFP_Fwd1/2 and pColdTFP_Rev are illustrated in purple. Complementary nucleotides are indicated in capital letters. Illustration created with Snapgene free online software (http://www.snapgene.com).

The primers pair pColdTFP1_Fwd and pColdTFP_Rev are selected for the generation of **P1**, while the couple pColdTFP2_Fwd and pColdTFP_Rev for **P2** (Fig. 3.3). The procedure afforded a fragment of 1000 nt (**P1**), and 958 nt (**P2**), delimited by NdeI and XhoI ends. Specifically, the reverse primer used for the amplification of both fragments **P1** and **P2** is complementary to the region of the gene delimited by nt_{1001} and nt_{1024} (the C-terminal of the protein) and introduces an XhoI restriction site (C^TCGAG) at the downstream of the stop codon. At the opposite end (N-terminal), pColdTFP1_Fwd covers the region of the gene from nt_{32} to nt_{56}, whereas pColdTFP2_Fwd the segment from nt_{77} to nt_{95}. Both Fwd primers introduce an NdeI restriction site (CA^TATG) upstream the gene, which also contains the triplet coding for the first methionine of the translated protein.

3.3.2 Generation of the expression vectors.

The PCR products **P1** and **P2** are purified and concentrated as described in §2.4.4. The vector is multiplied through an *E.coli* overnight culture (§2.3.4), and purified (§2.4.5). Restriction digestions of both inserts and vector are performed in one single reaction as described in §2.4.2. After removal of the enzymes (§2.4.4), the ligation reaction is performed (§2.4.3), generating the new expression vectors pColdTF-P1 and pColdTF-P2, with size approximately 6,7 kbp. The reaction mixtures for restriction digestions and ligation are summarised in Tab. 3.1. The

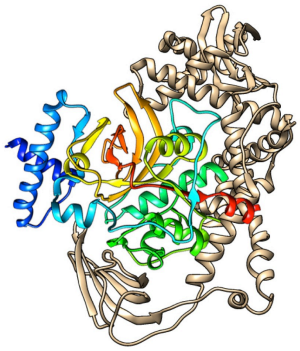
Table 3.1 | Reaction mixtures used for the molecular cloning of amplified P1 and P2 into the expression vector pCold.

Double restriction digestion (overnight, ~22°C)	**Volume (µl)**
Insert P1 or P2 (20 ng/µl)	10
Vector pCold (10 ng/µl)	4
Buffer 4 (NEB)	3
Enzyme NdeI (NEB – 10 U/ml)	1
Enzyme XhoI (NEB – 10 U/ml)	1
dH_2O	11
Total volume	30
Ligation reaction (overnight, 15°C)	**Volume (µl)**
DNA mixture (100 ng/µl)	10
DNA ligase buffer (NEB)	2
T4 DNA ligase (10 U/ml)	1
dH_2O	7
Total volume	20

recombinant expression vectors were verified by restriction digestion tests (DNA electrophoresis, §2.4.1) and DNA sequencing techniques (data not shown).

3.3.3 Structure of the fusion proteins: TF-Ps.

Through the DNA constructions obtained through the procedures described in §3.3.1 and §3.3.2, two versions of immature silicatein-α from *S. domuncula* are produced. The linear structures of the fusion polypeptides are illustrated in Fig. 3.4: the protein named "TF-P1" derives from the translation of the wild type gene ("**P1**") of silicatein-α (330 aminoacids); the second version, named "TF-P2", differs from TF-P1 because of the deletion of the N-terminal fragment of silicatein (from Val_2 to Ser_{15}), corresponding to the N-terminal signal peptide (estimated by ProtParam, http://web.expasy.org/protparam). As already mentioned, silicateins are separated from the fusion tag by an aminoacidic stretch containing recognition sites for proteases (Fig. 3.2). Predicted sizes of the possible cleavage products emerging from the hydrolysis of the TF-Ps are listed in the table enclosed in Fig. 3.4.



Construct	Size (kDa)	Cleavage products	Factor Xa (kDa)	Thrombin (kDa)
TF-P1	88,56	TF1	52,13	51,41
		P1	36,45	37,16
TF-P2	87,16	TF2	52,13	51,41
		P2	35,01	35,72

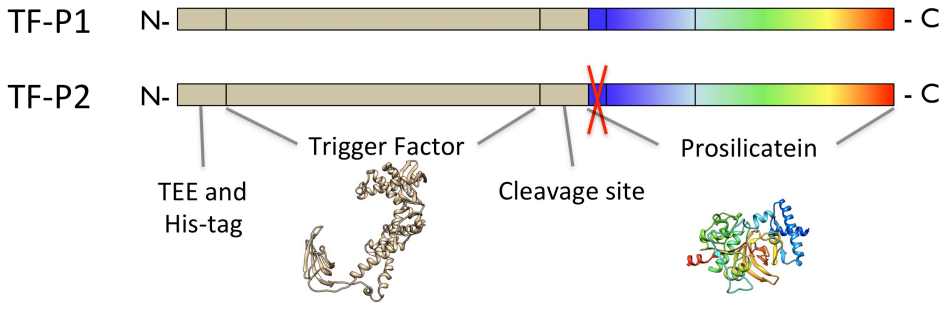


Figure 3.4 | Linear structures of the fusion proteins and sizes of the fragments obtained upon reaction with proteases. The cleavage site contains the recognition sequences for Factor Xa, thrombin and HRV 3C protease. TF-P2 differs from TF-P1 by the absence of the natural silicatein leader peptide (marked with a red cross). From the N-terminal the fusion protein is composed as follows: aa_1-aa_5: TEE element (§3.2.1); aa_6-aa_{11}: 6xHis tag; aa_{12}-aa_{443}: Trigger factor; aa_{444}-aa_{472}: cleavage site; aa_{473} – C-terminal (aa_{786}: TF-P2 / aa_{801}: TF-P1).

3.3.4 Production of TF-Ps proteins.

The plasmids pColdTF-P1 and pColdTF-P2 were used to transform *E.coli* cells (§2.4.8) and replicated by standard procedures (§2.3.4, §2.4.5). Selection of several potential transformant colonies are performed on agar plates by their capability of growing in presence of 50 µg/ml carbenicillin, indication of the resistance to the antibiotic conferred by the vector (Fig. 3.1). Several colonies from a set of strains of bacterial hosts (*E.coli* BL21(DE3)pLysS*, E.coli* BL21 Rosetta, and *E.coli* BL21 Rosetta Gami (Life technologies), as well as *E. coli* Shuffle (NEB) are tested for the expression of the recombinant fusion protein, giving slightly different results in terms of expression profile and protein production rate (data not shown). The strain *E.coli* BL21(DE3)pLysS was selected because it showed the best performance and homogeneous overexpression within a group of different clones from the same strain (Fig. 3.5A). Scale-up of the cell suspensions are obtained to culture volumes up to 150 ml, and a standard protocol for the cultivation, induction and protein extraction has been developed. Optimal production of the fusion proteins are obtained in batches of 150 ml Terrific Broth medium (Roth) containing 50 µg/ml carbenicillin, upon induction (OD_{600} = 0,5 – 0,7) with 0,4 mM IPTG and 40 hours of incubation under agitation (250 rpm) at 15°C, in sterile glass Erlenmeyer flasks of at least 500 ml. Bacterial pellets are collected by centrifugation 30'/5000 rpm/4°C, subjected to one cycle of freezing/thawing (-20°C/r.t.) and resuspended in 20 ml of Lysis Buffer (20 mM Tris pH 8, 150 mM NaCl), protein inhibitors cocktail (Roche). Extraction in native conditions is performed by sonication (5 cycles 15"/50 W on ice avoiding foaming on the surface) followed by lysate clearing through an additional centrifugation step (30'/5000 rpm/4°C). The production of soluble protein is estimated by densitometric analysis in the range of 10 mg of overexpressed fusion protein from a bacterial culture of 100 ml (volumetric yield = 100 mg / L). Negligible amounts of insoluble fusion proteins are detected in the cell pellets after repeated extractions with progressively increasing concentrations of denaturating agents, i.e. urea up to 6 M concentration (data not shown).

3.3.5 Purification of TF-P proteins.

Purification of the TF-Ps are performed by Metal Affinity Chromatography (IMAC) in native conditions as described in §2.6.1 within 48 hours after the extraction from the bacterial pellets. His-tagged fusion proteins showed high affinity and binding specificity to the stationary phase, and relatively pure soluble TF-P1 and TF-P2 could be isolated, as shown in Fig. 3.5B. Optimisation and standardisation of the method guarantees the best recovery of proteins through purification cycles on sample loads of 25 ml of cleared lysate on 5 ml cartridges (Bio-Scale Mini Profinity IMAC Cartridges, Bio-Rad), enabling the recovery of up to 5 mg of pure fusion protein (TF-Ps) in KH_2PO_4/KCl/imidazole buffer from a starting batch of 100 ml of bacterial culture. Improvement of the protein purity of the expressed fusion proteins was attempted by size exclusion chromatography (§2.6.2). A typical purification profile of TF-P1 is shown in Fig. 3.5C as an example.

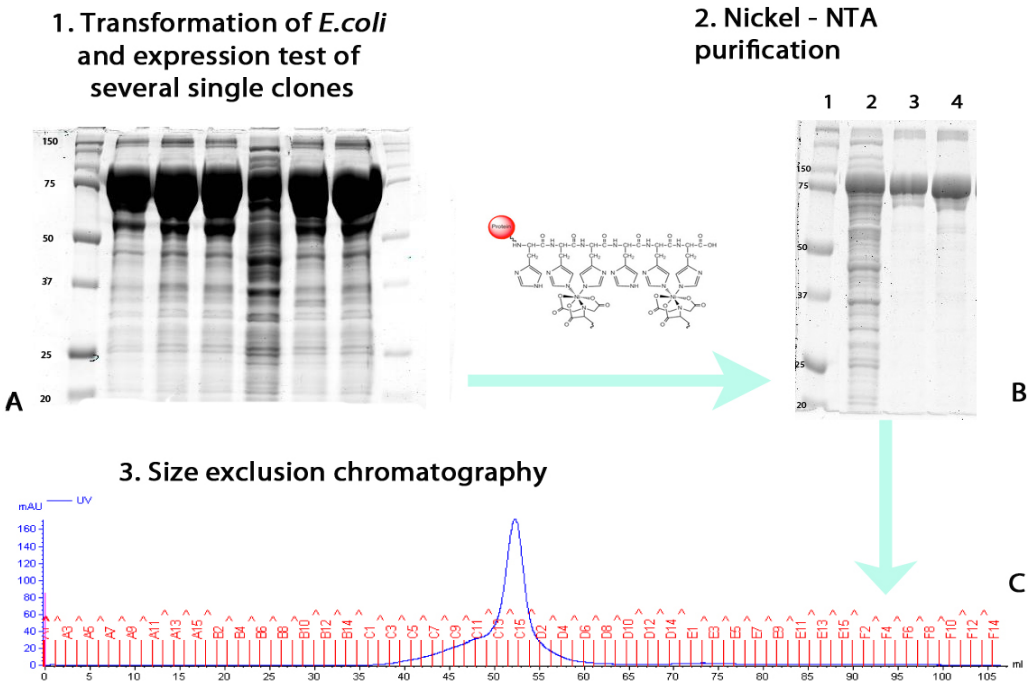


Figure 3.5 | Overview of the expression and purification standard methods implemented for TF-P fusion proteins. A) Expression of 6 different (randomly chosen) clones; B) Purification of TF-P1 by IMAC: Lane 1: Ladder PrecisionPlus (BioRad); Lane 2: clear lysate; Lane 3: Purified TF-P1; Lane 4: Dialysed and concentrated TF-P1 (§2.5.5-6). C) Chromatogram of TF-P1 purified by size exclusion chromatography in buffer 20 mM Tris pH 8 150 mM NaCl.

3.3.6 Cleavage of TF-P proteins.

The purified fusion protein were used as an intermediate to eventually isolate pro-silicatein (and silicatein). In this section the optimisation method is presented in order to afford the desired proteins from the TF constructs. As shown in Fig. 3.4 the fusion protein contains the aminoacidic sequence constituting the target site for the proteolytic activity of three different proteases: Herpes Rhino Virus 3C protease (HRV), Factor Xa (Xa), and thrombin (Thr). The efficiency of the enzymatic cleavage on TF-P1 and TF-P2 of two of the aforementioned proteases is tested (Xa and Thr). Several parameters are also monitored, in order to optimise the cleavage and the recovery of the target protein.

3.3.6.1 Choice of the protease.

In Fig. 3.6 is compared the kinetic of the proteolysis on TF-P1 (Fig. 3.6A and 3.6C) and TF-P2 (3.6B and 3.6D) as substrates of Xa (Fig. 3.6A and 3.6B) or Thr (Fig. 3.6C and 3.6D). The reaction course is followed and monitored by protein gel electrophoresis (§2.5.3). Solubility of the cleavage products is also assessed by centrifugation (60'/13000rpm/4°C) of the sample prior gel loading. We observed that equivalent amounts of protease (in terms of units) have a different effect on the cleavage of the fused proteins, although the reaction is performed at the same values of pH (7,5), temperature (23°C), and substrate concentration (1 mg/ml). Thr is able to cleave almost the totality of the fusion protein TF-P2 within 24 hours, confirmed by the appearance of the separated bands corresponding to TF (52 kDa) and P2 (36 kDa), and the disappearance of the initial band, belonging to the whole chimera (87 kDa). An additional band (24 kDa) is present in the reaction performed by Thr, most likely originating from the maturation process of pro-silicatein by loss of its pro-peptide, leading to the generation of the mature active form (Fig. 1.3). Xa showed a weaker activity, especially on TF-P1 where no cleavage is achieved (Fig. 3.6A). It is also evident that there is different progress of the reaction according to the nature of the protein: TF-P1 shows a slower processing with both proteases. This can only be depending by the presence of the N-terminal signal sequence (15 aa) of silicatein, which seems to have a remarkable impact on the enzymatic

activity of the proteases, probably due to masking of the target site. Moreover, it is noticed from this experiment that pro-silicatein (P2) cleaved with Thr from TF-P2 is not detected in the soluble fraction after 24 hours from the start of the reaction. This suggests a fast process of aggregation, and consequent precipitation, once the TF is separated. In TF-P1 this phenomenon is not observed, although additional experiments confirmed that the precipitation occurs in the same fashion, but requires more time, implying that the aggregation rate is concentration dependent (data not shown). The resolution of the reported gels and the protein amount loaded are not sufficient to visualise the pro-domain (11 kDa), generated from the maturation of pro-silicatein.

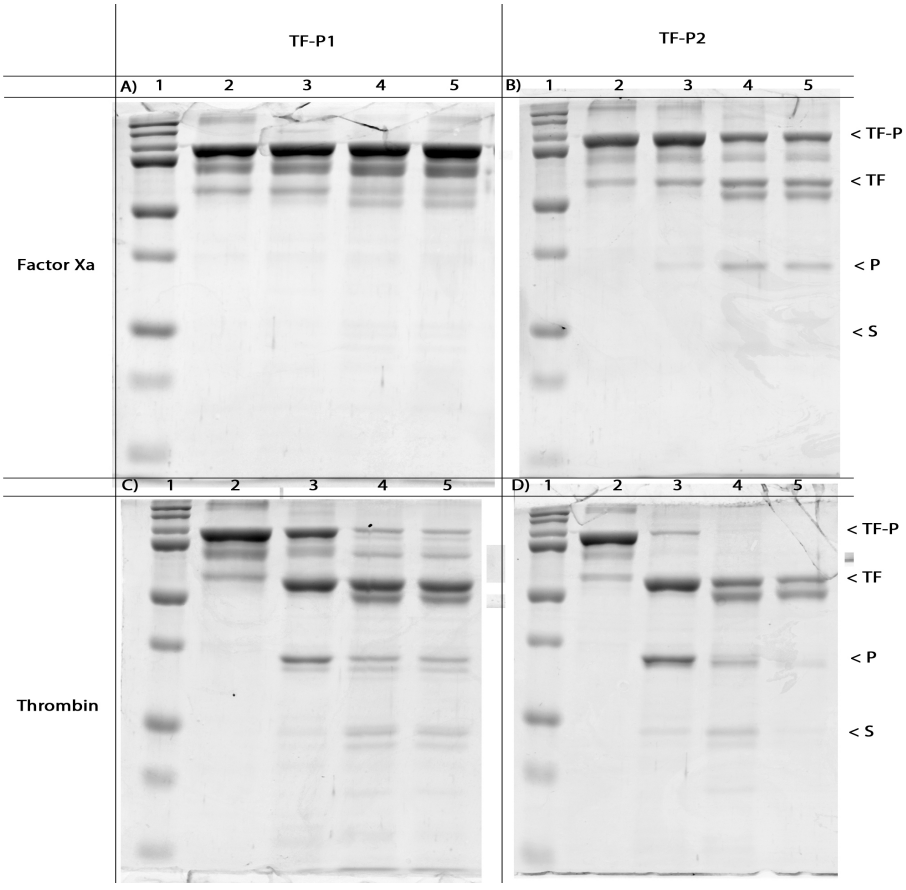


Figure 3.6 | Compared activity of Factor Xa and Thrombin on TF-P1 and TF-P2. Aliquots are collected before addition of 4 µg of each protease (Lanes 2A-D), after 1 hour (lanes 3A-D) and after 24 hours (lanes 4A-D). Centrifugation steps of 60'/13000 rpm are applied on the fraction collected at 24 hours, for the determination of the soluble content (Lanes 5A-D). In Lane 1 is shown the protein marker for the estimation of the molecular size (Bio-Rad Precision Plus, bands size from top: 250, 150, 100, 75, 50, 37, 25, 20, 15, 10 kDa). TF-P corresponds to the complete fusion protein, TF to the free trigger factor, P to the silicatein precursor (either P1 or P2) and S to the mature form of silicatein (24 kDa). Expected sizes of the cleavage products are listed in Figure 3.4. The reaction is performed at room temperature, pH 7,5, and protein concentration 1 mg/ml.

3.3.6.2 Effect of inhibitors and urea

The cleavage of the fusion proteins, especially in the case of TF-P2 leads to the non-specific hydrolysis and degradation of the products. Thus, ways to control the reaction are studied in this section. Experimentally, the effect of additives like chaotropes (urea), or protease inhibitors (PMSF and E-64) is investigated by SDS-PAGE, following the course of the reactions at different time periods. Reactions performed in presence of a defined concentration of additive are by default compared to a control constituted by the solvent used to solubilise the inhibitor. Similar values of temperature, pH, protein concentration, and amount of protease are applied. Urea promotes the denaturation of proteins, destroying the non covalent bonds (Sedlak and Robinson 2009). From our studies we found that, upon proteolysis, soluble P2 is still present in solution containing urea (1 M) after 3 days of reaction, compared to the control, where, instead, aggregation and aspecific hydrolysis of the proteins is noticed (data not shown). From these observations, we attribute to urea the cause of a decreased proteolytic activity of both Thr, and eventually silicatein itself. Moreover, due to its properties as a solubility enhancer, it hinders the spontaneous self-aggregation of silicateins, explaining the longer time in solution.

In a second experimental set, the effect of PMSF, a serine protease inhibitor is studied. Its mechanism of action is well described: it irreversibly inhibits a serine protease through a covalent bond with the hydroxyl group of the catalytic serine, with consequent release of the fluoridic anion. Once it is added in an aqueous solution, its stability decrease to zero within one hour (James 1978). Several experiments are set up, with the task of evaluating the optimal concentration for the modulation of the proteolytic activity. The inhibitor is dissolved in ethanol (concentration of 100 μM). Experimental data (shown in Fig. 3.7) reveal the impact of the inhibitor, that, similarly to urea, when applied at a concentration between 1 and 3 mM, prevents from the uncontrolled hydrolysis of the cleaved pro-silicatein (Fig. 3.7A). Additional experiments showed that excessive amounts of inhibitors cause instant aggregation of all the proteins, probably due to the denaturing effect of the solvent (data not shown).

Additional informations came from experiments of cleavage in presence of E-64, an inhibitor that doesn't affect Ser-proteases (like thrombin), but acts on a broad spectrum of Cys-proteases. Time-lapse characterisation of the fusion protein cleavage in presence of both inhibitors PMSF and E-64, showed a more controlled cleavage, and also the inhibition of the maturation of pro-silicatein into active silicatein. Such maturation event still occurs when only PMSF is added to the reaction mixture (data not shown).

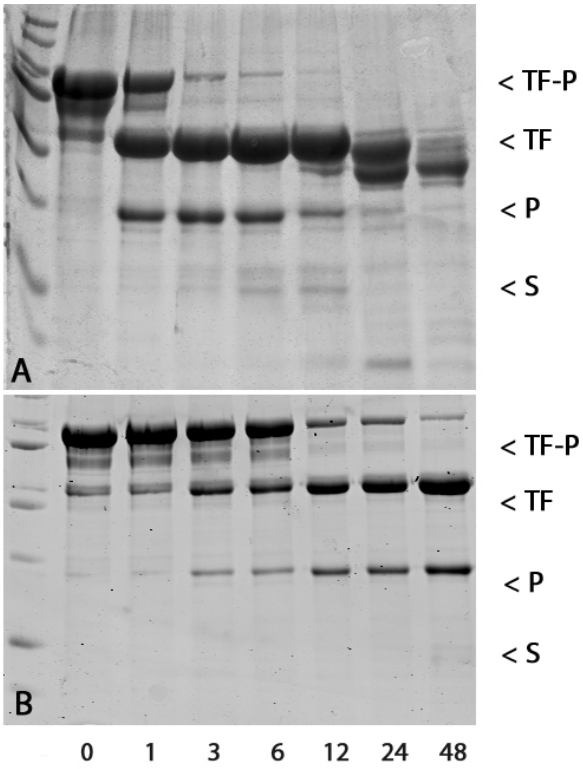


Figure 3.7 | Effect of inhibitors on the cleavage of TF-P2. Aliquots (10 µl) of protein are collected at 0, 1, 3, 6, 12, 24 and 48 hours after the addition of 4 µg of thrombin. The reaction is conducted in standard conditions (panel A) or upon pre-incubation with 2 mM PMSF and 15 µM E-64 (panel B). The notation is similar to Figure 3.6. The reaction is performed at room temperature, pH 7,5 and concentration 1 mg/ml

3.3.6.3 Effect of pH

The pH of the solution constitutes an important parameter to be tested in order to elucidate the characteristics of the activation mechanism of silicatein, and to define an optimal value for performing the cleavage of the recombinant fusion construct without inducing protein degradation. In this set of experiments, on PMSF-inhibited TF-P2 (pre-incubation overnight with 1 mM PMSF), the pH of the solution is regulated adding 100 mM of Tris maleate (prepared in 1 M stocks buffered at

different pH values). The buffering adjustment is performed immediately before addition of the biocatalyst (Thr). Analyses by SDS-PAGE revealed that the protein mixture undergoes remarkable precipitation at pH 5,5 (Fig. 3.8A), suggesting decreased solubility of the TF-Ps and relative post-cleavage fragments at pH close to the isoelectric point value of silicateins. Raising the pH determines a progressively increasing non-specificity in the proteolytic cleavage and reduced stability in solution of P2 (Fig. 3.8 B-F). In absence of inhibition, irrespectively from the pH value, pro-silicatein undergoes degradation (§3.3.6.2). It is observed that Thr showed a remarkable robustness, performing the cleavage of TF-P2 at any pH condition. In conclusion, it is selected as a standard an intermediate value (pH 7,5), which allows specificity of the reaction, and increased stability of the cleavage products.

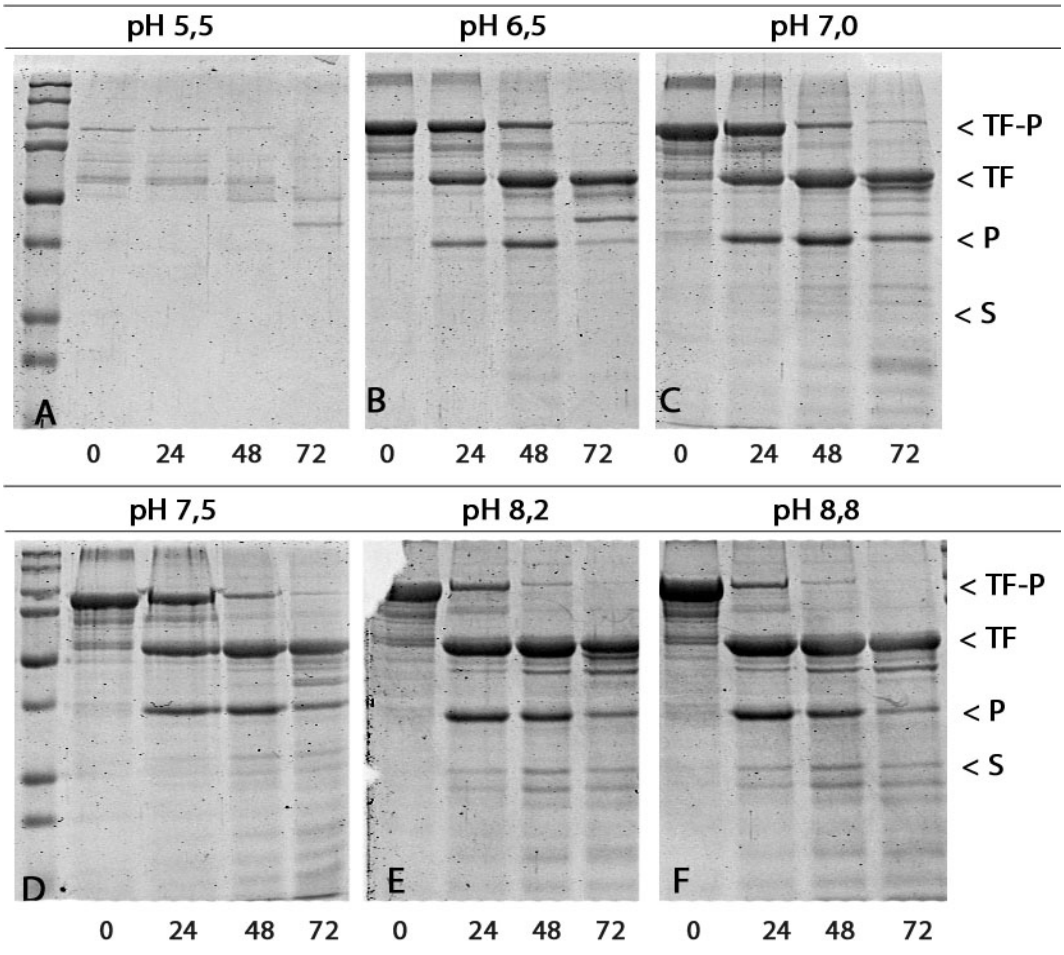


Figure 3.8 | Proteolytic cleavage of TF-P2 at different pH values. Aliquots of protein are collected at 0, 24, 48 and 72 hours after the addition of 4 µg of thrombin. In the panel A-F are presented the reaction at different pH (from 5,5 to 8,8). The notation is similar to Figure 3.6. The reaction is performed at room temperature and protein concentration 1 mg/ml

3.3.6.4 Effect of temperature

The temperature of the cleavage reaction is, together with the pH of the solution, a key parameter to be monitored. Therefore, experiments have been designed to monitor the course of the cleavage reaction on TF-P2 at four different temperatures of incubation. The reaction is performed in the same conditions described previously. In Fig. 3.9 is compared the time-lapse course of the proteolytic cleavage of the fusion protein at temperatures ranging from 0 to 37°C. A negative control (incubation at 70°C) was also prepared, resulting in the inactivation of the protease, and quick denaturation and precipitation of all the protein species (data not shown). Intermediate temperatures (22° and 37°C) show a fast catalysis with consequent appearance of side products and reduced solubility with time (Fig. 3.9 B-C). Surprisingly, when the reaction is performed at 0°C, the protease Thr maintains its activity, leading to almost complete catalysis within 24 hours. After the separation of TF from pro-silicatein, the latter one undergoes a clear maturation process, which at low temperature seems to be possible to control. Both forms of silicatein (precursor and mature) are soluble and stable in water solution for at least 6 days.

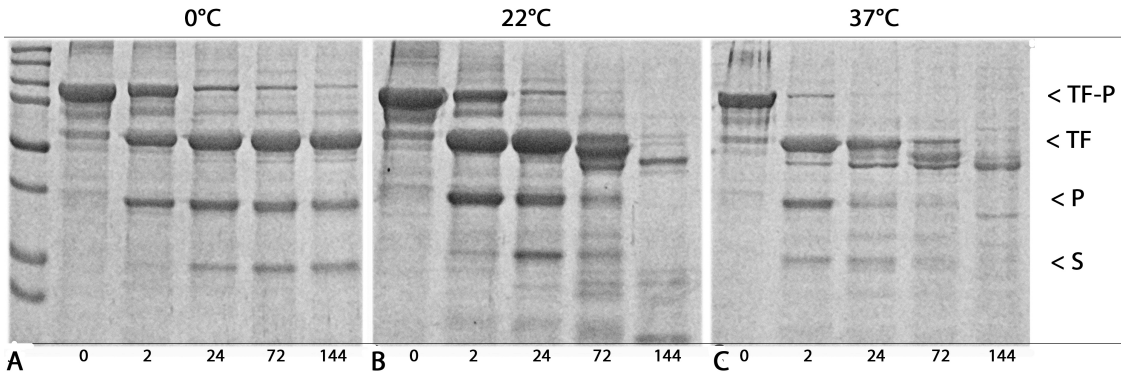


Figure 3.9 | Analysis of the soluble protein content upon proteolytic cleavage of TF-P2 at different temperatures. Aliquots of protein solution are collected at 0, 2, 24, 72 and 144 hours after the addition of 4 µg of thrombin. In the panel A is presented the reaction at 0°C; in panel B at 22°C; in panel C at 37°C. The notation is similar to Figure 3.6. pH of the reaction = 7,5; concentration of the protein = 1 mg/ml.

3.3.7 Crystallisation screening on TF-Ps fusion proteins.

The remarkable solubility and high stability of the uncleaved TF-Ps enabled studies about its propensity to generate protein crystals. High-throughput crystallisation screening is performed at the Hauptmann Woodward Institute (Buffalo, USA). By means of automated liquid-handling systems, one protein solution is tested against 1536 crystallisation conditions (Snell et al. 2008), based on the technique of the microbatch under oil (Chayen 1998). Each condition requires only 0,2 µl of concentrated protein solution. Visual analysis of the crystal growth is conducted on a set of sequential pictures taken at 1, 7, 14, 21, 28 and 42 days of crystal growth. In microbatch experiments, a small droplet containing both protein and precipitant is immersed in inert oil, which prevents evaporation. The principle is that a precipitating agent is instantaneously added to a protein solution, suddenly bringing it to a state of high supersaturation, the fundamental condition for nucleation of protein crystals. Upon purification in native conditions, samples are prepared by procedures of concentration with centrifugal devices described in §2.5.5. Batches of TF-P1 and TF-P2 in Tris/NaCl buffer (pH 8) at protein concentrations up to 15 mg/ml (assessed by the methods described in §2.5.1 and §2.5.2) are obtained. Repeated centrifugation cycles 60'/13000 rpm/4°C are applied to the high-concentration protein samples. No precipitation is observed. The samples submitted to the screening are: TF-P1 at 11 mg/ml, TF-P2 at 8,5 mg/ml, and TF-P2 at 5,5 mg/ml, prepared upon overnight dialysis against excess of buffer 20 mM Tris 100 mM NaCl pH 8. The latter sample (TF-P2 at 5,5 mg/ml) generates the best set of single crystals, while the more concentrated protein solutions lead to a more pronounced amorphous precipitation, masking any crystal growth. In Fig. 3.10 is shown a summary of the diverse morphologies of the crystals obtained, and the corresponding crystallisation conditions.

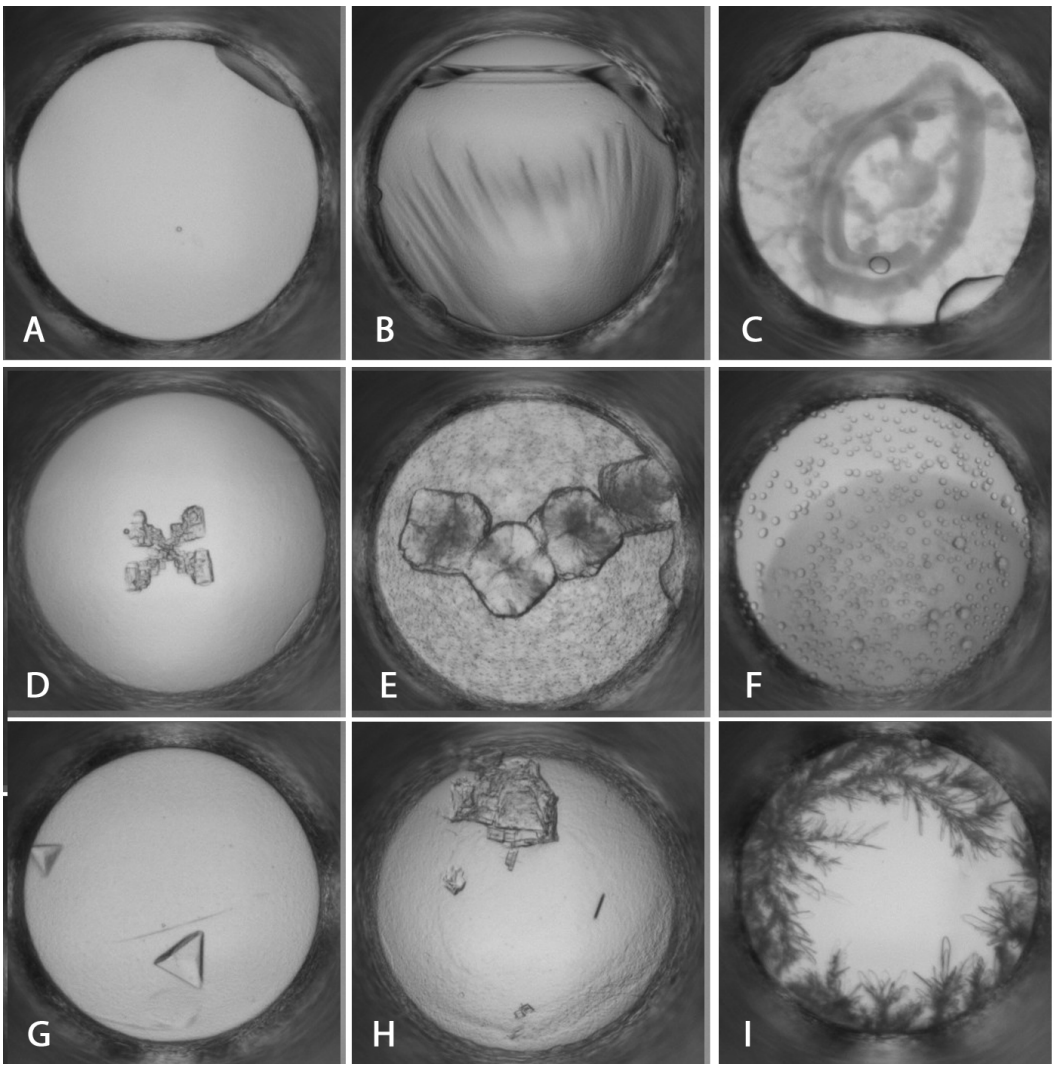


Figure 3.10 | High-throughput crystallisation screening on sample TF-P2 at 5,5 mg/ml (6 weeks crystal growth). The most crucial factor to be studied for the identification of promising crystallisation conditions is the chemical composition of the precipitation cocktail. Herein is summarised the effect of nine different crystallisation conditions, as an example of the possible outcomes of the screening. Pictures are original documentation provided by the Hauptmann Woodward Institute. TF-P proteins are produced (expressed, purified, concentrated and dialysed) in the Institute for Physiological Chemistry (Mainz) and delivered in icebox (0°C) by overnight fast service (DHL, Germany) to Buffalo (New York, USA) one day before the start of the screening.

The most common morphologies observed are:

A. Clear drop (2,32 M KCl 100 mM HEPES pH 7,5)
B. Skin (50% Silver Bullet Bio (HR2-088) A6 + 50% PEG/Tacsimate pH 6,8 (HR2-092) Silver Bullet Bio A6)
C. Precipitate (HR PEG/Ion HT-A6 0,2 M NaCl 20% (w/v) PEG 3350 pH 6,9
D. "Crosses" (3,2 M $(NH_4)_2HPO_4$ 100 mM Tris pH 8,0)
E. "Round edges" (0,1 M K_2CO_3 100 mM Tris pH 8,0 40% (w/v) PEG 8000)
F. Phase separation (0,1 M $(NH_4)_2HPO_4$ 100 mM Bis-Tris Propane pH 7,0 40% (w/v) PEG 1000)
G. Tetrahedric crystals (0,1 M $Li_2SO_4*H_2O$ 100 mM HEPES pH 7,5 40% (w/v) PEG 400
H. Irregular crystal formations (0,01 M Hexadecyltrimethylammonium Bromide 0,5 M NaCl 0,01 M $MgCl_2*6H_2O$)
I. "Branching crystals" (0,85 M K_2CO_3 100 mM CAPS pH 10)

3.4 DISCUSSION

3.4.1 A reliable method for expressing soluble fusion proteins in bacteria.

In this chapter is presented the establishment of a reliable and reproducible method for the expression in high yields of two variants of the recombinant precursor of silicatein α from *Suberites domuncula*. This work, published in the Journal of Biological Chemistry in April 2012 (Schröder et al. 2012a), represents the first attempt to obtain the pre-pro-silicatein (P1) and pro-silicatein (P2) proteins fused to a chaperone tag: the 48,2 kDa trigger factor from *E.coli*. Other fusion proteins containing silicatein have been described in the literature: the group of D. Morse produced mature silicatein α from *Tethya aurantia* coupled to MBP for functional characterisation (Cha et al. 1999), and also the fusion protein silicatein α - OmpA, for peptide affinity screenings (Curnow et al. 2005); more recently, with different purposes and applications, another group reported the production of mature silicatein α fused to GST (Ki et al. 2013). In most of the cases, overexpressed silicateins in *E.coli* were characterised by the accumulation in inclusion bodies, very dense particles with amorphous or paracrystalline structure that may reach sizes of μm, and contain up to 50% of the total cell protein. Although aggregated proteins may maintain a certain grade of secondary structure (Oberg et al. 1994), their complete renaturation require an intermediate step of solubilisation with strong denaturants (urea or guanidinium chloride), resulting in complete unfolding of the proteins. Subsequently, denaturants are removed by dialysis or dilution (Lilie et al. 1998), and depending on the properties of the expressed protein, recovery of the native folding may be a spontaneous process, or induced by a mix of additives and folding agents (Wisser and Gafni 1994). There is no correlation between the propensity of accumulation in inclusion bodies formation and the intrinsic properties (molecular weight, hydrophobicity, folding pathways) of a certain protein; however, it is known that proteins containing disulphide bonds (like silicatein) are not able to fold properly in the reducing environment of the bacterial cytosol. Recovery of functional silicatein-α from inclusion bodies was achieved thanks to the development of methods consisting of dialysis at low concentration against refolding systems containing glutathione redox pairs (GSH/GSSG), L-

arginine and bivalent cations (Müller et al. 2005b; Müller et al. 2013c; Natalio et al. 2010). Structural investigations on refolded *S. domuncula* silicatein α supported by biophysical approaches including circular dichroism and spectrofluorometry, confirmed the recovery of secondary structures (α-helices and β-sheets), and such analyses have been routinely used during the development of the renaturation methods (unpublished). Thus, by using the vector pCold, recombinant fusion proteins TF-Ps were obtained in high yields, distinguished by a high aqueous solubility and stability for extended time of storage at +4°C, with no necessity of performing any renaturation step. Hence, TF-Ps represented good candidates for the production of pro-silicatein samples to employ for further functional characterisation and enabled a preliminary study of crystallisation propensity.

3.4.2 Purification of silicateins before and after TF-P cleavage.

Multiple chromatographic steps are applied successfully to improve the purity of the uncleaved TF-Ps, resulting in minimum protein loss (Fig. 3.5). However, one important task is the isolation of the pre/pro-silicateins (P1 or P2) from the partner tag (TF). Therefore, the proteolytic cleavage of the fusion protein (TF-P1 and TF-P2), is studied extensively, by comparing the activity of two proteases (Thr and Xa). Thus, by assessing the influence of several parameters, like concentration of the "cleaving protease agent", addition of inhibitors, pH and temperature, a method to efficiently cleave the TF-Ps is developed. Subsequently, attempts to separate the partner proteins is also investigated. Due to the fact that the hexahistidine tag (6xH) is located at the N-terminal of the TF, a first approach is to perform, upon cleavage, the binding of the free TF to the Ni^{2+}-NTA resin (Fig. 2.1), and isolate cleaved P1 or P2 in the unbound fraction (substractive chromatography). Through this method it was not possible to have complete removal of the TF from the protein mixture, even after repeated loading cycles. Doubtless, it would have been advantageous to have the 6xH tag at the C-terminal of the fusion protein, directly connected to P1 or P2. Nevertheless, this would have required both the deletion of the N-terminal 6xH inserted in the vector (Fig. 3.1), and the shifting of such element at the downstream of the silicatein gene, requiring additional genetic

manipulation procedures. For different applications, fusion proteins TF-P-like, carrying octaglutamate tags (8xE) at the C-terminal were also produced, additionally to the standard 6xH (data not shown). The high affinity of the 8xE tag for hydroxyapatite (HyAp), can be exploited for performing chromatographic separation of proteins (Schröder et al. 2003a) on HyAp stationary phases. Elution of the bound protein is achieved through gradients of phosphate buffer, acting as a competitive agent similarly to the imidazole used for the IMAC purification procedure. By using this methodology on the 8xE tagged fusion proteins it was still not possible to recover pure P1 or P2 upon proteolytic cleavage of the TF-Ps due to co-elution of free TF. A similar experimental outcome was observed also when the disgested TF-P2 is separated by means of size exclusion chromatography (HPLC), either in native conditions or after denaturations by chaotropes in high concentrations (5 M urea or 4 M guanidinium chloride). Ionic strength and pH modifications of the protein solution showed no beneficial impact on the capability of the system to separate the cleaved silicateins from the TF. Attempts of purifications on desalted (by dialysis, §2.5.6) cleaved TF-Ps with the ion exchange chromatography technique (§2.6.3) revealed that it is possible to recover pure TF when the salt (NaCl) concentration is increased to 200 mM. Silicateins co-elutes at minimum ionic strength of 300 mM NaCl, together with the partner tag. Moreover, consistent loss of the target protein occurs during this procedure (data not shown). In conclusion, while the purification of TF-Ps is a relatively straightforward procedure, the separation of silicatein from the partner tag presents several difficulties. It is assumed that upon cleavage, TF retains strong affinity to silicateins, even if multiple evidences support the capability of TF to prevent aggregation and assist refolding (Kramer et al. 2004; Maier et al. 2001). It should be noticed that although its crystal structure (PDB ID 1W26) has been solved (Ferbitz et al. 2004), the dynamics of TF in solution are largely unknown; therefore, the patterns of association between pro-silicatein and TF after the cleavage need further investigations.

3.4.3 Considerations about the solubility of silicatein.

Through the several analyses conducted in order to optimise the cleavage of the fusion protein, it is possible to observe that both P1 and P2 undergo a maturation procedure (Fig. 3.7-9), originating the short active form of silicatein (S). The latter, upon separation from the N-terminal part (called pro-peptide), shows a phase transition (from soluble to insoluble), leading to a relatively rapid aggregation (Fig. 3.11). The large aggregates are composed mostly of mature silicatein units, as evidenced by SDS-PAGE analyses of urea-solubilised precipitates, in accordance with previous observations (data not shown). In fact, the polymerisation of silicatein α monomers from *Tethya a.* was observed *in vitro* and described as a self-

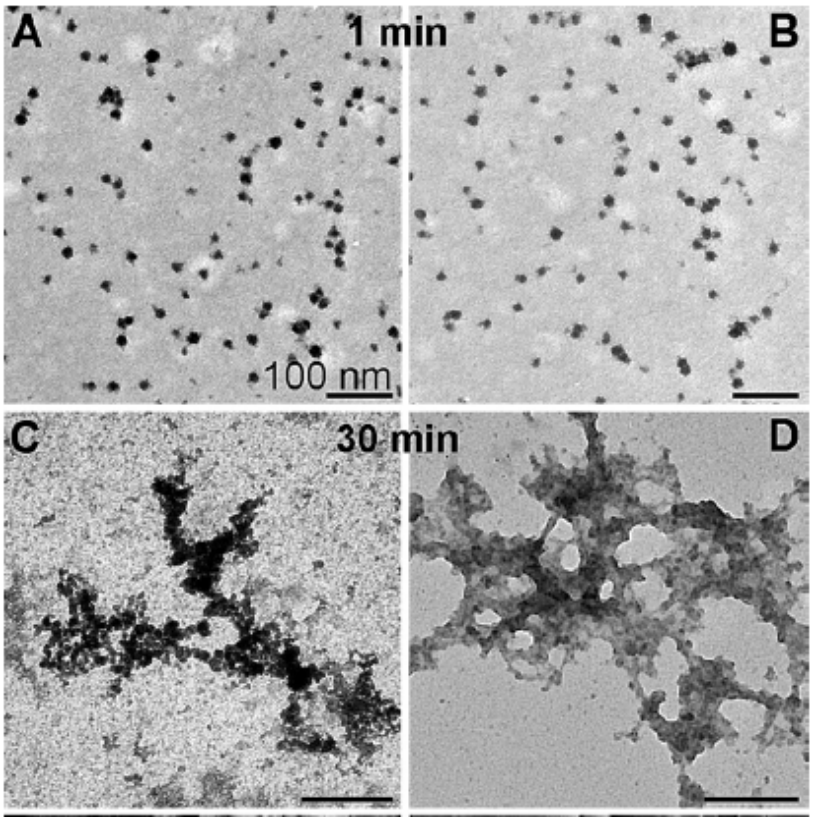


Figure 3.11 | Self-assembly of silicatein. TEM images of mature silicatein obtained 1 min (A); 30 min (C) after cleavage. Same pictures were obtained in experiments of co-incubation of silicatein and silintaphin-1 (ratio 4:1, respectively) (B and D). Size bar: 100 nm. Source of the picture: Schröder et al. 2012a.

assembly reaction, determining the formation of protein filaments fractally patterned, increasing their morphological complexity during time (Murr and Morse 2005). The same was observed for *Suberites domuncula* silicatein, with experimental evidences describing the structure of the protein filaments obtained by aggregation of recombinant mature silicatein α produced in *E. coli* (Müller et al. 2007a). These data explain the impossibility of recover soluble recombinant protein from heterologous expression of the wild-type protein in bacteria. The solubility of proteins in aqueous buffers depends generally on the overall distribution of hydrophilic and hydrophobic aminoacid residues

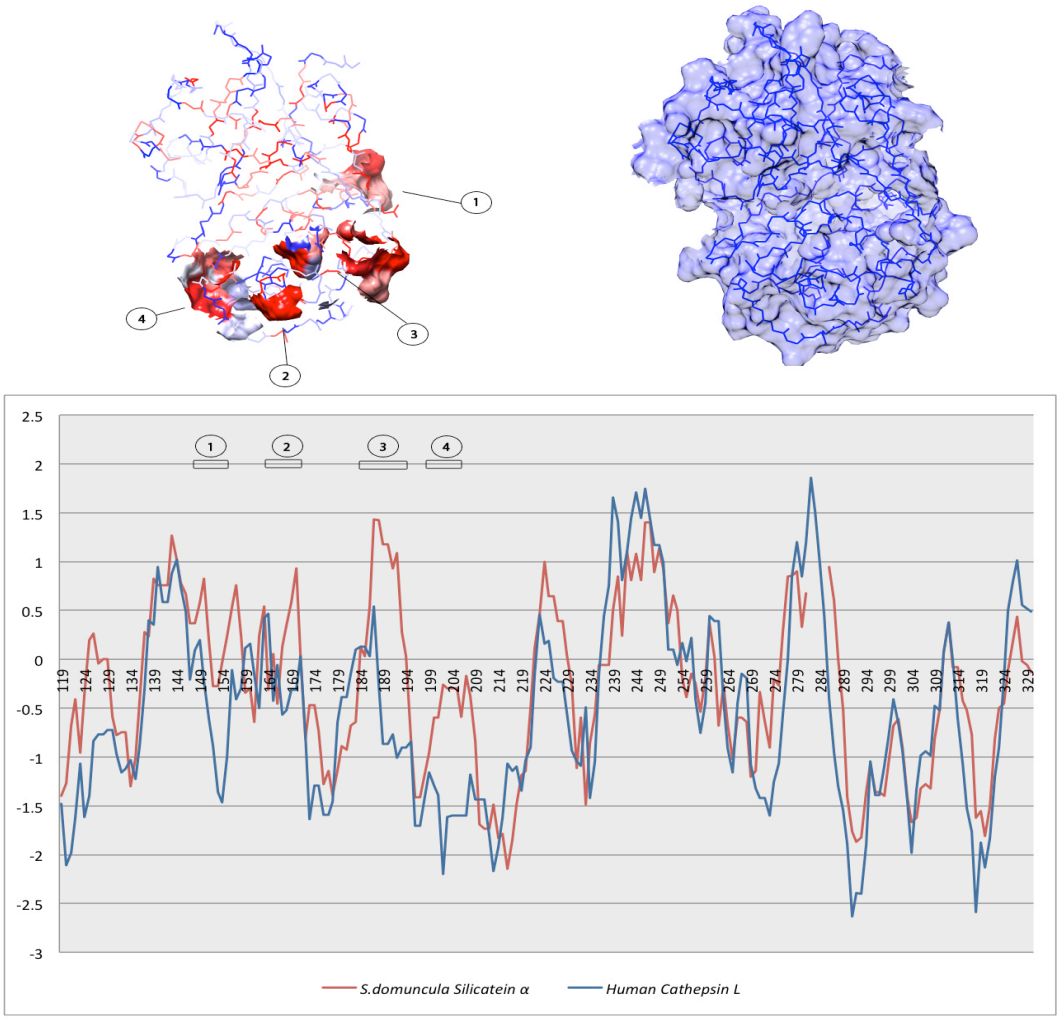


Figure 3.12 | Top: Homology models of mature silicatein based on the structure of cathepsin L. (Left) localisation of the solvent-accessible surfaces of the hydrophobic residues with highest hydropathy value difference (regions 1-4); (Right) Full 3D structure of mature silicatein-α. Hydropathy maps of human cathepsin L (blue) and *S. domuncula* silicatein-α (red), according to the Kyte-Doolittle scale (y-axis). Protein models were created with software Chimera (Salilab, USA). Plot was generated with Microsoft Excel. Illustration was created with Adobe Photoshop.

on the protein's surface. Hydrophobic residues predominantly occur in the globular protein core, whereas charged and polar surface residues increase protein solubility by interacting with ionic groups in the solvent. In some proteins, hydrophobic residues may be present as patches on the surface, usually for the purpose of interaction with other protein surfaces or biological substrates. This may result in self-aggregation by mutual interactions of the protein monomers into ordered aggregates (Akkermans et al. 2007) causing a decrease in aqueous solubility. In the case of silicatein, experimental data exclude a role of intermolecular disulfide bonds (Müller et al. 2005b) in guiding the assembly of the filament. To address the question why a remarkable sequence similarity between cathepsins and silicateins (calculated 50,2% identity; 59,7% similarity) is not translated in similar solubility properties, we focused into identifying the differences in terms of hydrophobicity of the residues. Hydropathy scales are frequently used to map the hydrophobicity/hydrophilicity patterns, ranking aminoacids on the basis of their transfer free energies from nonpolar to aqueous phases. Compared with cathepsin L hydropathy plot, silicatein is distinguished by a general higher hydrophobicity in certain areas, and exhibits at least four regions (1-4 in Fig. 3.12) of low sequence conservation where polar aminoacids (in cathepsin) are replaced by more hydrophobic ones (difference ≥ 1) in the corresponding aligned position. Using specific softwares for protein modelling (Modeller, Salilab UCSF, USA) we generated the 3D molecular structure of silicatein α, using human cathepsin L (PDB ID: 1CJL) as a template because it showed the highest sequence homology (E-value < 5 X 10^{-71}), as shown in 3.12. The silicatein domains having the highest hydropathy difference with cathepsins are clustered within a region of 60 aminoacids at the N-terminal of the mature protein, showing hydrophobic solvent-accessible surfaces in the outer shell. Thus, this may constitute the potential binding region for the assembly of the oligomer, and should be considered for potential mutagenesis approaches in the scope of solubility optimisation.

3.4.4 The maturation of silicatein.

Experiments described in §3.3.6 allowed the observation of the spontaneous processing of silicatein from the zymogen (P) to the active form (S), immediately after the fusion protein (TF-P) is cleaved by a protease. Similarly, the activation of cathepsins is known to proceed in two steps. Initially the N-terminal signal sequence is removed during the passage to the endoplasmic reticulum, together with a N-linked glycosylation; then, maturation occurs after removal of the pro-peptide. The latter acts as a inhibitor to prevent inappropriate proteolytic activity of the zymogen, contributes in keeping the protein folding, and targets cathepsins to the endosomes through a specific mannose-6-phosphate receptor pathway (Hasilik et al. 2009). Crystal structures of human procathepsin L (Coulombe et al. 1996), human procathepsin K (Sivaraman et al. 1999) and human procathepsin X (Sivaraman et al. 2000) showed that the pro-peptide comprises a globular domain (at the N-terminal), followed by a loop which runs through the active-site cleft of the mature enzyme, blocking the access to the substrate. Models of pro-silicatein show the same configuration (Fig. 3.13), and (like most of procathepsins) only hydrophobic interactions, salt bridges and hydrogen bonds are supposed to participate to the association between the pro-peptide and the mature enzyme. The transition from the immature to the active state of cathepsins has been described to be an autocatalytic process, occurring under acidic conditions (Rowan et al. 1992), whereas in other cases it needs the participation of an external protease (Nissler et al. 1998). In addition, if the maturation is autocatalytic, considerations whether the process is a unimolecular (Rowan et al. 1992), or an intra- and inter- molecular mechanism (Menard et al. 1998) should be taken into account. Similar assumptions are valid for silicatein, coexisting in both P- and S-form in the spicule extract, suggesting a tight mechanism of regulation controlling the biomineralisation in sponges, and the existence of a mechanism for the conversion from one form to the other. The experiments described in §3.3.6 strongly suggest that pro-silicatein activation is an autocatalytic process, since the purities of TF-P1 and TF-P2 used in the experiments are high (Fig. 3.5B-C). It is unlikely that other proteases with similar molecular weight have co-eluted with the TF-Ps during the multi-step purifications. Of course a possible objection could be

that since it has been reported that thrombin in certain cases shows limited specificity and may perform hydrolysis on secondary sites (Di Cera and Cantwell 2001), maturation of silicatein is not spontaneous, but mediated by the above mentioned endoprotease. As a matter of fact, neither the sequence of TF nor pro-silicatein contain any thrombin target site, especially in the region of the immature protein where the natural cleavage site of the pro-peptide is located, and in our opinion non-specific thrombin cleavage is an improbable occurrence in our experimental set, that certainly wouldn't allow to observe such a clear maturation event, shown in Fig. 3.9A, but a mix of non-specific fragments. Our assumptions are supported also from the assessment of the role of the protease inhibitors, described in §3.3.6.2. It was reported that the incubation with PMSF effectively reduces the rate of cleavage of TF-Ps, elongating the stability of the cleaved pre-/pro-silicateins in solution, but doesn't seem to interfere on the maturation. In contrast, the protein inhibitor E-64 showed an impact exclusively on the latter process. E-64 is known not to affect Ser-proteases (like thrombin) and earlier investigations reported its affinity with silicatein from *Monorhaphis chuni* (Müller et al. 2008a). Thus, assuming that the maturation event occurs as a single hydrolysis of the peptidic bond (connecting the last aminoacid of the pro-peptide and the first of the mature enzyme), such function should be attributed to the Ser/His catalytic site responsible also for the biomineralisation. Enzymology is actively contributing to the disclosure of the essential biochemical properties enabling proteins to perform reactions of hydrolysis on organosilicon substrates. Important factors are the active site geometry, the presence of hydrophobic environments, and the distribution of the binding sites (Frampton and Zelisko 2012). Notwithstanding the novelty of our findings, the set of conditions that trigger the autocatalysis in pro-silicatein (which likely occur by displacement of the pro-peptide, and exposure of the catalytic site) remains not completely resolved. According to our observations, silicatein activation is only slightly influenced by the pH of the environment, whereas the temperature displays a remarkable correlation with the solubility of the proteins in solution. Eventually, it should be taken in consideration the role of contingent post-translational modifications. The X-ray data of procathepsin B, obtained from crystals naturally grown in living insect

cells, revealed the importance of N-linked glycosides, which mediate several interactions between the propetide and the enzyme (Redecke et al. 2013). Unfortunately, at the present time it is not possible to say whether the native silicateins contain any glycosylation pattern, because of the harsh conditions required for their extraction from the natural sources, although prediction algorithms revealed a consensus polypeptide in the pro-domain. Finally, recombinant pro-silicatein produced in hosts naturally able to perform glycosylation patterns similar to the one produced in the sponges, could give new insights in the dynamics of this complex mechanism (§4).

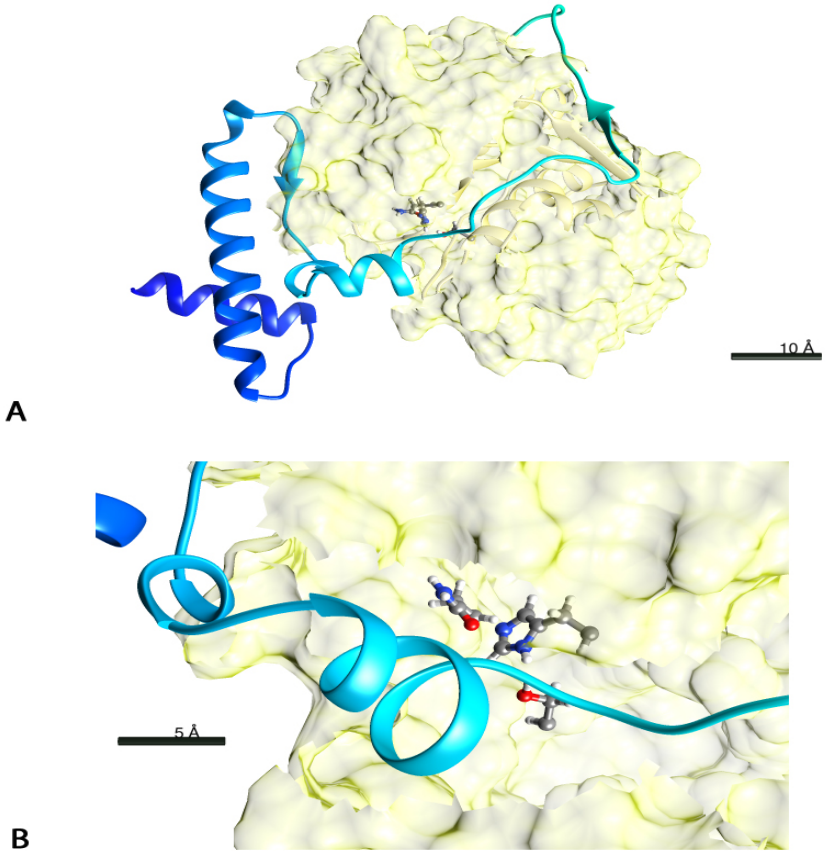


Figure 3.13 | 3D model of pro-silicatein α. A) View of the front side of pro-silicatein. The 97 aa long pro-peptide is constituted by a globular N-terminal domain, followed by a chain that passes through the cleft of the mature enzyme covering the active center. The pro-domain is colored in blue and the secondary structures depicted by helices and beta-strands. The surface of mature silicatein (yellow) is also included to show the general organisation of the protein. B) Zoom on the catalytic site (Ser138, His277, Asn297), in close proximity to one segment of the pro-peptide loop. Illustration created with Chimera software (Salilab).

3.4.5 Advances in the understanding of the formation of the axial filament.

Silicatein obtained after maturation of the precursor, previously fused with TF oligomerises, and flocculates out of solution. This feature, occurring at room temperature, is selective for the mature form, whereas the precursor remains soluble in comparable conditions. Interestingly, the precipitated silicatein could be washed and solubilised in buffers (Tris/NaCl) containing slight amounts of urea (1 M). This preparation was assayed for its ability to polymerise TEOS, by conventional methods already applied for recombinant mature silicatein α, resulting in a remarkable ability to form biosilica, to an extent described previously for the column-purified silicatein (Wang et al. 2011c). We assume that during spicule

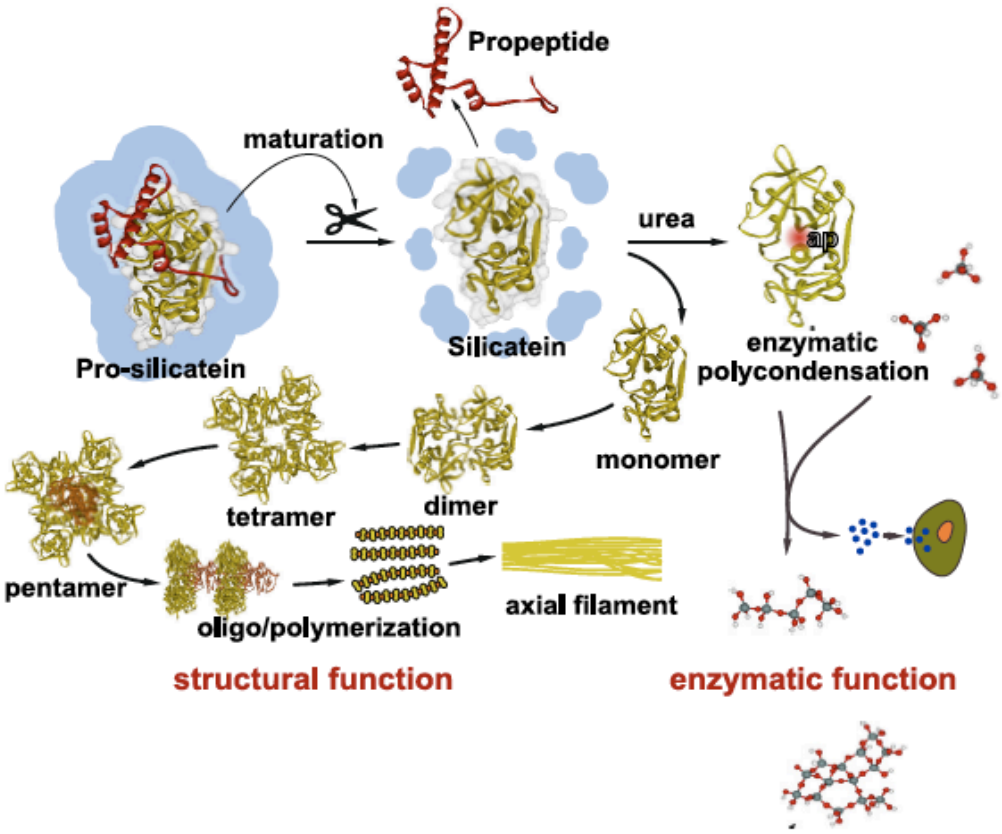


Figure 3.14 | Model of the maturation of silicatein and its supramolecular organisation. The autocatalytic cleavage and release of the N-terminal pro-peptide results in the exposure of hydrophobic patches on the surface of the molecule, resulting in the self-assembly of the mature silicatein monomers of silicatein-α to dimers and tetramers and 4:1 silicatein-α:silicatein-β pentamers which assemble to long filaments. The autocatalytic cleavage of pro-silicatein under release of the N-terminal pro-peptide also results in uncovering of the catalytic pocket (ap) of silicatein, enabling the enzymatic biosilica polycondensation reaction after solubilisation of the precipitated protein. The reaction water that accumulates during the polycondensation has been proposed to be removed by cells via an aquaporin mediated transport mechanism. Figure from Schröder et al. 2012a.

formation, the same transition from soluble precursor to insoluble mature form occurs. While retaining the bio-activity, the silicatein oligomers evolve in a fibrillar organisation (by fractal assembly of supramolecular structures composed of both isoforms of silicatein α and β), which guide the deposition and polymerisation of silica. This event confers to silicatein structure-guiding and structure-forming properties (Schröder et al. 2012a). In addition, experiments of co-incubation with silintaphins suggest that these interactors have an enhancing role in the catalysis. This work was supported for the first time by experimental data of a plausible organisation of the proteins responsible for the building of the sponge biosilica skeleton. A graphic scheme is given in Fig. 3.14.

3.4.6 Structure-forming and structure-guiding activity of silicatein.

The disclosure of the dual functionality of silicatein, encompassing a "structure-guiding" and a "structure-forming" activity (Schröder et al. 2012a), allowed further studies in order to better understand the features and applicability of the protein. The word "structure-forming" is related to the capability of silicatein to perform the deposition of inorganic silica in an enzymatic fashion. Alternatively, the term "structure-guiding" is used to define the competence of the protein to direct the polymerisation in an ordered manner, where a fine-tuned protein-protein interaction mechanism (involving the formation of proteinaceous complexes of mature silicatein α monomers) leads to organised fractally patterned assemblies. Immunobiochemical and immunohystochemical studies revealed that the catalytic activity of silicatein starts immediately after the processing from the immature to the mature form (Müller et al 2005), and the relation between such existing forms of silicatein found in the sponge spicules could be studied extensively through the generation of the TF-Ps fusion proteins. In an attempt to study separately both of the enzymatic properties of silicatein (structure-guiding / forming), recombinant mutants were generated. Synthetic DNA constructs allowed the production of the silicatein-α precursor, where the region containing the cleavage site between the precursor and the mature form of the protein is replaced by a stretch of glutamines (named SILICαNP). Additionally, a mutant of the WT mature protein was designed

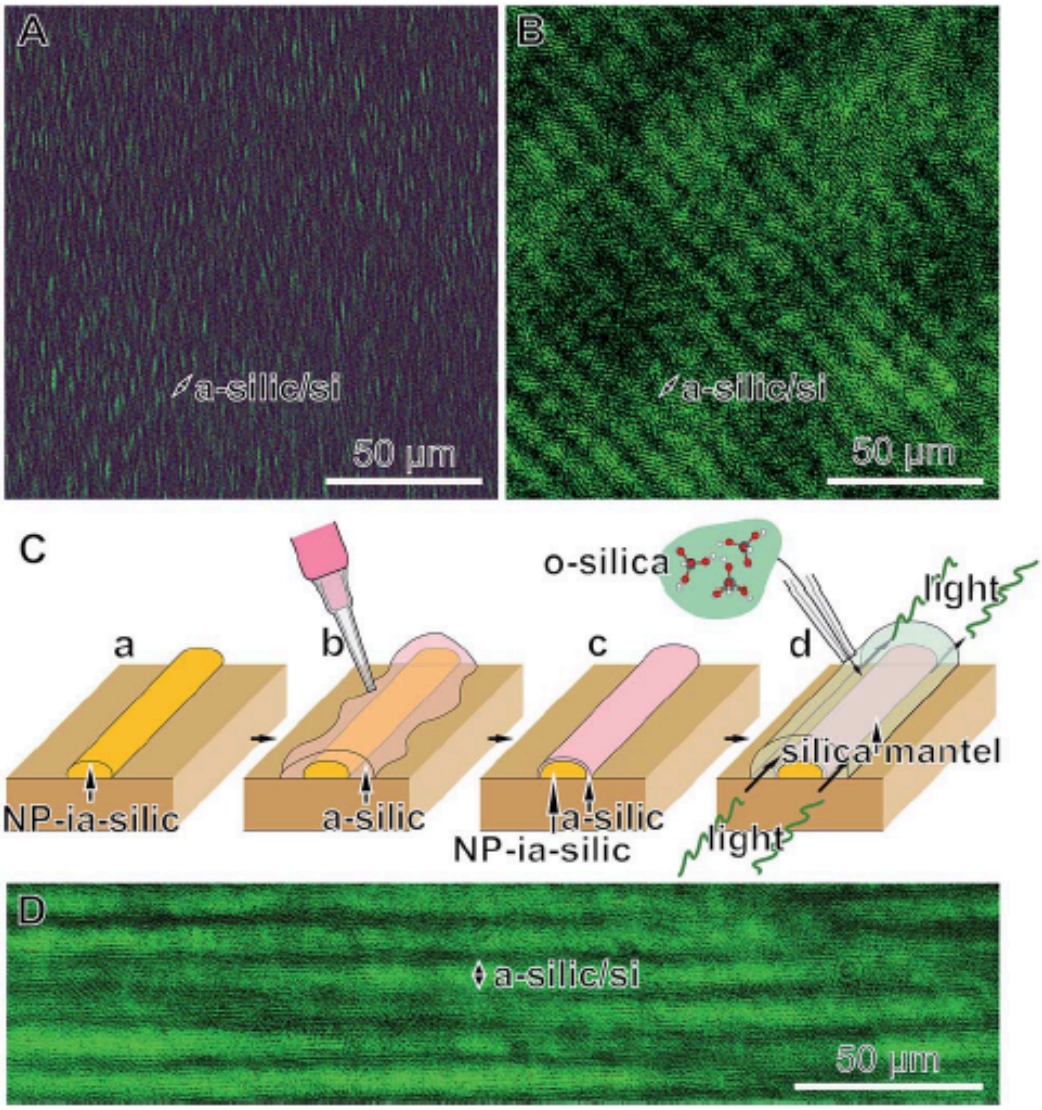


Figure 3.15 | Methods and results of the experiments conducted with silicatein mutants for the bioinspired fabrication of silica-based waveguides (B and D). NP-ia-silic corresponds to the non processed inactive silicatein possessing the structure guiding property (SILICαNP), whereas a-silic is the active form responsible for the silica deposition (SILICαQ/S). Light transmission (from a green laser source) is not detected when surfaces are not exposed to hydrolysed TEOS (A). A schematic outline of the biofabrication system is presented in (C): NP-ia-silic is printed onto gold plates using the micro-contact printing approach (µCP) patterned in parallel lines (500 x 10 µm) (a), and incubated with 20 µg/ml a-silic in Tris buffer for 5 hours at r.t. (b), to form an ordered organic template composed exclusively of silicatein proteins (c); after proteins in excess are removed through TBS buffer washing steps, the surfaces are bathed in solution 200 µM prehydrolysed TEOS for 14 h at r.t., resulting in the formation of a silica mantel showing the ability of transmitting light (d) in correspondance of the layered proteins (a-silic/si), a clear reproduction of the natural sponge spicules light guiding properties (Wang et al. 2011b). Optic fibres up to 800 µm have been fabricated (D). Characterisation of the organic matrix and the silica shell is performed by immunocomplex visualisation and EDX analyses, respectively (Müller et al., 2014b).

with the substitution of the poly Ser-stretch (aa_{267} – aa_{274}) also with a sequence of Gln (named SILICαQ/S). SILICαNP's use benefits of the advantage of being resistant to the auto-activation seen in the wild type pre-pro- (P1) and pro- (P2) silicateins described in §3.3.3 and Fig 3.6. On the other hand, SILICαQ/S is characterised by an improved capacity of mediating protein-protein associations (both homo- or hetero-typic). An enhanced biosilicification activity is also observed when SILICαQ/S is tested in presence of inert protein interactors (like BSA). Therefore, the two mutant proteins were employed in procedures of microcontact printing (μCP), with the goal of producing bioinspired optical fibres (Müller et al. 2014b). Briefly, inactive structure-guiding silicatein (SILICαNP) is printed on a gold surface in a pattern of parallel rods, followed by incubation with active structure-forming silicatein (SILICαQ/S). Upon addition of prehydrolised TEOS, a biosilica mantel is deposited on the organic template. Because of this ordered architecture of the organic matrix driving the silica deposition, the fabricated fibres show the capability of acting as light waveguides (Müller et al. 2014b). Notwithstanding the new horizons in the field of the production of optical fibres with improved characteristics under physiological conditions, this work represents an elegant strategy of protein engineering, that allowed the elucidation of new properties of the silicatein enzyme. In particular, the role of the pro-domain (present only in SILICαNP and not in SILICαQ/S) in the control of the disposition of the active protein is highlighted. Moreover, the replacement of the Ser-stretch resulted in an altered enzymatic activity in comparison with the wild-type proteins.

3.4.7 Are TF-Ps good candidates for "structure-to-function" studies?

Given the importance of silicatein and its role in the intricate mechanism of spicule formation, there are no doubts that the determination of its three-dimensional structure could be an important advance in the understanding at a molecular level of many features that presently may be interpreted only by *in silico* models, and comparisons with the well-known cathepsins. Luckily, such models can be considered very reliable (low RMSD) because of the remarkable similarity (up to 50% of aminoacid identitites) between cathepsin and silicatein, and the availability of several 3D structures of cathepsins. Primary structures comparisons highlight

the presence of regions with consistent level of aminoacidic conservation, suggesting common phylogeny and possible related biological activity. Through the same approach, differences in the aminoacidic sequence have been used to identify potential clusters distinctive in silicatein, and presumably involved in the unique biomineralising activity (e.g. the Ser- cluster). Nevertheless, X-ray diffraction of single protein crystals is considered the most robust depiction of the conformation of a macromolecule, resulting in the determination of the coordinates of every single atom in the space. Thus, the crystallisation of silicatein still represents a thrilling challenge in sponge biology. At the time of this writing, only two structures solved by X-ray crystallography of proteins from Demospongiae are deposited in the Protein Data Bank: geodin, a βγ-crystalline from *Geodia cydonium* (Vergara et al. 2013), and a tetrameric galectin from *Cinachyrella* sp. (Freymann et al. 2012). Because of the inherent properties of silicatein, single crystals of pure protein are unlikely to grow. In fact, success in the crystallisation strongly depends by the nature of the protein, and by its purity, monodispersity, high concentrations, and correct folding. For cytosolic proteins, solubility in aqueous solutions is mandatory for the correct nucleation, while membrane proteins require solubilisation in suitable detergents, able to mimic the phospholipidic layer typical of the cell membrane. In the literature, the impact of fusion tags in aiding crystallisation of proteins is extensively described. Advantages derive from solubility and folding improvements, mediation of crystal packing contacts, and facilitation of the structure determination of the target protein by molecular replacement methods (reviewed in Dale et al. 2003). From our work, TF-Ps were characterised by an extremely high solubility and stability of several months at +4°C, thus it was worthwhile to assess their propensity to crystallise through a high-throughput crystallisation approach. Notwithstanding the advantages of this platform, allowing the screening of thousands of crystallisation conditions within a time window of less than two months and with few milligrams of pure protein, the identification of hits is limited to visual analyses of the crystals morphologies, which need to be reproduced and verified in house for X-rays analyses. In fact, only by the latter technique, it is possible to elucidate whether a crystal grows exclusively from the target protein or from the solutes contained in

the crystallisation cocktail, instead. Moreover the limited understanding of physico-chemical conditions regulating protein crystallisation is impeding the development of any prediction method (Chayen 1998). Nevertheless, some hits from the HT screening showed growth of potential TF-Ps single crystals (Fig. 3.10). This suggests that the TF tag plays a beneficial role not only as a "solubility enhancer", but also in the promotion of TF-P crystallisation. Silicatein protein crystals have not been reported in the literature so far, whereas several cathepsins' structures have been elucidated through the successful generation of a diffractive protein crystal. It has been already discussed (§3.4.3) that although silicateins show a remarkable similarity in primary structure to cathepsins, the two protein families present differences in terms of solubility in aqueous solutions, and especially in biological functionality. Therefore, in the scope of growing diffractive silicatein crystals, the application of experimental strategies inspired by the successful outcomes from the cathepsins structural studies is not sufficient. New strategies have to be implemented, and our work on TF-Ps crystallisation is to be considered as a progress towards the elucidation of structures of important proteins from the marine world.

4 EUKARYOTIC SYSTEMS FOR THE PRODUCTION OF RECOMBINANT SILICATEIN

In this chapter is presented a comparative study aimed to assess the suitability of two eukaryotic hosts for the production of silicatein proteins: the insect cells and the yeast cells. The choice of the appropriate expression hosts is critical and determines the quantity and the quality of target protein produced. This is frequently a process that requires optimisation in terms of yield, folding, and solubility of the produced proteins. In the following, the initial phases in the development of the bioprocess for the functional production silicatein in eukaryotic systems is discussed.

4.1 HETEROLOGOUS EXPRESSION OF PROTEINS IN INSECT CELLS

At the time of this writing the interrogation of popular scientific literature databases (PubMed, Web of Science, etc.) using the term "Baculovirus" yields over 3000 hits, almost half of them published in the last decade. Although this is a significant number, it certainly underestimates the real extension of studies involving recombinant protein production based on baculovirus–insect cell expression systems, which includes investigations performed behind closed doors in the biotechnology industry, and experimental efforts not mentioned in the published literature. General agreement exists in attributing considerable importance to this expression system because of its capacity for high-level protein production, its ability to provide eukaryotic protein modifications, its biosafety, and its relatively moderate cost. High expression rates refer in average to yields above 100 mg of recombinant protein per litre of infected insect cell culture, although

amounts and solubility strongly depend on the inherent properties of the target protein, as well as its subcellular localisation, as cytosolic proteins are generally less arduous to express in comparison with membrane and secretory pathway proteins (Jarvis 2009).

4.1.1 Elements of Baculovirology

Baculoviruses are the most prominent viruses known to affect the insect population, causing a disease called polyhedrosis. They are taxonomically divided in two major groups based on the morphology of their capsid, also called occlusion body (OB): the nucleopolyhedroviruses (NPV), characterised by single or multiple crystalline rod-shaped nucleocapsids, and the granuloviruses (GV), so called for the presence of granular, ellipsoidal-shaped OBs. Both groups belong to the family of the Baculoviridae, and each species is named for its host. Particular interest in biotechnology is holded by *Ac*MNPV, the baculovirus associated to the alfalfa looper *Autographa californica* (*Ac*). Its genome is a double stranded, circular, supercoiled DNA of 134 kpb, packaged in bundles of 5 to 15 rod-shaped nucleocapsids that are 230-385 nm in lenght and 40-60 nm in diameter. The virions (constituted of protein shell and viral DNA) are packaged in polyhedra, highly stable elements that can maintain indefinitely the infectivity of the virus in most normal environmental conditions, thus allowing it to survive outside its host, especially when the latter has a seasonal cycle. Polyhedra have a diameter of 0,6-2 μm, and consist of a crystalline matrix composed of a protein called polyhedrin. These 30 kDa polyhedrin subunits form tightly packed trimers, which then arrange in dodecamers via disulfide bonds. This structure interlocks with another dodecamer to form the cubic-shaped unit cell of the crystal. Hydrophobic and salt bridge interactions between the cubes likely constitute the linkages at the crystal interfaces. A glycoproteic envelope (PE) surrounds the polyhedra and provides stability to the OB. The infection occurs when the insect ingests the OBs, and they reach the mid-gut, one of the three sections of the gastrointestinal trait of the insect. This is characterised by a gradient of pH from 7.0 up to 10.0 in the central region, and is the major site of food digestion. The alkaline environment, combined

with the action of specific proteinases, causes the disruption of the PE, the dissolution of the polyhedra and the release of the OBs into occlusion-derived virus (ODV). ODVs exhibit on the surface a complex glycoproteic envelope that permits the fusion with the membranes of the epithelial cells, the major cell type lining the midgut, and the release of the nucleocapsids into the cytoplasm. The virions are transported into the nuclei through the nuclear pores, by a mechanism involving the polymerisation of actin. Here the early infection phase takes place, which employs enhancers and transcriptional activators guided by the RNA polymerase of the host. Once the viral DNA is replicated, it is assembled with the capsids, and buds from the cell carrying a proteo-lipidic envelope obtained from the nuclear membrane. In the form of budded virus (BV), it propagates the infection from the basal side of the epithelium to the other cells of the insect (defined as systemic

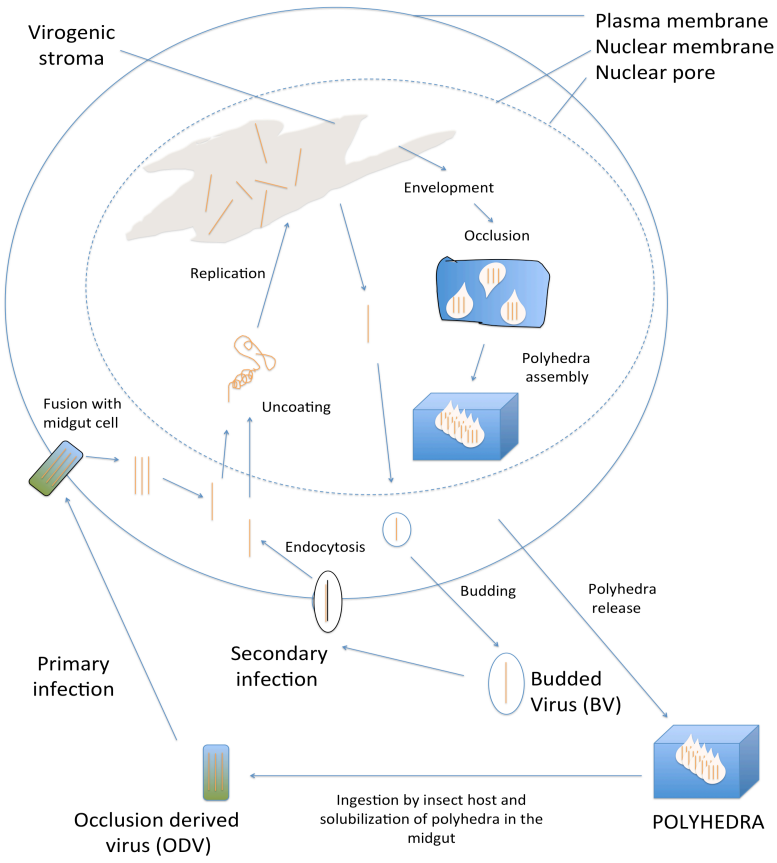


Figure 4.1 | Scheme of the replication cycle of the Baculovirus in nature. The Bac to Bac expression system allows to express a desired protein under the strong promoter originally controlling the expression of polyhedrins. Picture created with Microsoft Powerpoint.

infection or secondary infection). In a later stage of infection, under the control of the viral RNA polymerase, polyhedrin and the other components of the OBs are overexpressed, and packed with the viral DNA. The cells of the insect are blocked in the G2/M phase of the cell cycle, and induced to undertake an apoptosis program, which causes shrinkage, nuclear fragmentation, chromatin condensation and final disintegration. Eventually, the replicated OBs are released in the environment with the death of the insect. A schematic overview of the Baculovirus reproduction cycle is shown in Fig. 4.1.

4.1.2 The Bac-to-Bac system

Implementation of heterologous expression in insect cells is a multi-step procedure that starts with the creation of a recombinant virus. Several systems have been developed to serve a variety of different purposes, and to address the technical constraints associated to: the isolation of recombinant baculovirus, the efficiency of integration of the foreign gene, the reduction of the gene loss within the replication rounds, and the quality of produced protein. A system with improved efficiency in comparison with the previous approaches based on homologous recombination is the commercialised Bac-to-Bac. It relies on an efficient molecular mechanism of genetic transposition to produce recombinant baculovirus DNA (Luckow et al. 1993). The first step consists in cloning the genes encoding for the target protein (almost always a cDNA sequence) into a suitable vector. We used the 4,8 kbp transfer vector pBac, a customised version of the commercial plasmid pFastBac1 (Life Technologies). The expression cassette is delimited by the left (Tn7L) and right arm (Tn7R) of the bacterial transposon (Craig 1991), which enables site-specific transposition of the foreign genes into the baculovirus genome (bacmid). The transfer vector contains the following relevant elements: the *Autographa californica* multiple nuclear polyhedrosis virus (*Ac*MNPV) polyhedrin promoter for high level expression in insect cells (PH); an additional cis-element of 21 nucleotides positioned at the downstream of PH coming from the 5'-UTR region of the lobster tropomyosin, that showed to enhance the

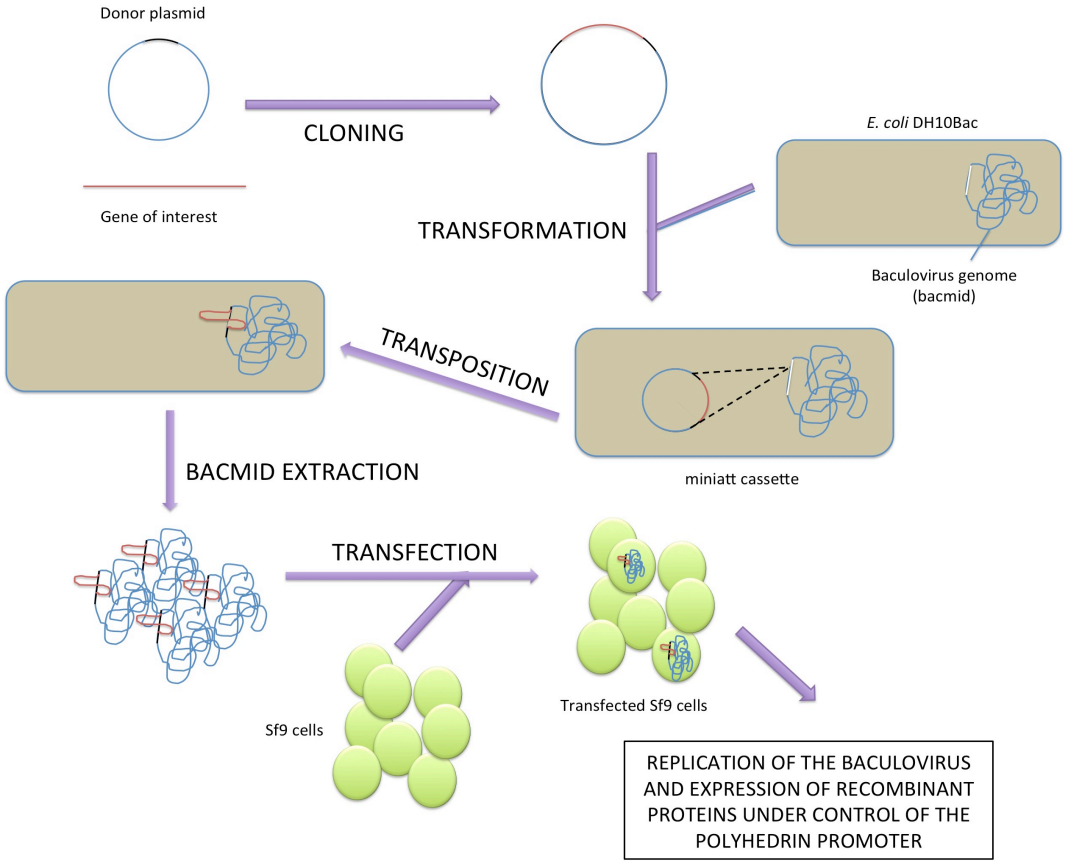


Figure 4.2 | Overview of the experimental steps for the establishment of the insect cells expression system. Silicatein genes are cloned into the transfer vector, used to vehicle the genes to the viral genome (bacmid), contained in *E.coli* DH10Bac. The bacmid is extracted, purified and transfected into insect cells. By cultivation in suspension high titre recombinant Baculovirus is amplified, and employed for protein expression. Picture created with software Microsoft Powerpoint.

expression of exogenous protein in Baculovirus-infected insect cells (Sano et al. 2002); the multicloning site (MCS) and other important features, listed in Fig. 4.3. The generation of the recombinant baculovirus occurs by transposition from transfer vector (donor) to the baculovirus shuttle vector (acceptor). The reaction is performed in *E.coli* (strain DH10Bac, Invitrogen), equipped with the 136 kbp bacmid vector bMON14272, and the 13,2 kbp helper plasmid pMON7124, which encodes for the specific transposase responsible for the gene transposition into the polyhedrin locus of the bacmid. The bacmid is an autonomously replicating plasmid, which includes a cloned copy of the entire baculovirus genome. The DH10Bac bacteria transformed with the recombinant transfer vector are identified

by blue/white selection, as the insertion of the foreign DNA disrupts the gene LacZα present in the bacmid, which complements a deletion of the gene previously induced in this special bacterial strain. In this way, the true positive colonies are not able to convert the substrate X-gal into the blue formazan salt and permit to isolate efficiently only the recombinant bacmids (white phenotype). These, upon purification, are transfected into insect cells resulting into the production of recombinant viral progeny. The main concept of this approach is to replicate *in vitro* the infection cycle outlined in §4.1.1 (Fig. 4.1), and using the strength of the polyhedrin promoter to induce overexpression (Fig. 4.2). This feature allows the production of large amounts of virtually every target protein, because the polyhedrin gene is not required for the replication of the virus in cultured insect cells. Translated into practice, insect cells are used as biofactories for the heterologous production of silicatein and pre-pro-silicatein (Fig. 4.6), overexpressed as a chimeric gene (polyhedrin promoter plus sponge genes) but following the natural mechanisms of the Baculovirus infection.

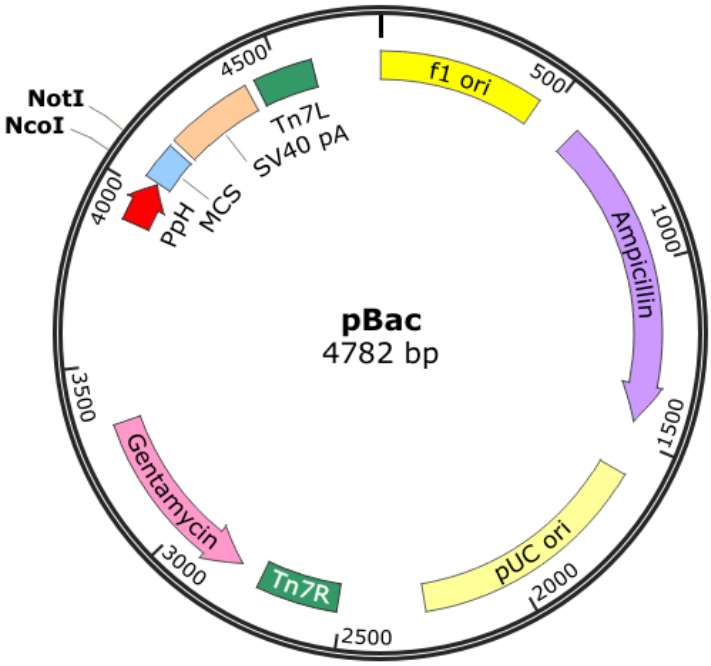


Figure 4.3 | Map of the transfer vector pBac. Tn7L and Tn7R: trasnposition sequences; P_{PH}: polyhedrin promoter; MCS: multi cloning site; Gentamycin: resistance gene; Ampicillin: resistance gene; SV40: polyadenilation signal; two origins of replication, one from f1 bacteriophage (f1 ori) and one pUC-type (pUC ori). Map created with Snapgene free software. (http://www.snapgene.com)

4.2 HETEROLOGOUS EXPRESSION OF PROTEINS IN THE YEAST *P. PASTORIS*

The use of the methylotrophic yeast *Pichia pastoris* as a heterologous protein expression system went through an exponential growth in the past two decades, from the decision of the proprietary corporation (Philips Petroleum Company) to release it in 1993 to the academic research laboratories. Numerous examples have appeared in the recent literature, describing the capability of the yeast to produce and secrete different proteins possessing the expected biological properties (Damasceno et al. 2012). The success in the use of this expression system is also testified by the first approvals and commercialisation of yeast-derived biopharmaceuticals in the European and American market, like Kalbitor® and JETREA® (Vogl et al. 2013).

4.2.1 Yeast biology.

Yeasts are eukaryotic microorganisms belonging to the kingdom of Fungi, known to mankind for at least 4 millennia and representing very early example of "traditional biotechnology", as reported by archaeological evidences of ancient Egyptian ruins of bakeries and breweries. The majority of the species are unicellular, with average diameters ranging from 3-4 to 40 µm, taxonomically divided in two phyla: Ascomycota and Basidiomycota. Yeasts reproduce mostly asexually by asymmetric division (budding) during the vegetative growth, although sexual cycles are undertaken under stress conditions with formation of meiotic spores. Yeasts are chemoorganotrophs, as they use sugars, alcohols, and organic acids as source of energy. Metabolic activities include aerobic respiration (obligate aerobes), or anaerobic fermentation (facultative anaerobes). Yeasts are well known organisms and represent a common model organism in biological research, as well as an important tool in many biotechnological and industrial applications (production of alcoholic beverages, baking, nutritional supplements, and bioremediation). The full genome of the yeast *Saccharomyces cerevisiae* was the first eukaryotic chromosomal DNA fully sequenced (Walsh and Barrell 1996).

4.2.2 The yeast *Pichia pastoris*.

Pichia pastoris is one of approximately a dozen yeast species able to metabolise methanol as a carbon source. The first step in the metabolic pathway is the oxidation to formaldehyde, with concurrent generation of peroxide, by the enzyme alcohol oxidase (AOX). The poor affinity of the enzyme for molecular oxygen (O_2) is a deficiency that the yeast compensates by synthetising large amounts of the enzyme (Cregg et al. 2000). Expression of the enzyme is tightly regulated at transcriptional level by methanol, and its accumulation in the cells reaches 5-30% of the total soluble protein content, depending on the growth conditions. Thus, the promoter of the *AOX1* gene (P_{AOX1}) has been selected as a tool for driving expression of recombinant proteins, as the production phase can be uncoupled by the biomass growth phase, a particularly important advance in case of toxicity of the heterologous proteins. In fact, P_{AOX1} is strongly repressed when *P.pastoris* is grown on various carbon sources, such as glucose, glycerol, or ethanol (Inan and Meagher 2001). An additional feature of the yeast is the ability to secrete proteins into the medium, which can improve the purification steps, as the culture medium contains no added proteins. Moreover, of primary importance is the potential of the yeast cells to perform many post-translational modifications to the expressed proteins. These include: processing of signal sequences, folding, disulfide bridges formation and both O- / N- linked glycosylation.

4.2.3 Biotechnological tools for protein production.

For the production of heterologous proteins, plasmid vectors were designed containing an expression cassette equipped with P_{AOX1}, followed by multiple restriction sites for the cloning of the foreign gene. These plasmids, upon linearisation, can generate stable *P. pastoris* transformants via homologous recombination within the genome. One of the most used plasmids is the 8 kb pPic9 (Fig. 4.4), which allows overexpression and secretion of the heterologous proteins. The expression cassette is designed for the expression of the target gene in frame and downstream of a secretion signal: the 89 aminoacids long pro-peptide of the *Saccharomyces cerevisiae* alpha-mating factor (α-MF). Processing of the obtained fusion protein is performed by the secretion machinery of the yeast (KEX2/STE13

gene products). While the KEX2 catalyses the hydrolysis of the protein between Arg84 and Glu85 of the tag, STE13 then remove the Glu-Ala repeats (Fig. 4.5) remained bound to the N-terminal of the target protein (Fuller et al. 1991). A detailed map is given in Fig. 4.4. The expression vector carries the auxotrophic *HIS4* gene, which encodes for histidinol dehydrogenase, an enzyme in the biosynthetic pathway of histidine. This feature is used as a selective marker, which complements knocked-out artificial *his4* yeast strains, and allows selective growth of transformants in minimal medium. Gene insertion events may occur both at *AOX1* or *HIS4* loci, both as a unique or a double cross-over event (gene replacement). The recombination type is related to the restriction site used for the plasmid linearisation and generates variations in the metabolising capacity of the transformant yeasts depending on the strain used. Multiple insertion events are also observed (Crane and Gould 1994).

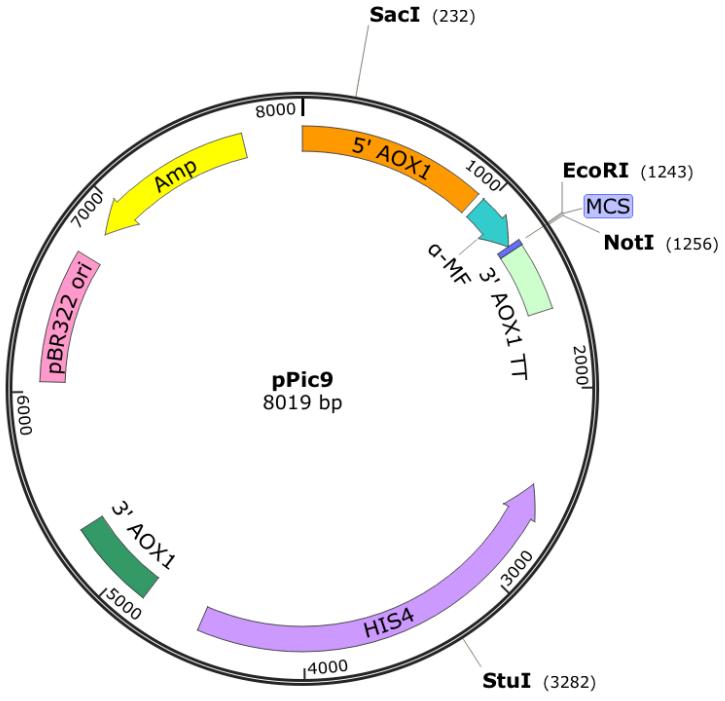


Figure 4.4 | Map of the vector pPic9. Elements: AOX1 promoter (5' AOX1); DNA sequence coding for the secretion signal sequence α-mating factor (α-MF); multiple cloning site (MCS); native transcription termination and polyadenilation signal from AOX1 gene (3' AOX1 TT), the *Pichia* wild-type gene coding for histidinol dehydrogenase (HIS4); sequences from AOX1 gene for the integration of the plasmid at the AOX1 locus (3'AOX1); the ampicillin resistance gene (Amp); the *E.coli* plasmid pBR322 origin of replication (pBR322). Restriction sites for the linearisation of the plasmid are indicated (SacI, StuI). Picture created with free software Snapgene (http;//www.snapgene.com).

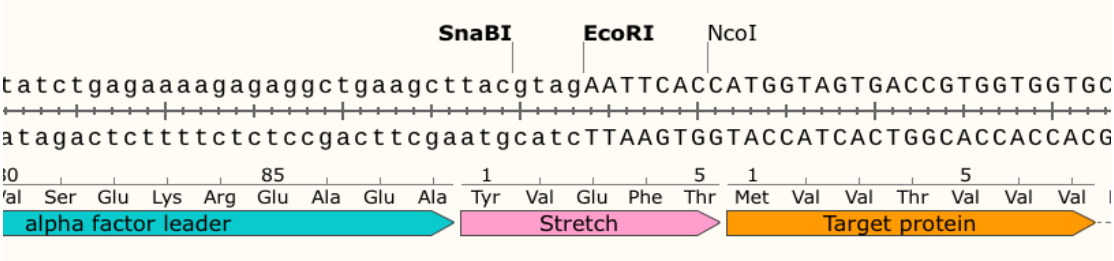


Figure 4.5 | Scheme of the cleavage site in the fusion protein expressed in *P. pastoris*. During secretion, KEX2 cleaves between Arg84 and Glu85. Consequently, STE13 removes the Glu-Ala repeats of the alpha-MF C-terminal portion. A stretch composed of five aminoacids (YVEFT) is expressed upstream the expressed protein because codons of alternative restriction sites present in the constructions are expressed in frame with the leader peptide (α-MF). Illustration created with Snapgene online free software (http://www.snapgene.com).

4.3 METHODS

4.3.1 Insect cells system.

All the procedures regarding insect cells requiring sterility are conducted with common aseptic techniques under vertical laminar flow hood equipped with HEPA (High Efficiency Particulate Air filters) 0,2 µm.

4.3.1.1 Initiation of a culture from cryo-stocks and cryopreservation.

Insect cells are stored indefinitely in liquid nitrogen in stocks of 1 ml in sterile cryotubes. In order to start new cultures from a frozen cryo-stock, cells are quickly thawn in a water bath (37°C), and aliquoted in a T-25 flask pre-filled with 4 ml of Sf900-II medium. After 30'-60' of incubation at 27°C the medium is replaced, and the procedure is repeated 24 hours later in order to recover maximum viability of the cells. Cryopreservation of insect cells is performed as follows: 5 ml suspension of cells in mid-logarithmic phase are centrifuged 10'//23°C/500 rpm, and resuspended in 1 ml of cryopreservation medium (45% v/v fresh Sf900 II + 45% v/v filtered culture medium + 10% DMSO). Cryovials are placed in polystyrene boxes and incubated overnight at -80°C, allowing a slow decrease of the temperature (-1°C/hour), before the definitive storage in liquid nitrogen.

4.3.1.2 Cultivation of cells in monolayers.

To generate a monolayer culture, insect cells from a suspension are diluted to 1,5*10^6 cells/ml in pre-warmed Sf900-II medium (Invitrogen) and poured into a sterile T-25 flask, followed by incubation in a humidified incubator at 27°C. Subcultures are set up every 3-4 days for Sf9 cells and every 2 days for High Five (Hi5) cells, dislodging the attached cells from the bottom of the plate pipetting the medium several time directly on the monolayer, or gently tapping the flask onto a hard surface. Once the cells are dislodged, 0,5 ml to 1 ml is transferred into a new plate, already containing pre-warmed growth medium. Larger monolayer cultures are also set up, using T-75 (20 ml medium) or T-150 (40 ml medium) flasks.

4.3.1.3 Cultivation of cells in suspension.

Cells are grown in suspension inoculating them in sterile Erlenmeyer flasks at a starting concentration of 1,5*10^6 cells/ml and incubating them on an orbital shaker at 27°C and 120 rpm agitation rate. The ratio between the volume of growth medium and volume of the flask is kept constant at 1:6. PTGE sterile flasks (Nunc, Wiesbaden, DE) of 100, 250, 500 ml are used in this work.

4.3.1.4 Cell culture parameters: cell density and viability.

Determination of the density of a culture (either in suspension, or dislodged cells from a monolayer) is performed by manual cell count. An aliquot of 10 µl is carefully placed onto a haemocytometer, at the interface between the cover slip and the counting grid. Several countings are performed for statistical robustness of the data. Viability/mortality of the cells is estimated by means of selective cell staining. Aliquots of cells are mixed with 0,4% Trypan Blue (a dye that penetrates only through the membranes of the dead cells) in PBS solution at a ratio 1:10 (dye : cells), and incubated at room temperature for 15'. The ratio of stained cells over the total number reflects the index of viability of the culture tested.

4.3.1.5 Bacmid extraction.

Recombinant bacmids contained in bacteria DH10Bac are selected by blue-white selection (§4.1.2). Successively, bacterial pellets from 6 ml culture are separated by centrifugation 10'/4°C/10000 g, resuspended in 800 µl of Resuspension Buffer (8% sucrose, 100 mM Tris pH 8, 10 mM EDTA, 50 µg/ml lysozyme, 10 µg/ml RNAase) and

incubated r.t/10'. Afterwards, 400 µl of 1% SDS in TE buffer are added, and the mixture incubated 37°C/10', followed by a third incubation on ice/30' after addition of 200 µl of 5 M potassium acetate and very gentle mixing. The mixture is centrifuged 20'/4°C/10000 g and the supernatant transferred in a new tube. An additional centrifugation step is applied to remove any debris, and the supernatant collected. An amount of 800 µl of iso-propanol is added and a DNA pellet obtained by centrifugation 15'/4°C/10000 g. A washing step with 70% ethanol is applied. Complete removal of ethanol is obtained by air drying, and the DNA pellet (the recombinant bacmid) resuspended in 50 µl of TE buffer. The recombinant bacmids produced are bacPRE and bacSIL, respectively from the transfer of the gene from pBac-PRE or pBac-SIL.

4.3.1.6 Transfection and virus amplification.

Bacmids (bacPRE and bacSIL) are transfected into Sf9 cells complexed with a lipidic cationic detergent (Turbofectin, Invitrogen). Turbofectin stocks are prepared diluting 15 µl of detergent in 250 µl of basal IPL41 culture medium, and incubating r.t/45'. Solubilisation of the detergent is achieved by vortexing. Sf9 cells in mid-logarithmic phase (3 ml culture) grown in monolayer onto the surface of a 6-well plate (viability > 90%) are selected for the transfection preferring the best in terms of morphology, and minimal presence of debris, and floaters. The supernatant is removed, and the cells washed twice with 1 ml of IPL41 medium. In the meanwhile, the transfection solution is prepared mixing the bacmid with Turbofectin solution at a final concentration of 5 µg/ml and incubating on ice for 15'. Transfection solution is distributed on the top of the monolayer, followed by an incubation 28°C/5 hours. Finally the solution is replaced by 2 ml of Sf900 medium containing 50 µg/ml gentamycin, and the medium refreshed after 24 hours. Inspection of the morphology of the cells is performed daily, until clear signs of infection appear (Fig. 4.9). The first generation of recombinant baculovirus released in the supernatant (P1) is harvested after one week of cell growth (centrifugation 20'/4°C/500 g) and filtered 0,22 µm. Further generations of higher titer stocks (P2, P3, P4) of virus are obtained by cyclical infection of Sf9 cells with 2,5% v/v of the parental virus generation, and harvested by separation from the cell pellets.

4.3.2 Yeast cells system.

4.3.2.1 Preparation of media.

For different purposes, different culture media are used in the development of the yeast system. Formulations of the culture broths are summarised in Tab. 4.1. BMGY and BMMY constitute the "production" medium, YPD the "growth medium", and RDB the "selection" medium. The usage of each type of medium is described in the following Paragraphs. All the liquid medium and agar plates for the cultivation of the yeast are prepared from a base solution containing peptone and yeast extract, sterilised by autoclaving. The separate stocks are added in sterile conditions. Seeding of *Pichia pastoris* cells is performed on agar plates, kept at incubation temperatures of 28-30°C. Suspension cultures are maintained as described in §2.3.4

Table 4.1: Components of the different media for the yeast cultivation. Concentrations are expressed in percentage weight on volume (% w/v) or molarity (M).

*YNB contains a nitrogen source for nitrogen assimilation, vitamins and trace elements (5 g/L ammonium sulfate, 2 µg/l biotin, 400 µg/l calcium pantothenate, 2 µg/l folic acid, 2 mg/l inositol, 400 µg/l nicotinic acid, 200 µg/l p-aminobenzoic acid, 400 µg/l pyridoxine HCl, 200 µg/l riboflavin, 400 µg/l thiamine HCl, 0,1 g/l citric acid, 500 µg/l boric acid, 40 µg/l copper sulfate, 100 µg/l potassium iodide, 200 µg/l ferric chloride, 400 µg/l manganese sulfate, 200 µg/l sodium molybdate, 400 µg/l zinc sulfate, 1 g/l potassium phosphate monobasic, 0,5 g/l magnesium sulfate, 0,1 g/l sodium chloride, 0,1 g/l calcium chloride);

**Aminoacids is a mixture of 5 g/l L-glutamic acid, 5 g/l L-methionine, 5 g/l L-lysine, 5 g/l L-leucine, 5 g/l L-isoleucine.

	BMGY	BMMY	YPD agar	RDB agar
Yeast Extract	1%	1%	1%	\
Peptone	2%	2%	2%	\
KH_2PO_4 buffer, pH 6,0	100 mM	100 mM	\	\
YNB*	1,34%	1,34%	\	1,34%
Biotin	$4*10^{-5}$ %	$4*10^{-5}$ %	\	$4*10^{-5}$ %
Glycerol	1%	\	\	\
Methanol	\	0,5%	\	\
Aminoacids**	\	\	\	0,005%
Sorbitol	\	\	\	1 M
Dextrose	\	\	2%	2%
Agar	\	\	2%	2%

in sterile Erlenmeyer flasks under agitation (250 rpm) and constant temperature (28°C). Generally the culture volumes never exceed 1/5 of the flask volume; improved aeration is achieved by sealing the flask neck with a porous tape (Invitrogen).

4.3.2.2 Transformation of competent yeast cells.

Competent *Pichia pastoris* cells are prepared using the kit PichiaComp (Invitrogen). GS115 or KM71 *Pichia pastoris* (Invitrogen) single colonies are inoculated in 10 ml sterile YPD medium (Tab. 4.1) and grown overnight at 30°C on an orbital shaker at 250 rpm agitation rate. The cells are diluted to an OD_{600} of 0,1 and grown until OD_{600} reaches 0,6-1. Cell pellets are separated by centrifugation 5'/r.t./500g and resuspended in 10 ml of Solution I; a second "centrifugation / resuspension step" in Solution I is repeated to a final volume of 1 ml. Aliquots of 50 µl are distributed in sterile tubes, and used for immediate transformation or cryopreservation at -80°C. The concentrated suspensions of competent cells are mixed separately either with 3 µg of linearised pPic9-PRE or pPic9-SIL; 1 ml of Solution II is added, mixed properly, and incubated for 1 hour at 30°C. The transformation reaction is mixed every 15 minutes. When the incubation is completed the reaction tubes are subjected to heat shock (10'/42°C). Cells are pelleted by centrifugation 5'/r.t./3000g, resuspended in 1 ml of Solution III, and (after a second centrifugation step), the solution volume reduced to 150 µl, before plating onto RDB agar (the selection medium, Tab. 4.1).

4.3.2.3 Transformation by electroporation.

Electrocompetent yeast is prepared as a sequence of resuspension steps of sedimented cells in sterile ice-cold solutions. Pellets are obtained by centrifugation cycles 5'/4°C/1500g. Suspensions of *P. pastoris* cells at OD_{600} of 1.3 – 1.5 are sequentially resuspended through the following solutions: 500 ml water, 250 ml water, 20 ml 1 M sorbitol, 1 ml of 1 M sorbitol. Aliquots of 80 µl of the electrocompetent cells are mixed with 5-20 µg of linearised plasmid pPic9-PRE or

pPic9-SIL and transferred into an ice-cold electroporation cuvette (0,2 cm). After incubation of 5' on ice, the cuvette is placed in the electroporation device, the cells pulsed (2 kV, 1 pulse, time constant = 4,5 ms) and immediately resuspended in 1 ml of 1 M sterile sorbitol. 200-600 µl of cells are distributed on the top of an RDB agar plate (Tab. 4.1), and incubated at 30°C for maximum 4 days.

4.3.2.4 Yeast genomic DNA extraction.

Genomic DNA from *Pichia pastoris* is extracted using the kit High Pure PCR Template Preparation Kit (Roche): 200 µl of suspension containing around 10^8 yeast cells is centrifuged 5'/r.t./3000g and resuspended in 200 µl PBS. Lysis of the cell wall is promoted by the addition of 10 µl of lyticase (0,5 mg/ml) and incubation 30'/37°C. Subsequently, 200 µl of Binding Buffer and 40 µl of Proteinase K are added to the mixture, before incubation 10'/70°C. Eventually, 100 µl of iso-propanol is added and the lysate mixed properly. The extracted DNA is bound to a silica resin contained in a centrifuge column, washed twice with Washing Buffer, and eluted in 25 µl of Elution Buffer, following the instructions of the manufacturer.

4.4 RESULTS

4.4.1 Design of the constructs.

The two gene donors used in this section, encoding for the pre-pro-silicaten and silicatein ("PRE" and "SIL", respectively) are synthesised from Geneart (Life Technologies), upon optimisation of the genetic sequence for the codon usage of the insect cells (*Spodoptera frugiperda*). They are provided as inserts in the plasmid pMK-T, carrying the resistance gene for kanamycin. The constructs are designed as follows: while SIL (767 bp) encodes for the wild-type gene (AJ272013) lacking of the sequence encoding for the pro-domain (aa_{16} to aa_{113}), PRE (1057 bp) covers the entire ORF, but with a point mutation on Ser_{138}, substituted by alanine (S138A, Fig. 4.6), one of the aminoacid belonging to the catalytic triad of the enzyme, and presumably also involved in the proteolytic activity. Such mutation was placed in order to knock out the proteolytic function, leading to spontaneous autocleavage of the protein (Schröder et al., 2012). Inactivation of the maturation process of the zymogen is intended to increase stability of the expressed protein, as reported in a similar approach on cathepsin S (Kaulmann et al. 2006). Restriction sites NcoI and NotI were added respectively at the 5' and 3' of the gene for molecular cloning into the Baculovirus transfer vector (Fig. 4.3). An additional restriction site EcoRI is introduced upstream the genes for the cloning in pPic9 (Fig. 4.4). In order to

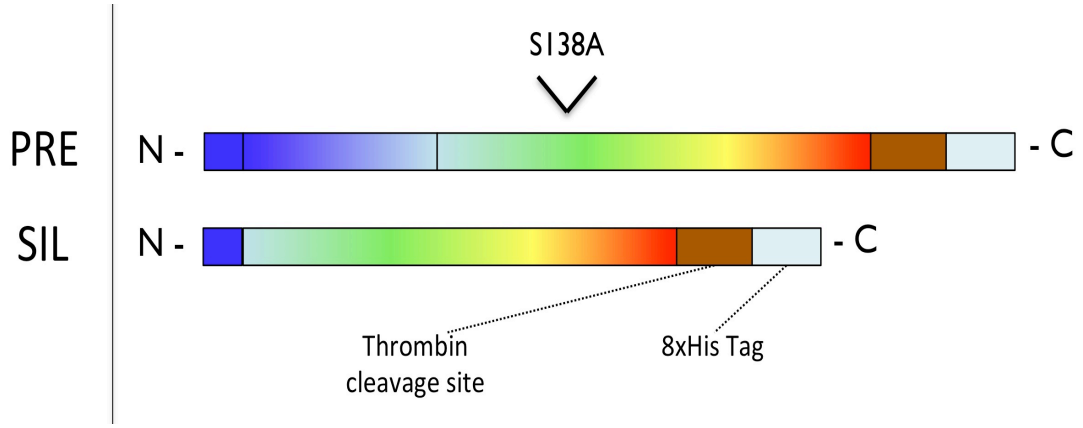


Figure 4.6 | Scheme of the constructs for the expression of PRE (precursor) and mature silicatein (SIL) in insect and yeast cells. PRE contains the entire silicatein gene, comprising the signal sequence, the pro-domain, and the mature enzyme. A point mutation on the catalytic serine is also evidenced in correspondance of the arrow; SIL is deprived of pro-peptide. Both constructs contains a thrombin cleavage site and a octahistidine tag at the C-terminal. Picture created with Microsoft Powerpoint.

facilitate the cloning into the insect expression vector an extra modification is applied. Restriction ends NcoI type generate an ATG initiator site followed by a guanine (G), but the second residue in silicateins (aa_2) is a Leu, which has no corresponding codons starting with G. Thus, the most similar aminoacid that can replace the natural one is a Val, encoded by consensus codon GUX. Eventually, codons encoding for 8 histidines is also added at the C-terminal, as well as a thrombin cleavage site for the removal of the purification tag (Fig. 4.6). The predicted molecular size (MW) of the expressed protein is 38,0 kDa for PRE and 26,5 kDa for SIL. Trimming of the N-terminal signal sequence reduce the MW of about 1,5 kDa.

Table 4.2 | Reaction mixtures prepared for the successful cloning of the silicatein gene inserts into the expression vectors for both expression systems. Six independent restriction digestions are performed on both inserts and vectors, followed by four combinations of ligation reactions.

Double restriction digestion (overnight, 37°C)	**Volume (µl)**					
	Insect system			Yeast system		
REACTION MIXTURE	PRE	SIL	pBac	PRE	SIL	pPic9
Insert PRE (50 ng/µl)	5			5		
Insert SIL (50 ng/µl)		5			5	
Vector pBac (50 ng/µl)			5			
Vector pPic9 (50 ng/µl)						5
Buffer TANGO (Fermentas)	2	2	2			
Buffer O (Fermentas)				2	2	2
Enzyme NcoI (NEB – 10 U/ml)	1	1	1			
Enzyme EcoRI (NEB – 10 U/ml)				1	1	1
Enzyme NotI (NEB – 10 U/ml)	1	1	1	1	1	1
dH_2O	11	11	11	11	11	11
Total volume	20	20	20	20	20	20
Ligation reactions (overnight, 15°C)	**Volume (µl)**					
REACTION	Insect system			Yeast system		
Digested PRE (10 ng/µl)	7			7		
Digested SIL (10 ng/µl)		7			7	
Digested pBac (10 ng/µl)	1	1				
Digested pPic9 (10 ng/µl)				3	3	
DNA ligase buffer (NEB)	2	2		2	2	
T4 DNA ligase (10 U/ml)	0,5	0,5		0,5	0,5	
dH_2O	9,5	9,5		9,5	9,5	
Total volume	20	20		20	20	
CONSTRUCT	**pBAC-PRE**	**pBAC-SIL**		**pPic9-PRE**	**pPic9-SIL**	

4.4.2 Generation of the expression vectors and diagnostic tests.

Molecular cloning procedures (summarised in Tab. 4.2) are employed for the genetic engineering of the expression vectors (pBac and pPic9). The inserts (PRE and SIL) were excised from the donor plasmids (pMK-T) and ligated into each expression vectors previously digested with the appropriate restriction enzymes, affording four novel plasmids: two for the expression in Baculovirus (pBac-PRE and pBAC-SIL), and two for the yeast (pPic9-PRE and pPic9-SIL). The correctness of the insertions are assessed performing a double cleavage reaction with the appropriate restriction enzyme couple (NcoI/NotI for pBac plasmids, and EcoRI/NotI for pPic9 plasmids) on samples of the recombinant vectors, followed by DNA agarose gel electrophoresis, as shown in Fig. 4.7. As expected, digestion of pBac plasmids result in two separate bands: one at around 5 kbp, corresponding to the free vector, and one at around 1 kb for PRE, or 0,75 kb for SIL. Digestions of pPic9 vectors generate a band at around 8 kb, size of the plasmid, plus the associated inserts. Since in both types of plasmids two different restriction ends are created, it is unlikely that the insertion occurs in the wrong direction.

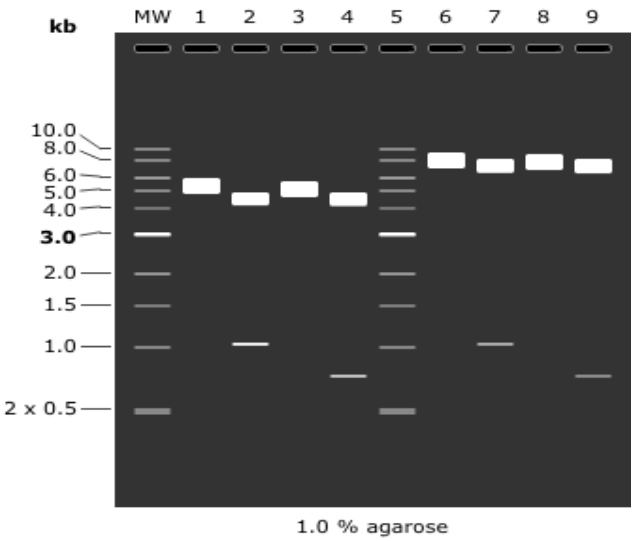


Figure 4.7 | Agarose gel (1%) electrophoresis representation of the diagnostic restriction digestions of the constructed expression plasmids. Digestion of pBac-PRE (5,7 kbp, lane 1) with enzymes NcoI/NotI generates two fragments of 4,7 kbp and 1 kbp (Lane 2), corresponding to the empty vector and PRE insert; digestion of the plasmid pBac-SIL (5,5 kbp, lane 3) leads to the separation of the SIL insert (0,75 kbp, Lane 4) from the empty vector pBac; digestion of the plasmid pPic9-PRE (9,1 kbp, Lane 6) originates the fragments at 8 kbp (empty vector pPic9) and 1 kbp (PRE sequence), visible in Lane 7; reaction on the plasmid pPic9-SIL (8,8 kbp, Lane 8) permits the separation of the insert SIL (0,75 kbp) from the expression vector pPic9 (Lane 9). The simulations are created for a better clarity and data visualisation quality with Snapgene Trial Version (http://www.snapgene.com). Original documentation of the agarose gels is available.

4.4.3 Generation and characterisation of the recombinant viruses

Plasmids pBac constitute the vehicle of the foreign genes, which are expected to integrate into the viral genome (the bacmid) by site-specific transposition. The bacmid is contained in a special strain of bacteria (*E.coli* DH10Bac), and isolation of transformant colonies (where positive integration of PRE and SIL genes have occurred) is performed by blue/white selection. Bacteria are transformed as already described in §2.4.8 and seeded on agar plates containing the chromogenic substrate X-gal and IPTG, in addition to a mixture of antibiotics. Acquisition of resistance to kanamycin, gentamycin, and tetracyclin derives from the bacmid, the transfer vector and the helper plasmid, respectively. After 48 hours of cultivation, colonies with white phenotype are identified and selected with an average yield of 6% for bacteria transformed with pBac-PRE and 11% for pBac-SIL. Recombinant bacmids (bacPRE and bacSIL) are extracted and purified as described in §4.3.1.5. The presence of the correct insert is monitored by PCR analysis using primers complementary to segments from the bacmid flanking the insert (Fig. 4.8). The amplified DNA fragments match with the expected sizes (about 3,3 kbp for bacPRE and 3,0 kbp for bacSIL), as it is determined by agarose gel electrophoresis, shown in Fig. 4.8. The PCR reaction is performed in volume of 20 µl using the HiFi PCR kit (KAPA Biosystems), and the following set of cycles: 1 cycle 95°C / 180″ + 35 cycles (98°C /20″ + 60°C / 15″ + 72 °C / 180″) + 1 cycle 72°C / 5′. Bacmids are then used for transfection of Sf9 cells, in order to produce the first generation of infective Baculovirus progenies (Vir1PRE and Vir1SIL). Starting from the initial transfection, the titre of the virus must be increased by several cycles of infection of insect cells, as described in §4.3.1.6. Every generation (Vir2, Vir3, Vir4… Vir_n) is collected separately and used for infection (2,5% v/v) of Sf9 cells cultivated in parallel. Evaluation of the most productive viral stock is conducted by SDS-PAGE and Western blot of cell lysates at 4 days post-infection. The generation Vir4 showed for both PRE (Vir4PRE) and SIL (Vir4SIL) a higher intracellular protein production compared to the parental virus generations (data not shown). Further generations were not generated in order to avoid accumulation of mutations, or loss of the genes.

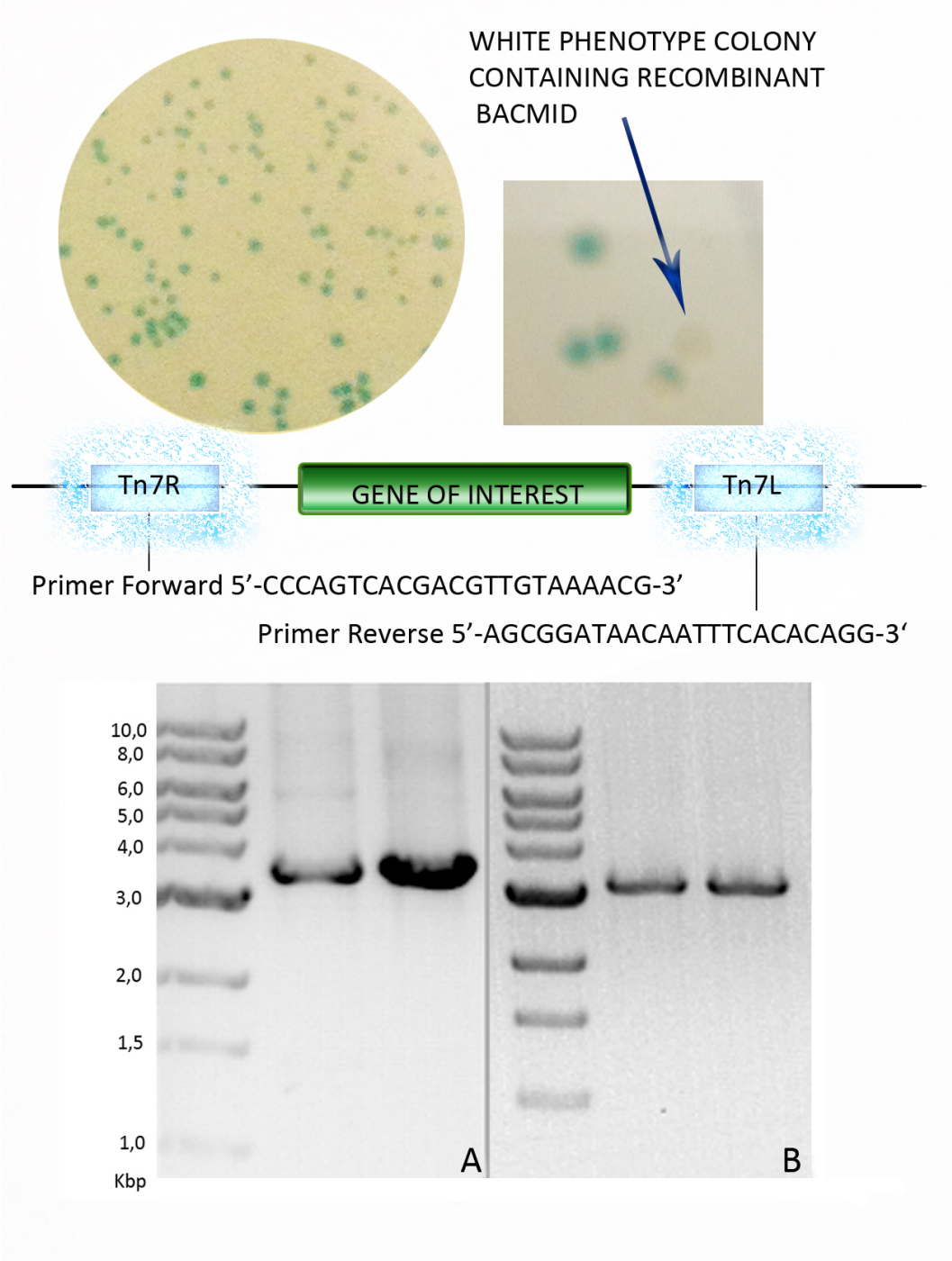


Figure 4.8 | Blue/white selection of bacteria for the production of the recombinant bacmids. Upper Left: LB agar plate with DH10Bac colonies grown in presence of antibiotics and X-gal. Upper Right: Zoom of an area containing colonies with different phenotypes: 3 colonies with blue colour, 1 colony with slight blue, and 1 colony with white phenotype. The latter is a potential transformant to be selected for the extraction of the recombinant bacmid. Center: a scheme of the region selected for the detection of the insertion of the silicatein genes into the bacmid is shown. Bottom: characterisation of the purified bacmids. PCR performed on DNA extracted from four different white colonies (two transformed with pBac-PRE and two with pBac-SIL) employing the primers Fwd and Rev. Fragments of 2300 bp plus the length of the gene (3357 bp for PRE sequence (panel A) and 3067 bp for SIL sequence (panel B)) are amplified.

4.4.4 Insect cells morphology and cultivation.

Two insect cell lines (Sf9 and High Five) are maintained for at least 20 sub-culturing cycles at a viability of 90-95%. In case of viability reduction or contaminations from external biological agents, new cultures are established from the cryopreserved cells as described in §4.3.1.1, with total recovery of cell functionality after two sub-culturing cycles. Healthy insect cells generally appear as rounded singular units with a diameter of around 15 µm. In later growth stages the cells exhibit the tendency to form small cell clusters, especially the High Five strain. Baculovirus infection progression is assessed through frequent observations by optical microscopy. Typical infection signs are the enlargement and deformation of the cell volume, the decrease of the growing rate, the appearance of granules and the expansion of the nucleus (Fig. 4.9). These observations are of primary importance, especially for the determination of the proper harvesting time of the first generation of Baculoviruses.

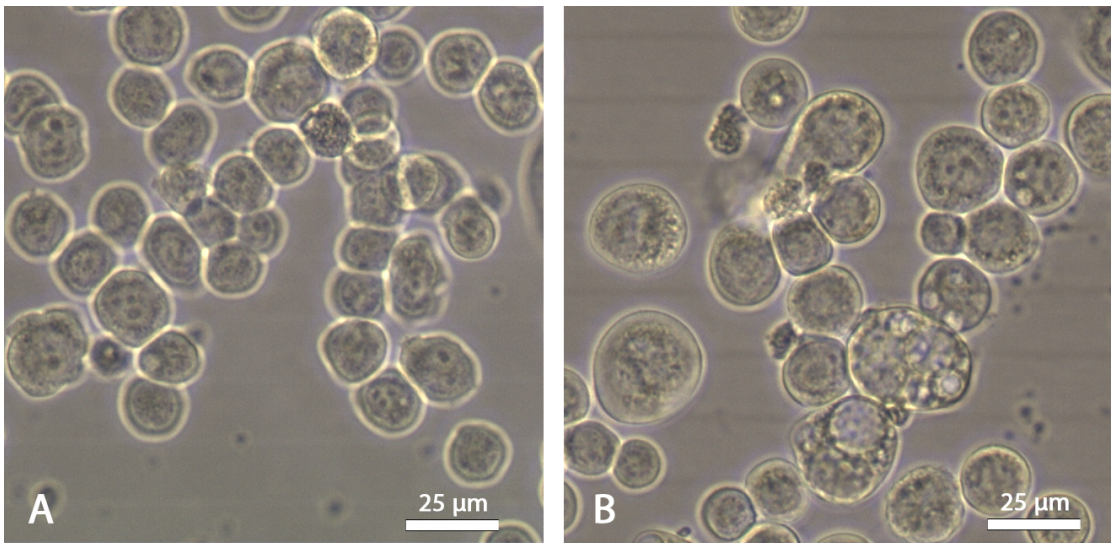


Figure 4.9 | Morphology of Sf9 cells. Micrographs of a cluster of Sf9 cells growing on a monolayer. Comparison between non-infected (A) and infected cells with 5% v/v Vir3SIL Baculovirus stock (B). Observations are performed after 3 days from the infection. Normally, Sf9 cells not exposed to the virus show a doubling time of 48-72 hours, and a quite uniform morphology. The infection causes in some cells abnormal enlargement, increase of the presence of granules, and a looser inter-cellular packing. Micrographs are taken on EVOS xl digital inverted microscope (Fisher Scientific GmbH, Nidderau, DE).

4.4.5 Expression of silicateins in insect cells

The development of a method for the production of recombinant proteins in insect cells starts with small-scale expression tests. Two cell lines (Sf9 and High Five) are compared. Cells are cultivated in 20 ml suspensions (100 ml vessel) and infected with a variable amount (% v/v) of Vir4 Baculovirus stocks (Vir4PRE and Vir4SIL); cell pellets are collected at different time points, and processed under equal lysis conditions in order to evaluate the performance within a wide time-window, and identify the optimal harvesting time. In Sf9 cells both PRE and SIL start to accumulate only after the second day of infection, increasing during the 3rd and 4th day. No intracellular proteolysis events are observed. High Five cells show a faster productivity, as both proteins are already detected after 24 hours from the start of the infection. Nevertheless, this cell line is characterised by a higher growth rate compared to Sf9 cells, concomitant with more pronounced cell decay. By densitometric analyses on Coomassie blue stained SDS-PAGE of the cell lysates, the production yields are estimated in the range of 10-30 mg/L of overexpressed protein, with optimum at 48 hours post-infection for High Five cells, and 96 hours

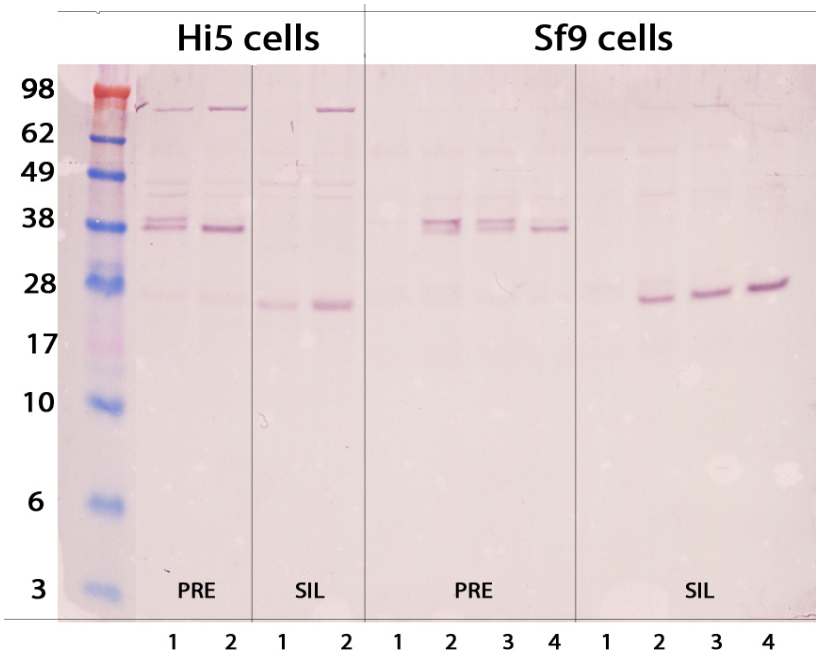


Figure 4.10 | Expression profile in Hi5 and Sf9 cells for PRE and SIL. Time lapse experiment. Identification of the presence of silicateins at 1,2,3 and 4 days post-infection for Sf9, and at 1 and 2 days for High Five cells (Hi5); the latter suspensions were stopped earlier due to the faster growth rate, and decrease of culture quality and cell viability. Test conducted by Western blot obtained from 4%-12% NuPAGE. Detection with anti-silicatein α antibody.

post-infection for Sf9 cells (Fig. 4.10). The majority of the detected protein is accumulated intracellularly, whereas only negligible amounts of secreted proteins are detected in the supernatants. Both PRE and SIL are extracted from frozen cell pellets (-20°C) in Tris/NaCl buffer after three cycles freeze/thawing in soluble conditions. Neither detergents nor chaotropes are required to stabilise the proteins in solution at this stage. No degradation events were observed within 7 days after extraction and storage at +4°C.

4.4.6 Purification of silicateins in insect cells

Several attempts to purify recombinant PRE and SIL proteins from the insect cellular extracts are performed. Multiple techniques and parameters are studied: composition and amount of the washing and elution buffers (pH, ionic strength), binding time and temperature. Moreover, manual procedures employing Ni^{2+}-NTA resin (Invitrogen) are compared to automated systems (Profinia, BioRad). Generally, in native conditions it is observed very poor retention of the extracted His-tagged proteins to the stationary phase, as almost the totality of target proteins are found in the unbound fraction, even after repeated binding cycles (data not shown). Extraction and purification in denaturing conditions (6 M urea) appeared to be necessary to enable the binding to the Nickel-agarose matrix and achieve sufficient recovery of proteins. The yield is estimated at a level of 2-3 mg of pure SIL from 100

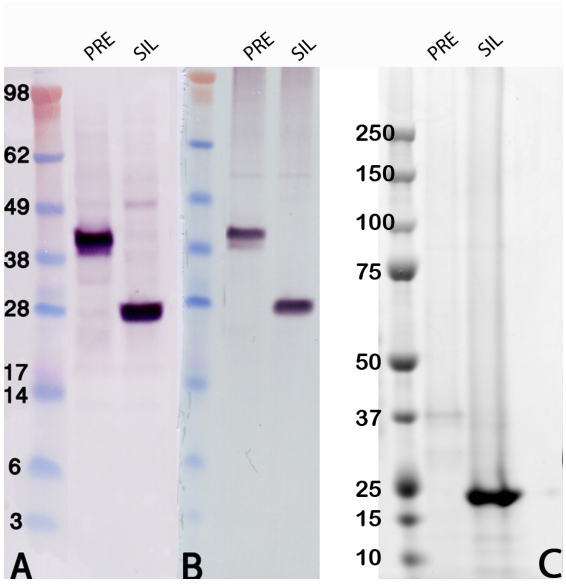
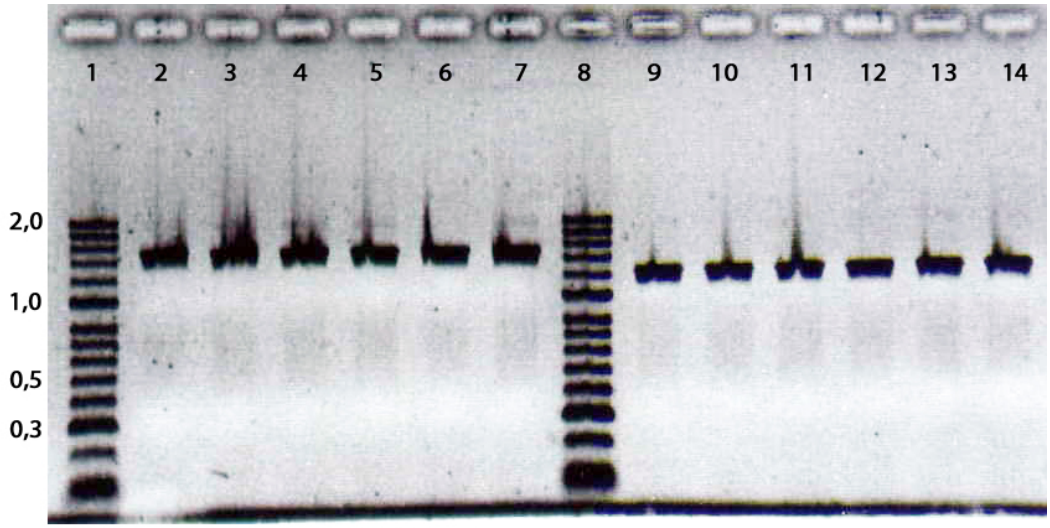


Figure 4.11 | Detection of PRE (≈ 38 kDa) and SIL (≈ 26,5 kDa) with antibody anti-silicatein (A) and anti-His antibody (B) in Sf9 cell extracts 4 days post-infection with Vir4PRE and Vir4SIL C) Pure fractions of PRE and SIL in denaturing conditions visualised by Coomassie blue stained SDS-PAGE.

ml of Sf9 culture (Bradford method). In contrast PRE is detected after purification in denaturing conditions only in traces in the elution fraction, as it is shown in Fig. 4.11C.

4.4.7 Generation and characterisation of transformant *P. pastoris*.

Transformation techniques for the insertion of the expression vectors into the yeast cells are performed by chemical transformation of competent yeasts, and electroporation. The latter resulted in a better transformation yield for the *Pichia* strain GS115. Poor efficiency is seen instead with both methods for the strain KM71. Recombination of the transfer vector into the genome occurs upon linearisation operated by restriction enzymes. The endonucleases SacI and StuI are selected



Primer pAOX1Fwd: 5'-GACTGGTTCCAATTGACAAGC-3'

Primer pAOX1Rev: 5'-GCAAATGGCATTCTGACATCC-3'

Figure 4.12 | PCR screening on extracted genomes from random yeast clones. Fragments amplified from integrated PRE (1544 bp) are shown on the left sector (lanes 2-7); fragments amplified from SIL (1259 bp) are shown on the right (lanes 9-14). Sequence of the oligomers used for the PCR initiations are shown.

because no associated restriction sites are present on the silicatein gene sequences. Purified plasmids pPic9-PRE and pPic9-SIL are digested following recommendation of the manufacturer and aliquots of digested DNA are analysed by agarose gel electrophoresis. Linearisation of the plasmid is observed as an increase in the band size (data not shown), generated from the transition from the supercoiled to the linear state (Vologodskii et al. 1979). Positive transformations of a group of clones of *P. pastoris* are investigated at a genetic level. The yeast genome is extracted using the procedure described in §4.3.2.4. Silicatein genes, integrated by homologous recombination, are amplified by PCR using the primers in pAOXFwd and pAOXRev (Fig. 4.12), the High Fidelity PCR kit (Fermentas) and the following set of temperature cycles: 1 cycle 95°C / 120″ + 35 cycles (95°C / 30″ + 56°C / 30″ + 72 °C / 120″) + 1 cycle 72°C / 5′. The amplified fragments obtained match with the expected sizes in all the clones tested (Fig. 4.12).

4.4.8 Identification of highly productive yeast clones.

The identification of productive yeast clones is screened comparing the expression level of the silicateins (PRE and SIL) both in the intra-cellular fraction, and in the culture medium from about 50 independent clones per type (PRE or SIL). Suspension cultures are initiated from every colony to be tested in BMGY medum (Tab. 4.1), and the expression induced with 1% methanol every 24 hours in BMMY medium, according to the instructions of the manufacturer. The harvesting is performed at the fifth day of production, and the protein expression level assayed by dot-blot technique (§2.5.4). In Fig. 4.13 the different performance of a group of clones (strain GS115) resulted all positive at the PCR test (Fig. 4.12) is shown. From the screening it is possible to identify clones with several level of expression. Detection with anti-silicatein antibody showed reactivity in almost every spot, but these data were contradicted by Western blot analyses, which revealed a certain grade of non-specificity of this antibody in the yeast system. The usage of the anti-His antibody instead guaranteed good level of specificity and no reactivity with yeast proteins, enabling reliable identification of clones overexpressing either PRE or SIL. Detectable secreted proteins were identified in clones P4, P7 (expressing

PRE), S1 and S3 (expressing SIL). This result suggests that analysis of the genome by PCR for the identification of the genes is not sufficient to confirm the ability of the yeast to successfully express the desired proteins.

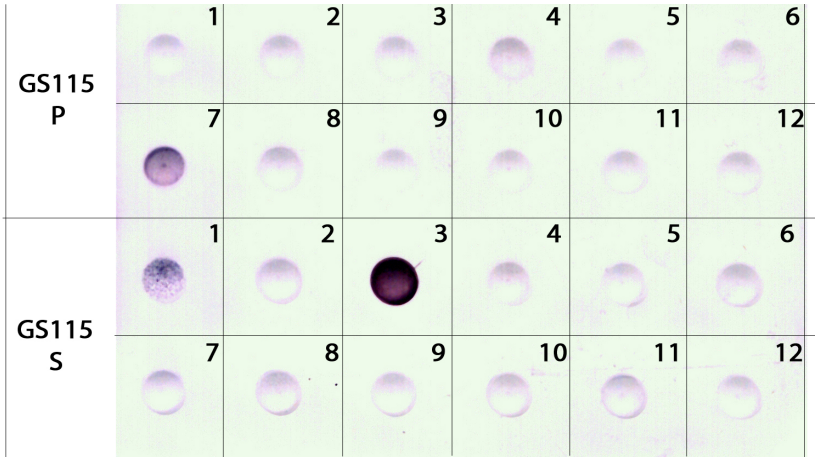


Figure 4.13 | Detection of secreted silicatein at the 5th day of cultivation from 24 cultures of *Pichia pastoris* (strain GS115) obtained from different clones (transformed either with pPic9-PRE (pre-pro-silicatein serie -P1, P2,…P12-) or pPic9-SIL (silicatein serie -S1, S2,…S12-).

4.4.9 Optimisation of the expression levels in yeast.

After identification of productive clones, different parameters affecting the expression levels of yeast were examined and the dot-blot technique has represented a straightforward method to perform such investigations. Clone #S3 expressing SIL (Fig. 4.13) is selected as model for the tests. Protein production levels depends on cultivation time, as shown in Fig. 4.14A, with an increase in the intensity of the reactive spot, obtained from equal amounts of supernatants. A separate experiment showed decrease of the overexpressed protein concentration of SIL in solution after the 5th day of fermentation (data not shown). Immuno-blot experiment confirmed that the signal measured in the dot blot derives exclusively from the His-tagged SIL (band at around 26,5 kDa), and not from other contaminants (Fig. 4.14B). Important parameters like temperature and pH are monitored for their effect on the protein production. The temperature tested (22°C, 26°C, and 30°C) didn't give appreciable differences (data not shown), while the pH represented an important factor determining the outcome of the yeast production.

Based on our observations, silicatein secretion is stimulated more intensively in acidic conditions (Fig. 4.14A). Densitometric analyses calibrated on known concentrations of SIL revealed a maximum productivity of ≈30 mg/L from the clone #S3.

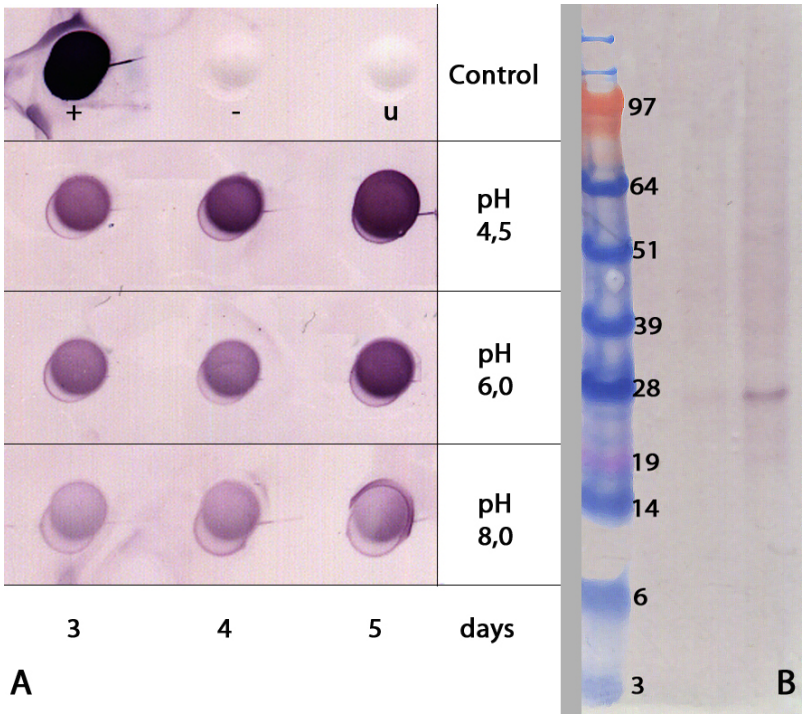


Figure 4.14 | A) Secreted His-tagged silicatein in medium BMMY at pH values of 4,5, 6 and 8 at different time periods (3,4 and 5 days after the inoculation of the cultiure in the production medium). Detection with monoclonal antibody anti-His tag. Positive (+), negative (-) and untransformed yeast (u) are included in the test; B) Western blot: 15 µl of *Pichia pastoris* culture medium is loaded into SDS-PAGE and transfered into nitrocellulose membrane. A unique band at MW < 28 kDa is detected, corresponding to secreted His-tagged silicatein.

4.5 DISCUSSION

4.5.1 Implementation of the expression strategies.

Availability of a highly pure, homogeneous, and soluble protein is an indispensable requirement for any experimental study e.g. spectroscopic, crystallographic, biophysical, and biochemical research. Recombinant DNA methodologies are constantly improving and have set the stage for the production of recombinant proteins in multiple hosts, bypassing tedious procedures of isolation from complex biological systems and enable the creation of artificial mutants with improved features (Graslund et al. 2008). However, accomplishment of all the aforementioned features is strictly dependent on the nature of the target protein, and can be developed only experimentally by trial-and-error procedures. Structural Genomics Centres offer valuable statistics on the success of every stage in the protein production pipeline, calculated on the base of the results achieved on hundreds of thousands of target proteins tested, in the frame of a global structural biology project. It is estimated that, although *E. coli* is the most used expression host (78% of the 3D structures deposited between 2004 and 2013), and significant improvements were achieved in the expression of difficult targets, only 10% of the proteins of higher eukaryotes can be expressed soluble in this system (Brown et al. 2011). Hence, solubility remains one of the major difficulties to overcome for large-scale investigation of protein function (Ahn et al. 2011). Nevertheless, many strategies to tackle this problem have been developed, as use of fusion tags, choice of alternative bacterial hosts, optimal growth conditions, refolding, co-expression of chaperones etc. Moreover, enhanced opportunities come from recent new developments in eukaryotic expression platforms, including insect, yeast and mammalian-based systems (Assenberg et al. 2013), witnessed by their increase of two-fold in the last ten years (Xiao et al. 2010). The protein studied in this work, silicatein-α from *S. domuncula,* is under investigation for more than a decade, and the elucidation of its properties and structure is in progress. The wild-type enzyme, as well as mutants and fusion proteins have been produced *in vitro* in a functional way (Bawazer et al. 2012; Cha et al. 1999; Curnow et al. 2005; Ki et al. 2013; Krasko et al. 2002; Natalio et al. 2010) and recovery of its catalytic activity from the

aggregated unfolded inclusion bodies is extensively studied, as it constitutes the only natural catalyser discovered so far in nature directing biosilica formation. Notwithstanding the successful achievements on biomineralisation processes elucidation and the several applications in biotechnology found by the refolded protein (Schröder et al. 2009), its production in soluble conditions remains a technical challenge, which must be tackled in order to proceed to a full structural and functional characterisation. Towards this goal, presently we describe an approach to produce silicatein (SIL) and a mutant of pre-pro-silicatein (PRE) from eukaryotic expression hosts: two insect cell lines and the yeast *P. pastoris*. Evaluation of each host regarding quantity and quality of the produced proteins is performed. Synthetic DNA sequences (sequences attached in Appendices §8) were carefully designed for the facile construction of the expression vectors (§4.4.1-2) and we succeeded in the generation of recombinant Baculoviruses (Fig. 4.8) able to infect and induce intracellular accumulation of PRE and SIL in insect cells (Fig. 4.10-11), as well as in the isolation of productive yeast clones expressing SIL (Fig. 4.12-13). Protein expression rate between the two systems is comparable in terms of volumetric yield, with the difference that the production in the insect cells is almost exclusively intracellular, whereas in yeast cells is secretory.

4.5.2 Production of silicateins in insect cells.

The insect expression system is chosen because it provides the eukaryotic post-translational modifications (including glycosylation) that promote secretion of the produced proteins. However, in the above described application for silicatein and pre-pro-silicatein, the accumulation of the overexpressed proteins occurs mainly with packing into inclusion bodies as in the bacterial expression (Takahashi et al. 2003). Analyses of the supernatants revealed only traces of secreted proteins (data not shown), detectable only by more sensitive immuno-blotting techniques. According to our results, the insect derived proteins could be extracted in native conditions from the cell pellets, keeping their solubility and stability in the cell lysate. The lysis procedure has been developed after testing commercial and custom solutions, and investigation on the impact of factors such as pH, addition of detergents (NP-40, Triton X-100), and reducing agents (DTT or β-mercaptoethanol).

Fine-tuning of the procedure included studies about the impact of sonication, freezing/thawing, and temperature. Despite silicateins in the insect cell extracts appeared to be soluble in native conditions, the purification of the proteins required their denaturation to succeed (Fig. 4.11C). His-tags were expressed in frame at the C-terminal of the expressed proteins, and their presence verified by immuno-blotting with specific antibodies (Fig. 4.11A-B). Buffer sets containing 0, 2, or 6 M urea both in washing and elution steps were prepared according to the standard procedures described in §2.6.1 for native, semi-denaturing, and denaturing purifications, respectively. The highest yield of pure protein was obtained with SIL, purified in denaturing condition. Our experiments reveal poor affinity of the recombinant proteins for the immobilised nickel ions in native and semi-denaturing conditions, suggesting several possible reasons: (a) the His-tag is buried internally in the folded protein (even in presence of low concentrations of urea); (b) the His-tag is masked because of oligomerisation or (c) the proteins aggregate during early stages of purification, which is a well-known phenomenon that may occur either on the top of the column or during the elution steps (Lebendiker and Danieli 2014). Factors playing a role in protein aggregation reside both in the intrinsic properties of the expressed protein, as well as the interactions with the solvation system (Churion and Bondos 2012). Reasons for this behaviour can be disordered regions without secondary structure, which usually leads to a decreased solubility of the protein, in agreement with predictions attributing to only 50% of the residues in silicatein a stable organisation in structures such as helices or sheets. Moreover, it is also known that mature silicatein *in vitro* spontaneously oligomerises in a fractal fashion (Müller et al. 2007; Murr and Morse 2005), with formation of heteropentamers (Müller et al. 2007) as basic unit for the development of the axial filament and it has been already discussed in this work (§3.4.3) the role of hydrophobic patches on the surface of the proteins as potential driving force (Akkermans et al. 2007) for the self-assembly *in vivo*. In addition to the characteristics of the protein, the physico-chemical environment surrounding the extracted proteins influences their overall folding, and a systematic approach of optimisation of the buffering system (pH, ionic strength, addition of osmolytes,

detergents and reducing agents) at the pre-purification stage is currently under study.

4.5.3 Protein processing.

From our analyses of the cell extracts, it is interesting to observe that PRE undergoes a size modification, which gradually increases during time. In fact, both in High Five and Sf9 overexpressed protein appears as a double band, with progressive conversion of the higher into the the lower band (Fig. 4.10). One of the main differences between the prokaryotic and the eukaryotic cell is the compartmentalisation of the proteins, and their targeting towards different sub-cellular organelles during and after the biosynthesis. We assume that this observation is connected to the mechanism of translocation into the endoplasmic reticulum of the insect cell, with the removal of the 15 aa N-terminal signal sequence (Lodish 2008). Reflected to the biology of sponges, this event represents an experimental evidence of the targeting of silicatein precursors to the sclerocytes, where the initial processes related to the formation of spicules take place (Wang et al. 2011), confirming the role of the N-terminal sequence as a subcellular targeting leader peptide. Moreover, the *in vitro* spontaneous maturation observed in TF-Ps (§3.4.4) does not happen in cell extracts containing PRE, irrespective of the presence or not of protein inhibitors in the lysis buffer. The protein might be unable to activate itself by autocatalysis (McQueney et al. 1997), possibly due to the point mutation (S138A) introduced for this specific construct, without excluding alternative causes such as repression of the mechanism by the intra-cellular conditions, or by interactions with inhibiting agents.

4.5.4 Multisubunit complexes.

In relation to the recent discovery of proteins interacting with silicatein, the silintaphins (Wiens et al. 2009; Wiens et al. 2011), new questions about the mechanisms regulating spicule construction, and dynamics and roles of the different components have arisen. In this context, the insect cells expression system constitutes a promising tool in order to proceed to the dissection of the

quaternary structure of such an elaborate system (Amunugama et al. 2013; Ohno et al. 2013; Shrive et al. 2014). Detailed descriptions of structure and function of complex biological multisubunit machineries have been reported through procedures of co-infection or co-expression in insect cells (Imasaki et al. 2011; Schreiber et al. 2011). Therefore, simultaneous expression of silicatein α with its natural interactors (i.e. silicatein β or silintaphins) mediated by tailored Baculoviruses could mark the basis for profiling a new perspective in the way to approach structural studies on silicatein-based multiprotein complexes.

4.5.5 Production of silicateins in yeast.

In parallel to the insect system, we studied the feasibility of a protein production system in *P. pastoris* and demonstrated that mature silicatein can be expressed and secreted in the culture medium (Fig. 4.13B). Through a simple query in the Protein Data Bank (PDB), it turned out that the yeast seems the most successful system to produce the cathepsins, prone to be grown as crystals which led to successful 3D structure elucidation. In Tab. 4.3 are shown the variants of cathepsins sharing the highest aminoacidic similarity to silicatein. The majority were obtained from the

Table 4.3 | 3D structures of cathepsins. List of the X-ray elucidated structures of cathepsins deposited in PDB showing the maximum similarity to silicatein. In Bold is in evidence the expression system used.

Protein	Specie	Expression system	PDB ID
Cathepsin V	*Homo s.*	***S. cerevisiae***	1FH0
Procathepsin L	*Homo s.*	***Pichia pastoris***	1CS8
Procathepsin L	*Homo s.*	***Pichia pastoris***	1CJL
Procathepsin K	*Homo s.*	*Escherichia coli*	1BY8
Procathepsin S	*Homo s.*	High Five	2C0Y
Cathepsin S	*Homo s.*	*Escherichia coli*	3IEJ
Cathepsin S	*Homo s.*	Sf9	2FQ9
Cathepsin S	*Homo s.*	***Pichia pastoris***	2FYE
Cathepsin K	*Homo s.*	*Escherichia coli*	3H7D
Cathsilicatein	*Homo s.*	***Pichia pastoris***	2VHS
Cathepsin K	Rabbit	***Pichia pastoris***	2F7D
Cathepsin K	*Homo s.*	***Pichia pastoris***	3C9E
Cathepsin L	*Fascicola*	***Pichia pastoris***	2O6X

yeast *P. pastoris*, hence making it an attractive host also for the silicateins production. Several other reasons elicited the interest in investigating this expression system: the rather easy handling of the microorganisms (in comparison with the insect cells), the possibility of scaling-up the culture volumes, the low cost of the equipment and reagents required (Morrow 2007; Schilling et al. 2001). *Pichia pastoris*, being an eukaryotic cell, offers also the advantages of post-translational modifications of proteins compared to the bacterial system (Demain and Vaishnav 2009); moreover, the possibility to recover the desired proteins from the culture medium, permits to bypass additional steps involving the lysis of the cell, which for the yeast case can be compromised due to the presence of a cell wall (Yang et al. 2011).

4.5.6 Considerations on the secretion of proteins in yeast.

Proteins enter into the secretory pathway by co- or post-translational translocation through the Sec61 translocon into the endoplasmic reticulum (ER) lumen. In this environment, the ER-resident protein-folding machinery operates a strict quality control while directing the folding of the nascent proteins through covalent modifications of the native structures, such as signal sequence processing, disulfide bonds formation, N-glycosylation, glycosyl-phosphatidyl-inositol addition, degradation and sorting (Anelli and Sitia 2008). In heterologous protein production, active secretion is a good marker of correct processing, and an indication that the expressed protein has passed the quality control of the eukaryotic cell (Yang et al. 2013). In fact, in those cases of expressed proteins presenting aberrant folding properties, secretion might be prevented by a cellular mechanism known as unfolded protein response (UPR), a system able to recognise misfolded proteins and target them for ER-associated protein degradation (ERAD) via ubiquitin-proteasome (Travers et al. 2000), as well as suppressing the related transcription. Evidence of silicatein retention in the yeast cells is revealed by time-lapse Western blot analyses on the yeast cell extracts, where we observed antibody recognition on proteins at molecular sizes different from the expected. Specifically, a unique band detected during the early hours of expression (likely silicatein fused to α-MF), progressively degrades to a smear covering low molecular weights in the late

stages of yeast cultivation, which likely indicate an intra-cellular proteolytic processing of the unsecreted protein (data not shown), due either to incorrect folding, or insufficient capability of the secretion system to clip the signal sequence and translocate the mature protein in the extra-cellular space. Strategies are being developed to tackle such bottleneck arising at any level of the protein secretion pathway, towards the development of engineered strains with altered features in protein trafficking, glycosylation mechanisms, proteolytic activity and protein folding (reviewed in Idiris et al. 2010).

4.5.7 Optimisation of the production yields in yeast.

An array of strategies for the enhancement of protein expression in yeast includes the targeting of different components of the system, like nature of the expression hosts, choice of promoters and signal sequences (He et al. 2012), generation of fusion proteins, introduction of chaperones and proteases, and rational design of the target proteins (Wang et al. 2014). Overexpression levels in the yeast systems can be also altered by both the chromosomal locus of integration, as well as the gene copy number (Mansur et al. 2005; Vassileva et al. 2001), implying that numerous clones has to be screened (Lin-Cereghino and Lin-Cereghino 2007). Therefore, it is of particular importance the restriction enzyme used for the linearisation of the expression vector, which determines the site of integration of the foreign gene into the yeast genome (Cregg et al. 1989). This can be the reason why in our experiments it was never possible to isolate positively expressing clones after linearisation with the enzyme SacI, while good results came only upon the usage of the restriction enzyme StuI. Another aspect to be considered regards the physiological requirements of the yeast. Multiple factors, such as pH, temperature, quantity of inducer and oxygenation can significantly affect the production yields and the secretion efficiency, therefore accurate experimental tests of all of the above play a crucial role in the development of the protein production method. Statistical approaches similar to those designed for the optimisation of bacterial bioprocesses (Papaneophytou and Kontopidis 2014) can be applied also in yeast, due to the ease of preparation of the cultivation medium; this is an advantage compared to the insect cells, where only pre-formulated media are used.

Preliminary results of medium optimisation showed an increased productivity of SIL in BMMY medium at slightly acidic pH values (Fig. 4.13A). The pH of the cultivation medium is usually related to promote high-density cell growth and address the stability of the target protein, and not with the specific purpose of inducing stress on the productive host (Mattanovich et al. 2004). Knowledge about the molecular basis for cellular pH stress response is still fragmentary, although some genes have been identified as specifically regulated in a pH manner in *S. cerevisiae*. Low-pH protein production improvements has been ascribed to a reduced protease activity of the host cells (Jahic et al. 2003), even if this correlation is not applicable in all cases, as demonstrated by optimums of production at pH 8 (Shi et al. 2003). This leads to the conclusion that every cultivation parameter is case specific, and requires experimental validation. As already mentioned the yeast cells are known to operate remarkable protein glycosylation. Therefore, we assume that secreted silicatein produced from the yeast is likely to have been covalently modified by glycosylation, known to improve the general solubility and stability of proteins (Dixon 1991). The latter constitutes an advantage in the production strategy, and may set the stage for a better understanding and eventually the disclosure of new functionalities and interaction patterns of silicatein.

5 SUMMARY

The marine world is an immense source of biodiversity that provides substances with striking potentials in medicinal chemistry and biotechnology. Sponges (Porifera) are marine animals that represent the most impressive example of organisms possessing the ability to metabolise silica through a family of enzymes known as silicateins. Complex skeletal structures (spicules) made of pure biogenic silica (biosilica) are produced under physiological conditions. Biosilica is a natural material comprising inorganic and organic components with unique mechanical, optical, and physico-chemical properties, including promising potential to be used for development of therapeutic agents in regenerative medicine. Unravelling the intimate physiological mechanisms occurring in sponges during the construction of their siliceous spicules is an on-going project, and several questions have been addressed by the studies proposed by our working group .

In this doctoral work, the recombinant DNA technology is exploited for functional and structural characterisation of silicatein. Its precursors are produced as fusion proteins with a chaperone tag (named TF-Ps), and a robust method for the overexpression of native soluble proteins in high concentrations has been developed. In addition, it is observed and proven experimentally that the maturation of silicatein is an autocatalytic event that: (i) can be modulated by rational use of protease inhibitors; (ii) is influenced by the temperature of the environment; (iii) only slightly depends on the pH. In the same experimental framework, observations on the dynamics in the maturation of silicateins allow a better understanding of how the axial filaments form during the early stages of spicule construction. In addition, the definition of new distinct properties of silicatein (termed “structure-guiding” and “structure-forming”) is introduced. By homology models and through comparisons with similar proteins (the cathepsins),

domains with significant surface hydrophobicity are identified as potential self-assembly mediators. Moreover, a high-throughput screening showed that TF-Ps could generate crystals under certain conditions, becoming promising for further structural studies. With the goal of optimise the properties of the recombinant silicatein, implementation of new production systems are tried for the first time. Success in the expression of silicatein-type proteins in insect and yeast cells, constitute a promising basis for further development, towards the establishment of an efficient method for the production of a high-value pure and soluble protein.

6 ZUSAMMENFASSUNG

Die marine Welt stellt eine immense Quelle an Biodiversität dar, die Stoffe mit beachtlichem Potential für die medizinische Chemie und die Biotechnologie bereitstellt. Schwämme (Porifera) sind marine Organismen und besitzen die einzigartige Fähigkeit, Glas mit Hilfe von Enzymen, den Silicateinen zu metabolisieren. Die komplexen, aus reinem biogenen Glas (Biosilikat) produzierten Skelettstrukturen (Spiculae) werden unter physiologischen Bedingungen produziert. Bei Biosilikat handelt es sich um ein natürliches Material, aus anorganischen und organischen Bestandteilen mit einzigartigen mechanischen, optischen und physikochemischen Eigenschaften. Darüber hinaus besitzt es ein vielversprechendes Potential in der Entwicklung therapeutischer Mittel innerhalb der regenerativen Medizin. Die physiologischen Mechanismen, welche an der Konstruktion der Glasspiculae beteiligt sind, konnten von unsere Arbeitsgruppe bereits weitgehend aufgeklärt werden. Im Rahmen dieser Arbeit wurde mit Hilfe der rekombinanten DNA-Technologie eine funktionelle und strukturelle Charakterisierung des Silicateins durchgeführt. Dessen Vorläuferproteine wurden über eine Fusion zwischen dem Protein und einem Chaperon-Tag (TF-Ps) hergestellt. Für die Überexpression des nativen und löslichen TF-Ps in hohen Konzentrationen wurde eine spezielle Methode entwickelt. Des Weiteren wurde beobachtet und experimentell bestätigt, dass die Reifung des Silicateins ein autokatalytischer Vorgang ist, welcher: (i) durch den entsprechenden Einsatz von Protease-Inhibitoren moduliert werden kann, (ii) durch die umgebende Temperatur beeinflusst werden kann, und (iii) nur leicht vom pH-Wert abhängig ist. Beobachtungen eines Phasenwechsels während der Reifung des Silicateins erlauben zusätzlich ein besseres Verständnis darüber, wie sich die Axialfilamente während der frühen Phase der Spikulum-Konstruktion formieren. Dies führte

zusätzlich zu Definitionen bestimmter Eigenschaften des Silicateins, die im Rahmen dieser Arbeit als ''structure-guiding'' und ''structure forming'' bezeichnet werden. Mit Hilfe von Homologie Modellen und durch den Vergleich mit ähnlichen Proteinen (Cathepsine), wurden Domänen mit signifikanter Oberflächenhydrophobizität als potentielle selbstassemblierende Mediatoren identifiziert. Zusätzlich zeigte ein high-throughput screening, dass TF-Ps unter bestimmten Bedingungen, Kristalle generieren können, die sich für weitere strukturelle Studien eignen. Neben der Optimierung der Eigenschaften des rekombinanten Silicateins, konnten erstmalig neue Produktionssysteme implementiert werden. Der Erfolg der Expression von Silikatein-artigen Proteinen in Insekten- und Hefezellen, liefert eine vielversprechende Basis für die weitere Entwicklung einer effizienten Methode zur Produktion von hochwertigen, rekombinanten, puren und löslichen Proteinen.

7 REFERENCES

Ahn, J. H., J. W. Keum and D. M. Kim (2011). "Expression screening of fusion partners from an *E. coli* genome for soluble expression of recombinant proteins in a cell-free protein synthesis system." PloS one **6**(11): e26875.

Akkermans, C., A. J. Van der Goot, P. Venema, H. Gruppen, J. M. Vereijken, E. Van der Linden and R. M. Boom (2007). "Micrometer-sized fibrillar protein aggregates from soy glycinin and soy protein isolate." Journal of Agricultural and Food Chemistry **55**(24): 9877-9882.

Allison, J. H., H. C. Agrawal and B. W. Moore (1974). "Effect of N,N,N',N'-tetramethylethylenediamine on the migration of proteins in SDS polyacrylamide gels." Analytical Biochemistry **58**(2): 592-601.

Amunugama, R., J. Groden and R. Fishel (2013). "The HsRAD51B-HsRAD51C stabilizes the HsRAD51 nucleoprotein filament." DNA repair **12**(9): 723-732.

Anelli, T. and R. Sitia (2008). "Protein quality control in the early secretory pathway." EMBO Journal **27**(2): 315-327.

Armirotti, A., G. Damonte, M. Pozzolini, F. Mussino, C. Cerrano, A. Salis, ... M. Giovine (2009). "Primary structure and post-translational modifications of silicatein beta from the marine sponge *Petrosia ficiformis* (Poiret, 1789)." Journal of Proteome Research **8**(8): 3995-4004.

Assenberg, R., P. T. Wan, S. Geisse and L. M. Mayr (2013). "Advances in recombinant protein expression for use in pharmaceutical research." Current Opinion in Structural Biology **23**(3): 393-402.

Bawazer, L. A., M. Izumi, D. Kolodin, J. R. Neilson, B. Schwenzer and D. E. Morse (2012). "Evolutionary selection of enzymatically synthesized semiconductors from biomimetic mineralisation vesicles." Proceedings of the National Academy of Sciences of the United States of America **109**(26): E1705-1714.

Bennett, J. and K. J. Scott (1971). "Quantitative staining of fraction I protein in polyacrylamide gels using Coomassie brillant blue." Analytical Biochemistry **43**(1): 173-182.

Bergquist, P. R. (1998). "Porifera". Invertebrate Zoology. D. T. Anderson. Melbourne, Oxford University Press.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Analytical Biochemistry **72**: 248-254.

Braun, P. and J. LaBaer (2003). "High throughput protein production for functional proteomics." Trends in Biotechnology **21**(9): 383-388.

Brown, W. C., J. DelProposto, J. R. Rubin, K. Lamiman, J. Carless and J. L. Smith (2011). "New ligation-independent cloning vectors compatible with a high-throughput platform for parallel construct expression evaluation using

baculovirus-infected insect cells." Protein Expression and Purification **77**(1): 34-45.

Cattaneo-Vietti, R., G. Bavestrello, C. Cerrano, M. Sarà, U. Benatti, M. Giovine and E. Gaino (1996). "Optical Fibres in an Antarctic Sponge." Nature **383**: 397-398.

Cha, J. N., K. Shimizu, Y. Zhou, S. C. Christiansen, B. F. Chmelka, G. D. Stucky and D. E. Morse (1999). "Silicatein filaments and subunits from a marine sponge direct the polymerisation of silica and silicones *in vitro*." Proceedings of the National Academy of Sciences of the United States of America **96**(2): 361-365.

Chang, A. C. and S. N. Cohen (1974). "Genome construction between bacterial species *in vitro*: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*." Proceedings of the National Academy of Sciences of the United States of America **71**(4): 1030-1034.

Chayen, N. E. (1998). "Comparative studies of protein crystallisation by vapour-diffusion and microbatch techniques." Acta Crystallographica D **54**(Pt 1): 8-15.

Churion, K. A. and S. E. Bondos (2012). "Identifying solubility-promoting buffers for intrinsically disordered proteins prior to purification." Methods in Molecular Biology **896**: 415-427.

Cohen, S. N., A. C. Chang, H. W. Boyer and R. B. Helling (1973). "Construction of biologically functional bacterial plasmids *in vitro*." Proceedings of the National Academy of Sciences of the United States of America **70**(11): 3240-3244.

Collins-Racie, L. A., J. M. McColgan, K. L. Grant, E. A. DiBlasio-Smith, J. M. McCoy and E. R. LaVallie (1995). "Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA." Bio/technology **13**(9): 982-987.

Coulombe, R., P. Grochulski, J. Sivaraman, R. Menard, J. S. Mort and M. Cygler (1996). "Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment." The EMBO journal **15**(20): 5492-5503.

Craig, N. L. (1991). "Tn7: a target site-specific transposon." Molecular Microbiology **5**(11): 2569-2573.

Crane, D. I. and S. J. Gould (1994). "The *Pichia pastoris* HIS4 gene: nucleotide sequence, creation of a non-reverting his4 deletion mutant, and development of HIS4-based replicating and integrating plasmids." Current Genetics **26**(5-6): 443-450.

Cregg, J. M., K. R. Madden, K. J. Barringer, G. P. Thill and C. A. Stillman (1989). "Functional-Characterisation of the 2 Alcohol Oxidase Genes from the Yeast *Pichia pastoris*." Molecular and Cellular Biology **9**(3): 1316-1323.

Cregg, J. M., J. L. Cereghino, J. Shi and D. R. Higgins (2000). "Recombinant protein expression in *Pichia pastoris*." Molecular Biotechnology **16**(1): 23-52.

Croce, G., A. Frache, M. Milanesio, L. Marchese, M. Causa, D. Viterbo, ... H. Amenitsch (2004). "Structural characterisation of siliceous spicules from marine sponges." Biophysical Journal **86**(1 Pt 1): 526-534.

Curnow, P., P. H. Bessette, D. Kisailus, M. M. Murr, P. S. Daugherty and D. E. Morse (2005a). "Enzymatic synthesis of layered titanium phosphates at low temperature and neutral pH by cell-surface display of silicatein-alpha." Journal of the American Chemical Society **127**(45): 15749-15755.

Currie, H. A. and C. C. Perry (2007). "Silica in plants: biological, biochemical and chemical studies." Annals of Botany **100**(7): 1383-1389.

Dale, G. E., C. Oefner and A. D'Arcy (2003). "The protein as a variable in protein crystallisation." Journal of Structural Biology **142**(1): 88-97.

Damasceno, L. M., C. J. Huang and C. A. Batt (2012). "Protein secretion in *Pichia pastoris* and advances in protein production." Applied Microbiology and Biotechnology **93**(1): 31-39.

De Marco, V., G. Stier, S. Blandin and A. de Marco (2004). "The solubility and stability of recombinant proteins are increased by their fusion to NusA." Biochemical and Biophysical Research Communications **322**(3): 766-771.

Demain, A. L. and P. Vaishnav (2009). "Production of recombinant proteins by microbes and higher organisms." Biotechnology Advances **27**(3): 297-306.

Demaster, D. J. (1981). "The Supply and Accumulation of Silica in the Marine-Environment." Geochimica Et Cosmochimica Acta **45**(10): 1715-1732.

Di Cera, E. and A. M. Cantwell (2001). "Determinants of thrombin specificity." Annals of the New Yorc Academy of Science **936**: 133-146.

Dixon, B. (1991). "Enzyme Expression - Glycosylation Enhances Stability." Bio-Technology **9**(5): 418-418.

Eckert, C., H. C. Schröder, D. Brandt, S. Perovic-Ottstadt and W. E. G. Müller (2006). "Histochemical and electron microscopic analysis of spiculogenesis in the demosponge *Suberites domuncula*." The Journal of Histochemistry and Cytochemistry **54**(9): 1031-1040.

Epstein, E. (1999). "Silicon." Annual review of Plant Physiology and Plant Molecular Biology **50**: 641-664.

Fairhead, M., K. A. Johnson, T. Kowatz, S. A. McMahon, L. G. Carter, M. Oke, ... C. F. van der Walle (2008). "Crystal structure and silica condensing activities of silicatein alpha-cathepsin L chimeras." Chemical Communications (15): 1765-1767.

Ferbitz, L., T. Maier, H. Patzelt, B. Bukau, E. Deuerling and N. Ban (2004). "Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins." Nature **431**(7008): 590-596.

Frampton, M. B. and P. M. Zelisko (2012). "A Comparison of Protease Active Sites and their Ability to Process Silicon-Based Substrates." Silicon **4**(1): 51-56.

Freymann, D. M., Y. Nakamura, P. J. Focia, R. Sakai and G. T. Swanson (2012). "Structure of a tetrameric galectin from *Cinachyrella* sp. (ball sponge)." Acta Crystallographica D **68**(Pt 9): 1163-1174.

Fuller, R. S., C. Brenner, P. Gluschankof and C. A. Wilcox (1991). "The Yeast Prohormone-Processing Kex2 Protease, an Enzyme with Specificity for Paired Basic Residues." Methods in Protein Sequence Analysis **1**: 205-214.

Garant, P. R. (2003). "Oral cells and tissues", Quintessence Publishing Co.

Garrone, R., T. L. Simpson and J. Pottu-Boumendil (1981). "Ultrastructure and deposition of silica in sponges". New York, Springer.

Graslund, S., P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, ... K. C. Gunsalus (2008). "Protein production and purification." Nature Methods **5**(2): 135-146.

Hartl, F. U., A. Bracher and M. Hayer-Hartl (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.

Hasilik, A., C. Wrocklage and B. Schröder (2009). "Intracellular trafficking of lysosomal proteins and lysosomes." International Journal of Clinical and Pharmacological Therapeutics **47 Suppl 1**: S18-33.

He, Z. Y., Y. K. Huang, Y. F. Qin, Z. G. Liu, D. L. Mo, P. Q. Cong and Y. S. Chen (2012). "Comparison of Alpha-Factor Preprosequence and a Classical Mammalian Signal Peptide for Secretion of Recombinant Xylanase xynB from Yeast *Pichia pastoris*." Journal of Microbiology and Biotechnology **22**(4): 479-483.

Hochuli, E., H. Dobeli and A. Schacher (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues." Journal of Chromatography **411**: 177-184.

Hoffmann, A., F. Merz, A. Rutkowska, B. Zachmann-Brand, E. Deuerling and B. Bukau (2006a). "Trigger factor forms a protective shield for nascent polypeptides at the ribosome." Journal of Biological Chemistry **281**(10): 6539-6545.

Hoffmann, F., M. Cornelius, J. Morell and M. Froba (2006b). "Silica-based mesoporous organic-inorganic hybrid materials." Angewandte Chemie **45**(20): 3216-3251.

Huang, G. C., Z. Y. Li and J. M. Zhou (2000). "Conformational specificity of trigger factor for the folding intermediates of alpha-lactalbumin." Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology **1480**(1-2): 77-82.

Idiris, A., H. Tohda, H. Kumagai and K. Takegawa (2010). "Engineering of protein secretion in yeast: strategies and impact on protein production." Applied Microbiology and Biotechnology **86**(2): 403-417.

Iler, R. K. (1979). "The Chemistry of Silica: solubility, polymerisation, colloid and surface properties and biochemistry". New York, Wiley.

Imasaki, T., G. Calero, G. Cai, K. L. Tsai, K. Yamada, F. Cardelli, ... Y. Takagi (2011). "Architecture of the Mediator head module." Nature **475**(7355): 240-243.

Imsiecke, G., R. Steffen, M. Custodio, R. Borojevic and W. E. G. Müller (1995). "Formation of spicules by sclerocytes from the freshwater sponge *Ephydatia muelleri* in short-term cultures in vitro." *In vitro* Cellular & Developmental Biology. Animal **31**(7): 528-535.

Inan, M. and M. M. Meagher (2001). "Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*." Journal of Bioscience and Bioengineering **92**(6): 585-589.

Jahic, M., M. Gustavsson, A. K. Jansen, M. Martinelle and S. O. Enfors (2003). "Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes." Journal of Biotechnology **102**(1): 45-53.

James, G. T. (1978). "Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers." Analytical Biochemistry **86**(2): 574-579.

Jarvis, D. L. (2009). "Baculovirus-Insect Cell Expression Systems." Guide to Protein Purification, Second Edition **463**: 191-222.

Kaluzhnaya, O. V., S. I. Belikov, H. C. Schröder, M. Wiens, M. Giovine, A. Krasko, ... W. E. G. Müller (2005). "Dynamics of skeleton formation in the Lake Baikal sponge *Lubomirskia baicalensis*. Part II. Molecular biological studies." Naturwissenschaften **92**(3): 134-138.

Kapust, R. B. and D. S. Waugh (1999). "*Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused." Protein science: a publication of the Protein Society **8**(8): 1668-1674.

Kaulmann, G., G. J. Palm, K. Schilling, R. Hilgenfeld and B. Wiederanders (2006). "The crystal structure of a Cys25 -> Ala mutant of human procathepsin S elucidates enzyme-prosequence interactions." Protein Science **15**(11): 2619-2629.

Keeping, M. G. and O. L. Kvedaras (2008). "Silicon as a plant defence against insect herbivory: response to Massey, Ennos and Hartley." The Journal of Animal Ecology **77**(3): 631-633.

Ki, M. R., K. B. Yeo and S. P. Pack (2013). "Surface immobilisation of protein via biosilification catalyzed by silicatein fused to glutathione S-transferase (GST)." Bioprocess and Biosystem Engineering **36**(5): 643-648.

Kisailus, D., Q. Truong, Y. Amemiya, J. C. Weaver and D. E. Morse (2006). "Self-assembled bifunctional surface mimics an enzymatic and templating protein for the synthesis of a metal oxide semiconductor." Proceedings of the National Academy of Sciences of the United States of America **103**(15): 5652-5657.

Kramer, G., A. Rutkowska, R. D. Wegrzyn, H. Patzelt, T. A. Kurz, F. Merz, ... B. Bukau (2004). "Functional dissection of *Escherichia coli* trigger factor: unraveling the function of individual domains." The Journal of Bacteriology **186**(12): 3777-3784.

Krasko, A., B. Lorenz, R. Batel, H. C. Schröder, I. M. Müller and W. E. G. Müller (2000). "Expression of silicatein and collagen genes in the marine sponge *Suberites domuncula* is controlled by silicate and myotrophin." European Journal of Biochemistry **267**(15): 4878-4887.

Krasko, A., H. C. Schröder, R. Batel, V. A. Grebenjuk, R. Steffen, I. M. Müller and W. E. G. Müller (2002). "Iron induces proliferation and morphogenesis in

primmorphs from the marine sponge *Suberites domuncula*." DNA and cell biology **21**(1): 67-80.

Kröger, N., S. Lorenz, E. Brunner and M. Sumper (2002). "Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis." Science **298**(5593): 584-586.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature **227**(5259): 680-685.

LaVallie, E. R., E. A. DiBlasio-Smith, L. A. Collins-Racie, Z. Lu and J. M. McCoy (2003). "Thioredoxin and related proteins as multifunctional fusion tags for soluble expression in *E. coli*." Methods in Molecular Biology **205**: 119-140.

Lebendiker, M. and T. Danieli (2014). "Production of prone-to-aggregate proteins." FEBS Letters **588**(2): 236-246.

Lehman, I. R. (1974). "DNA ligase: structure, mechanism, and function." Science **186**(4166): 790-797.

Lieberman, R. L., M. E. Peek and J. D. Watkins (2013). "Determination of soluble and membrane protein structures by X-ray crystallography." Methods in Molecular Biology **955**: 475-493.

Lilie, H., E. Schwarz and R. Rudolph (1998). "Advances in refolding of proteins produced in *E. coli*." Current Opinion in Biotechnology **9**(5): 497-501.

Lin-Cereghino, J. and G. P. Lin-Cereghino (2007). "Vectors and strains for expression." Methods in Molecular Biology **389**: 11-26.

Lodish, H. (2008). "Molecular Cell Biology", W. H. Freeman.

Luckow, V. A., S. C. Lee, G. F. Barry and P. O. Olins (1993). "Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*." Journal of Virology **67**(8): 4566-4579.

Ludlam, A. V., B. A. Moore and Z. Xu (2004). "The crystal structure of ribosomal chaperone trigger factor from *Vibrio cholerae*." Proceedings of the National Academy of Sciences of the United States of America **101**(37): 13436-13441.

Lutgens, F. K., E. J. Tarbuck and D. Tasa (2009). "Essentials of geology. Upper Saddle River", N.J., Pearson Prentice Hall.

Maier, R., C. Scholz and F. X. Schmid (2001). "Dynamic association of trigger factor with protein substrates." The Journal of Molecular Biology **314**(5): 1181-1190.

Malakhov, M. P., M. R. Mattern, O. A. Malakhova, M. Drinker, S. D. Weeks and T. R. Butt (2004). "SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins." Journal of Structural and Functional Genomics **5**(1-2): 75-86.

Mansur, M., C. Cabello, L. Hernandez, J. Pais, L. Varas, J. Valdes, ... E. Martinez (2005). "Multiple gene copy number enhances insulin precursor secretion in the yeast *Pichia pastoris*." Biotechnology Letters **27**(5): 339-345.

Marbach, A. and K. Bettenbrock (2012). "lac operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA." Journal of Biotechnology **157**(1): 82-88.

Marko, M. A., R. Chipperfield and H. C. Birnboim (1982). "A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder." Analytical Biochemistry **121**(2): 382-387.

Mattanovich, D., B. Gasser, H. Hohenblum and M. Sauer (2004). "Stress in recombinant protein producing yeasts." Journal of Biotechnology **113**(1-3): 121-135.

Mayer, G. (2005). "Rigid biological systems as models for synthetic composites." Science **310**(5751): 1144-1147.

McQueney, M. S., B. Y. Amegadzie, K. D'Alessio, C. R. Hanning, M. M. McLaughlin, D. McNulty, ... C. S. Jones (1997). "Autocatalytic activation of human cathepsin K." The Journal of Biological Chemistry **272**(21): 13955-13960.

Menard, R., E. Carmona, S. Takebe, E. Dufour, C. Plouffe, P. Mason and J. S. Mort (1998). "Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L *in vitro*." The Journal of Biological Chemistry **273**(8): 4478-4484.

Morrow, K. J. (2007). "Improving protein production strategies." Genetic Engineering News **27**(3): 37-39.

Müller, W. E. G. (2001). "How was metazoan threshold crossed? The hypothetical Urmetazoa." Comparative Biochemistry Physiology A **129**(2-3): 433-460.

Müller, W. E. G., A. Krasko, G. Le Pennec, R. Steffen, M. Wiens, M. S. Ammar, ... H. C. Schröder (2003). "Molecular mechanism of spicule formation in the demosponge *Suberites domuncula*: silicatein--collagen--myotrophin." Progress in Molecular and Subcellular Biology **33**: 195-221.

Müller, W. E. G., M. Rothenberger, A. Boreiko, W. Tremel, A. Reiber and H. C. Schröder (2005). "Formation of siliceous spicules in the marine demosponge *Suberites domuncula*." Cell and Tissue Research **321**(2): 285-297.

Müller, W. E. G., S. I. Belikov, W. Tremel, C. C. Perry, W. W. Gieskes, A. Boreiko and H. C. Schröder (2006). "Siliceous spicules in marine demosponges (example *Suberites domuncula)*." Micron **37**(2): 107-120.

Müller, W. E. G., A. Boreiko, U. Schlossmacher, X. Wang, M. N. Tahir, W. Tremel, ... H. C. Schröder (2007a). "Fractal-related assembly of the axial filament in the demosponge *Suberites domuncula*: relevance to biomineralisation and the formation of biogenic silica." Biomaterials **28**(30): 4501-4511.

Müller, W. E. G., U. Schlossmacher, C. Eckert, A. Krasko, A. Boreiko, H. Ushijima, ... H. C. Schröder (2007b). "Analysis of the axial filament in spicules of the demosponge *Geodia cydonium*: different silicatein composition in microscleres (asters) and megascleres (oxeas and triaenes)." European Journal of Cell Biology **86**(8): 473-487.

Müller, W. E. G., A. Boreiko, U. Schlossmacher, X. Wang, C. Eckert, K. Kropf, ... H. C. Schröder (2008a). "Identification of a silicatein(-related) protease in the giant spicules of the deep-sea hexactinellid *Monorhaphis chuni*." The Journal of Experimental Biology **211**(Pt 3): 300-309.

Müller, W. E. G., U. Schlossmacher, X. Wang, A. Boreiko, D. Brandt, S. E. Wolf, ... H. C. Schröder (2008b). "Poly(silicate)-metabolizing silicatein in siliceous spicules and silicasomes of demosponges comprises dual enzymatic activities (silica polymerase and silica esterase)." The FEBS Journal **275**(2): 362-370.

Müller, W. E. G., X. Wang, K. Kropf, A. Boreiko, U. Schlossmacher, D. Brandt, ... M. Wiens (2008c). "Silicatein expression in the hexactinellid *Crateromorpha meyeri*: the lead marker gene restricted to siliceous sponges." Cell and Tissue Research **333**(2): 339-351.

Müller, W. E. G., X. H. Wang, V. A. Grebenjuk, M. Korzhev, M. Wiens, U. Schlossmacher and H. C. Schröder (2012). "Nocturnin in the demosponge *Suberites domuncula*: a potential circadian clock protein controlling glycogenin synthesis in sponges." Biochemical Journal **448**: 233-242.

Müller, W. E. G., E. Mugnaioli, H. C. Schröder, U. Schlossmacher, M. Giovine, U. Kolb and X. Wang (2013a). "Hierarchical composition of the axial filament from spicules of the siliceous sponge *Suberites domuncula*: from biosilica-synthesizing nanofibrils to structure- and morphology-guiding triangular stems." Cell and Tissue Research **351**(1): 49-58.

Müller, W. E. G., H. C. Schröder, Z. Burghard, D. Pisignano and X. Wang (2013b). "Silicateins--a novel paradigm in bioinorganic chemistry: enzymatic synthesis of inorganic polymeric silica." Chemistry **19**(19): 5790-5804.

Müller, W. E. G., H. C. Schröder, S. Muth, S. Gietzen, M. Korzhev, V. A. Grebenjuk, ... X. Wang (2013c). "The silicatein propeptide acts as inhibitor/modulator of self-organisation during spicule axial filament formation." The FEBS journal **280**(7): 1693-1708.

Müller, W. E. G., H. C. Schröder, D. Pisignano, J. S. Markl and X. Wang (2013d). "Metazoan circadian rhythm: toward an understanding of a light-based zeitgeber in sponges." Integrative and Comparative Biology **53**(1): 103-117.

Müller, W. E. G., E. Tolba, H. C. Schröder, B. Diehl-Seifert, T. Link and X. Wang (2014a). "Biosilica-loaded poly(-caprolactone) nanofibers mats provide a morphogenetically active surface scaffold for the growth and mineralisation of the osteoclast-related SaOS-2 cells." Biotechnology Journal DOI: 10.1002/biot.201400277.

Müller, W. E. G., T. Link, H. C. Schröder, M. Korzhev, M. Neufurth, D. Brandt and X. Wang (2014b). "Dissection of the structure-forming activity from the structure-guiding activity of silicatein: a biomimetic molecular approach to print optical fibers." Journal of Materials Chemistry B **2**: 5368-5377

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich (1986). "Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction." Cold Spring Harbor Symposia on Quantitative Biology **51 Pt 1**: 263-273.

Murr, M. M. and D. E. Morse (2005). "Fractal intermediates in the self-assembly of silicatein filaments." Proceedings of the National Academy of Sciences of the United States of America **102**(33): 11657-11662.

Natalio, F., T. Link, W. E. G. Müller, H. C. Schröder, F. Z. Cui, X. Wang and M. Wiens (2010a). "Bioengineering of the silica-polymerizing enzyme silicatein-alpha for a targeted application to hydroxyapatite." Acta biomaterialia **6**(9): 3720-3728.

Neilson, J. R., N. C. George, M. M. Murr, R. Seshadri and D. E. Morse (2014). "Mesostructure from hydration gradients in demosponge biosilica." Chemistry **20**(17): 4956-4965.

Nissler, K., S. Kreusch, W. Rommerskirch, W. Strubel, E. Weber and B. Wiederanders (1998). "Sorting of non-glycosylated human procathepsin S in mammalian cells." Biological Chemistry **379**(2): 219-224.

Oberg, K., B. A. Chrunyk, R. Wetzel and A. L. Fink (1994). "Native-Like Secondary Structure in Interleukin-1-Beta Inclusion-Bodies by Attenuated Total Reflectance Ftir." Biochemistry **33**(9): 2628-2634.

Ohno, K., J. Sawada, S. Takiya, M. Kimoto, A. Matsumoto, T. Tsubota, ... Y. Suzuki (2013). "Silk gland factor-2, involved in fibroin gene transcription, consists of LIM homeodomain, LIM-interacting, and single-stranded DNA-binding proteins." The Journal of Biological Chemistry **288**(44): 31581-31591.

Papaneophytou, C. P. and G. Kontopidis (2014). "Statistical approaches to maximize recombinant protein expression in *Escherichia coli*: a general review." Protein Expression and Purification **94**: 22-32.

Perry, C. C. and T. Keeling-Tucker (2000). "Biosilicification: the role of the organic matrix in structure control." Journal of Biological Inorganic Chemistry **5**(5): 537-550.

Pozzolini, M., L. Sturla, C. Cerrano, G. Bavestrello, L. Camardella, A. M. Parodi, ... M. Giovine (2004). "Molecular cloning of silicatein gene from marine sponge *Petrosia ficiformis* (Porifera, Demospongiae) and development of primmorphs as a model for biosilicification studies." Marine Biotechnology (NY) **6**(6): 594-603.

Prado Figueroa, M., F. Barrera and N. N. Cesaretti (2008). "Chalcedony (a crystalline variety of silica): biogenic origin in electric organs from living *Psammobatis extenta* (family Rajidae)." Micron **39**(7): 1027-1035.

Qing, G., L. C. Ma, A. Khorchid, G. V. Swapna, T. K. Mal, M. M. Takayama, ... M. Inouye (2004). "Cold-shock induced high-yield protein production in *Escherichia coli*." Nature Biotechnology **22**(7): 877-882.

Redecke, L., K. Nass, D. P. DePonte, T. A. White, D. Rehders, A. Barty, ... H. N. Chapman (2013). "Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser." Science **339**(6116): 227-230.

Roberts, R. J. (1976). "Restriction endonucleases." Critical Reviews in Biochemistry **4**(2): 123-164.

Rowan, A. D., P. Mason, L. Mach and J. S. Mort (1992). "Rat procathepsin B. Proteolytic processing to the mature form *in vitro*." Journal of Biological Chemistry **267**(22): 15993-15999.

Sano, K., K. Maeda, M. Oki and Y. Maeda (2002). "Enhancement of protein expression in insect cells by a lobster tropomyosin cDNA leader sequence." FEBS Letters **532**(1-2): 143-146.

Schilling, B. M., J. C. Goodrick and N. C. Wan (2001). "Scale-up of a high cell density continuous culture with *Pichia pastoris* X-33 for the constitutive expression of rh-chitinase." Biotechnology Progress **17**(4): 629-633.

Schreiber, A., F. Stengel, Z. Zhang, R. I. Enchev, E. H. Kong, E. P. Morris, ... D. Barford (2011). "Structural basis for the subunit assembly of the anaphase-promoting complex." Nature **470**(7333): 227-232.

Schröder, E., T. Jonsson and L. Poole (2003a). "Hydroxyapatite chromatography: altering the phosphate-dependent elution profile of protein as a function of pH." Analytical Biochemistry **313**(1): 176-178.

Schröder, H. C., A. Krasko, G. Le Pennec, T. Adell, M. Wiens, H. Hassanein, ... W. E. G. Müller (2003b). "Silicase, an enzyme which degrades biogenous amorphous silica: contribution to the metabolism of silica deposition in the demosponge *Suberites domuncula*." Progress in Molecular and Subcellular Biology **33**: 249-268.

Schröder, H. C., S. Perovic-Ottstadt, M. Rothenberger, M. Wiens, H. Schwertner, R. Batel, ... W. E. G. Müller (2004). "Silica transport in the demosponge *Suberites domuncula*: fluorescence emission analysis using the PDMPO probe and cloning of a potential transporter." Biochemical Journal **381**(Pt 3): 665-673.

Schröder, H. C., X. Wang, W. Tremel, H. Ushijima and W. E. G. Müller (2008). "Biofabrication of biosilica-glass by living organisms." Natural Product Reports **25**(3): 455-474.

Schröder, H. C., U. Schlossmacher, A. Boreiko, F. Natalio, M. Baranowska, D. Brandt, ... W. E. G. Müller (2009). "Silicatein: nanobiotechnological and biomedical applications." Progress in Molecular and Subcellular Biology **47**: 251-273.

Schröder, H. C., X. Wang, A. Manfrin, S. H. Yu, V. A. Grebenjuk, M. Korzhev, ... W. E. G. Müller (2012a). "Acquisition of structure-guiding and structure-forming properties during maturation from the pro-silicatein to the silicatein form." The Journal of Biological Chemistry **287**(26): 22196-22205.

Schröder, H. C., X. H. Wang, M. Wiens, B. Diehl-Seifert, K. Kropf, U. Schlossmacher and W. E. G. Müller (2012b). "Silicate modulates the cross-talk between osteoblasts (SaOS-2) and osteoclasts (RAW 264.7 cells): inhibition of osteoclast growth and differentiation." The Journal of Cellular Biochemistry **113**(10): 3197-3206.

Schröder, H. C., M. Wiens, U. Schlossmacher, D. Brandt and W. E. G. Müller (2012c). "Silicatein-Mediated Polycondensation of Orthosilicic Acid: Modeling of a Catalytic Mechanism Involving Ring Formation." Silicon **4**(1): 33-38.

Sedlak, E. and N. C. Robinson (2009). "Sequential dissociation of subunits from bovine heart cytochrome C oxidase by urea." Biochemistry **48**(34): 8143-8150.

Serwer, P. and M. E. Pichler (1978). "Electrophoresis of bacteriophage T7 and T7 capsids in agarose gels." Journal of Virology **28**(3): 917-928.

Shapiro, A. L. and J. V. Maizel, Jr. (1969). "Molecular weight estimation of polypeptides by SDS-polyacrylamide gel electrophoresis: further data concerning resolving power and general considerations." Analytical Biochemistry **29**(3): 505-514.

Shi, X. Z., T. Karkut, M. Chamankhah, M. Alting-Mees, S. M. Hemmingsen and D. Hegedus (2003). "Optimal conditions for the expression of a single-chain antibody (scFv) gene in *Pichia pastoris*." Protein Expression and Purification **28**(2): 321-330.

Shimizu, K., J. Cha, G. D. Stucky and D. E. Morse (1998). "Silicatein alpha: cathepsin L-like protein in sponge biosilica." Proceedings of the National Academy of Sciences of the United States of America **95**(11): 6234-6238.

Shrive, A. K., J. B. Moeller, I. Burns, J. M. Paterson, A. J. Shaw, A. Schlosser, ... U. Holmskov (2014). "Crystal structure of the tetrameric fibrinogen-like recognition domain of fibrinogen C domain containing 1 (FIBCD1) protein." The Journal of Biological Chemistry **289**(5): 2880-2887.

Sivaraman, J., M. Lalumiere, R. Menard and M. Cygler (1999). "Crystal structure of wild-type human procathepsin K." Protein Science **8**(2): 283-290.

Sivaraman, J., D. K. Nagler, R. Zhang, R. Menard and M. Cygler (2000). "Crystal structure of human procathepsin X: a cysteine protease with the proregion covalently linked to the active site cysteine." The Journal of Molecular Biology **295**(4): 939-951.

Snell, E. H., J. R. Luft, S. A. Potter, A. M. Lauricella, S. M. Gulde, M. G. Malkowski, ... G. T. Detitta (2008). "Establishing a training set through the visual analysis of crystallisation trials. Part I: approximately 150,000 images." Acta Crystallogrica D **64**(Pt 11): 1123-1130.

Striegel, A. M. (2008). "Size-exclusion chromatography: smaller, faster, multi-detection, and multi-dimensions." Analytical and Bioanalytical Chemistry **390**(1): 303-305.

Tahir, M. N., P. Theato, W. E. G. Müller, H. C. Schröder, A. Janshoff, J. Zhang, ... W. Tremel (2004). "Monitoring the formation of biosilica catalysed by histidine-tagged silicatein." Chemical Communications **24**: 2848-2849.

Tahir, M. N., P. Theato, W. E. G. Müller, H. C. Schröder, A. Boreiko, S. Faiss, ... W. Tremel (2005). "Formation of layered titania and zirconia catalysed by surface-bound silicatein." Chemical Communications **44**: 5533-5535.

Tahir, M. N., P. Theato, P. Oberle, G. Melnyk, S. Faiss, U. Kolb, ... W. Tremel (2006). "Facile synthesis and characterisation of functionalized, monocrystalline rutile TiO2 nanorods." Langmuir **22**(12): 5209-5212.

Takahashi, M., C. Goto, K. Ishikawa, I. Matsuda, S. Toriyama and K. Tsuchiya (2003). "Rice stripe virus 23.9 K protein aggregates and forms inclusion bodies in cultured insect cells and virus-infected plant cells." Archives of Virology **148**(11): 2167-2179.

Towbin, H., T. Staehelin and J. Gordon (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." Proceedings of the National Academy of Sciences of the United States of America **76**(9): 4350-4354.

Travers, K. J., C. K. Patil, L. Wodicka, D. J. Lockhart, J. S. Weissman and P. Walter (2000). "Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation." Cell **101**(3): 249-258.

Uhlen, M. (2008). "Affinity as a tool in life science." Biotechniques **44**(5): 649-654.

Uriz, M. J., X. Turon and M. A. Becerro (2000). "Silica deposition in Demosponges: spiculogenesis in *Crambe crambe*." Cell and Tissue Research **301**(2): 299-309.

Vassileva, A., D. A. Chugh, S. Swaminathan and N. Khanna (2001). "Effect of copy number on the expression levels of hepatitis B surface antigen in the methylotrophic yeast *Pichia pastoris*." Protein Expression and Purification **21**(1): 71-80.

Vergara, A., M. Grassi, F. Sica, E. Pizzo, G. D'Alessio, L. Mazzarella and A. Merlino (2013). "A novel interdomain interface in crystallins: structural characterisation of the betagamma-crystallin from *Geodia cydonium* at 0.99 A resolution." Acta Crystallographica D **69**(Pt 6): 960-967.

Voet, D. and J. G. Voet (2004). "Biochemistry". 3rd edition J. Wiley&Sons

Vogl, T., F. S. Hartner and A. Glieder (2013). "New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*." Current Opinion in Biotechnology **24**(6): 1094-1101.

Vologodskii, A. V., A. V. Lukashin, V. V. Anshelevich and M. D. Frank-Kamenetskii (1979). "Fluctuations in superhelical DNA." Nucleic Acids Research **6**(3): 967-982.

Walsh, S. and B. Barrell (1996). "The S*accharomyces cerevisiae* genome on the World Wide Web." Trends in Genetics **12**(7): 276-277.

Wang, P., L. Huang, H. Jiang, J. Tian, X. Chu and N. Wu (2014a). "Improving the secretion of a methyl parathion hydrolase in *Pichia pastoris* by modifying its N-terminal sequence." PloS one **9**(5): e96974.

Wang, S.-F., X. Wang, M. Wiens, H. C. Schröder and W. E. G. Müller (2011a). "Biosilica-glass formation using enzymes from sponges [silicatein]: Basic aspects and application in biomedicine [bone reconstitution material and osteoporosis]." Frontiers of Materials Science **5**(3): 266-281.

Wang, X., L. Gan, K. P. Jochum, H. C. Schröder and W. E. G. Müller (2011b). "The largest Bio-Silica Structure on Earth: The Giant Basal Spicule from the Deep-Sea Glass Sponge *Monorhaphis chuni*." Evidence-Based Complementary and Alternative Medicine **2011**: 540987.

Wang, X., M. Wiens, H. C. Schröder, U. Schlossmacher, D. Pisignano, K. P. Jochum and W. E. G. Müller (2011c). "Evagination of cells controls bio-silica formation and maturation during spicule formation in sponges." PloS one **6**(6): e20523.

Wang, X., H. C. Schröder and W. E. G. Müller (2014b). "Enzyme-based biosilica and biocalcite: biomaterials for the future in regenerative medicine." Trends in Biotechnology. DOI: 10.1016/j.tibtech.2014.05.004

Wang, X. H., L. S. Zeng, M. Wiens, U. Schlossmacher, K. P. Jochum, H. C. Schröder and W. E. G. Müller (2011d). "Evidence for a biogenic, microorganismal origin of rock varnish from the Gangdese Belt of Tibet." Micron **42**(5): 401-411.

Waring, M. J. (1965). "Complex formation between ethidium bromide and nucleic acids." The Journal of Molecular Biology **13**(1): 269-282.

Wiens, M., M. Bausen, F. Natalio, T. Link, U. Schlossmacher and W. E. G. Müller (2009). "The role of the silicatein-alpha interactor silintaphin-1 in biomimetic biomineralisation." Biomaterials **30**(8): 1648-1656.

Wiens, M., H. C. Schröder, X. Wang, T. Link, D. Steindorf and W. E. G. Müller (2011). "Isolation of the silicatein-alpha interactor silintaphin-2 by a novel solid-phase pull-down assay." Biochemistry **50**(12): 1981-1990.

Wiens, M., T. Link, T. A. Elkhooly, S. Isbert and W. E. G. Müller (2012b). "Formation of a micropatterned titania photocatalyst by microcontact printed silicatein on gold surfaces." Chemical Communications **48**(92): 11331-11333.

Wisser, K. C. and A. Gafni (1994). "Spontaneous and Chaperonin-Assisted Refolding of Rabbit Muscle Glyceraldehyde-3-Phosphate Dehydrogenase." Biophysical journal **66**(2): A134-A134.

Wolf, S. E., U. Schlossmacher, A. Pietuch, B. Mathiasch, H. C. Schröder, W. E. G. Müller and W. Tremel (2010). "Formation of silicones mediated by the sponge enzyme silicatein-alpha." Dalton transactions **39**(39): 9245-9249.

Xiao, R., S. Anderson, J. Aramini, R. Belote, W. A. Buchwald, C. Ciccosanti, ... T. B. Acton (2010). "The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium." The Journal of Structural Biology **172**(1): 21-33.

Yang, F., S. Zhang, G. Jin, X. Lin and Z. K. Zhao (2011). "Purification and characterisation of a beta-1,3-glucomannanase expressed in *Pichia pastoris*." Enzyme and Microbial Technology **49**(2): 223-228.

Yang, S., Y. Kuang, H. Li, Y. Liu, X. Hui, P. Li, ... D. Wu (2013). "Enhanced production of recombinant secretory proteins in *Pichia pastoris* by optimizing Kex2 P1' site." PloS one **8**(9): e75347.

Young, C. L., Z. T. Britton and A. S. Robinson (2012). "Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications." The Biotechnology Journal **7**(5): 620-634.

Zhou, Y., K. Shimizu, J. N. Cha, G. D. Stucky and D. E. Morse (1999). "Efficient catalysis of polysiloxane synthesis by silicatein alpha requires specific hydroxy and

imidazole functionalities." Angewandte Chemie-International Edition **38**(6): 780-782.

Zlotnikov, I., P. Werner, H. Blumtritt, A. Graff, Y. Dauphin, E. Zolotoyabko and P. Fratzl (2014). "A perfectly periodic three-dimensional protein/silica mesoporous structure produced by an organism." Advanced Materials **26**(11): 1682-1687.

Zou, Z., L. Cao, P. Zhou, Y. Su, Y. Sun and W. Li (2008). "Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*." Journal of Biotechnology **135**(4): 333-339.

8 APPENDICES

8.1 ABBREVIATIONS

§ = paragraph / chapter

3D = three dimensional

6xH = hexahistidine tag

8xE = octaglutamate tag

8xH = octahistidine tag

aa = aminoacid

CD = circular dichroism

cDNA = complementary DNA

CV = column volume

dH_2O = distilled water

DLS = dynamic light scattering

DNA = deoxyribo-nucleic acid

dNTPs = deoxynucleotides

dsDNA = double strand DNA

E. coli = *Escherichia coli*

Fig. = Figure

fwd = forward primer

HEPA = High Efficiency Particulate Air filter

Hi5 = High Five cells

HPLC = high performance liquid chromatography

HT = high throughput

HyAp = hydroxyapatite

IMAC = immobilisation on metal affinity chromatography

LB = Lysogenic Broth (Miller) medium

MCS = multi-cloning site

mRNA = messenger RNA

nt = nucleotide

OD = optical density

ORF = open reading frame

PCR = polymerase chain reaction

PRE = pre-pro-silicatein (from synthetic DNA)

P1 = pre-pro-silicatein

P2 = pro-silicatein

PDB = Protein Data Bank

r.t. = room temperature

rev = reverse primer

RMSD = root-mean-square deviation of atomic positions

RNA = ribo-nucleic acid

rpm = rounds per minute

S-S bond = disulfide bond

SAXS = small angle X-ray scattering

SDS-PAGE = sodium dodecyl sulphate – polyacrylamide gel electrophoresis

SIL = mature silicatein (from synthetic DNA)

Tab. = Table

TEM = transmission electron microscopy

TEOS = tetraethoxysilane

TF = Trigger Factor

TF-P = fusion protein TF - P1 or TF-P2

Thr = thrombin

UV = ultraviolet

Vir# = Viral stock #

Xa = factor Xa

8.2 LIST OF FIGURES

8.3 LIST OF TABLES

8.4 DNA CODING SEQUENCES

P1 (used in § 3)

990 coding nucleotides + stop codon

cDNA

ATGCTTGTCACAGTGGTAGTACTGGGTCTACTGGGGTTTGCTTCTGCAGCCCAGCCCAAGT
TTGAATTTGTAGAAGAATGGCAGCTGTGGAAGTCCACTCACTCTAAGATGTACGAGTCAC
AGTTAATGGAACTCGAAAGACATCTGACGTGGCTCTCCAATAAGAAATATATCGAGCAAC
ACAATGTCAACTCACACATTTTCGGTTTTACTCTGGCAATGAACCAGTTTGGAGATCTGAG
TGAATTGGAGTATGCTAACTATCTTGGCCAGTATCGCATTGAGGATAAAAAATCTGGCAA
CTACTCAAAGACTTTTCAGCGTGATCCTCTACAGGACTACCCTGAAGCTGTAGACTGGAG
AACCAAAGGAGCTGTCACGGCTGTCAAGGACCAGGGAGACTGTGGTGCTAGCTATGCTT
TCAGTGCTATGGGTGCTTTGGAGGGTGCTAATGCTTTAGCCAAGGGAAATGCAGTATCTC
TCAGTGAACAGAACATCATTGATTGCTCGATTCCTTACGGTAACCACGGTTGTCATGGAG
GCAATATGTATGATGCTTTTTTGTATGTCATCGCTAACGAGGGGGTCGATCAGGACAGTG
CATATCCATTTGTAGGAAAGCAATCCAGCTGCAACTATAATAGTAAATACAAAGGTACAT
CAATGTCGGGGATGGTGTCAATCAAAAGTGGTAGTGAGTCTGACTTACAAGCAGCTGTTT
CAAACGTTGGCCCTGTATCTGTTGCTATTGATGGTGCTAACAGTGCCTTCAGGTTTTACTA
CAGTGGTGTCTATGACTCATCACGATGCTCTAGTAGTAGTCTTAACCACGCAATGGTAGTC
ACTGGATACGGATCATACAATGGGAAAAAATACTGGCTGGCCAAGAATAGCTGGGGAAC
TAACTGGGGTAACAGTGGCTATGTGATGATGGCTCGCAACAAGTACAACCAGTGTGGAAT
TGCTACCGATGCATCTTATCCCACCCTATGA

P2 (used in § 3)

948 coding nucleotides + stop codon

cDNA

ATGGCAGCCCAGCCCAAGTTTGAATTTGTAGAAGAATGGCAGCTGTGGAAGTCCACTCAC
TCTAAGATGTACGAGTCACAGTTAATGGAACTCGAAAGACATCTGACGTGGCTCTCCAAT
AAGAAATATATCGAGCAACACAATGTCAACTCACACATTTTCGGTTTTACTCTGGCAATGA
ACCAGTTTGGAGATCTGAGTGAATTGGAGTATGCTAACTATCTTGGCCAGTATCGCATTGA
GGATAAAAAATCTGGCAACTACTCAAAGACTTTTCAGCGTGATCCTCTACAGGACTACCC
TGAAGCTGTAGACTGGAGAACCAAAGGAGCTGTCACGGCTGTCAAGGACCAGGGAGACT
GTGGTGCTAGCTATGCTTTCAGTGCTATGGGTGCTTTGGAGGGTGCTAATGCTTTAGCCAA
GGGAAATGCAGTATCTCTCAGTGAACAGAACATCATTGATTGCTCGATTCCTTACGGTAAC
CACGGTTGTCATGGAGGCAATATGTATGATGCTTTTTTGTATGTCATCGCTAACGAGGGGG
TCGATCAGGACAGTGCATATCCATTTGTAGGAAAGCAATCCAGCTGCAACTATAATAGTA
AATACAAAGGTACATCAATGTCGGGGATGGTGTCAATCAAAAGTGGTAGTGAGTCTGACT
TACAAGCAGCTGTTTCAAACGTTGGCCCTGTATCTGTTGCTATTGATGGTGCTAACAGTGC
CTTCAGGTTTTACTACAGTGGTGTCTATGACTCATCACGATGCTCTAGTAGTAGTCTTAACC
ACGCAATGGTAGTCACTGGATACGGATCATACAATGGGAAAAAATACTGGCTGGCCAAG
AATAGCTGGGGAACTAACTGGGGTAACAGTGGCTATGTGATGATGGCTCGCAACAAGTA
CAACCAGTGTGGAATTGCTACCGATGCATCTTATCCCACCCTATGA

PRE (used in § 4)

1032 coding nucleotides + stop codon

synthetic DNA

ATGGTAGTGACCGTGGTGGTGCTGGGCCTGCTGGGTTTCGCTTCTGCTGCTCAGCCCAAG
TTCGAGTTCGTGGAAGAGTGGCAGCTGTGGAAGTCCACCCACTCCAAGATGTACGAGTCC
CAGCTGATGGAACTCGAGCGTCACCTGACCTGGCTGTCCAACAAGAAGTACATCGAGCA
GCACAACGTGAACTCCCACATCTTCGGTTTCACCCTGGCTATGAACCAGTTCGGCGACCT
GTCCGAGCTGGAATACGCTAACTACCTGGGCCAGTACCGTATCGAGGACAAGAAGTCCG
GCAACTACTCCAAGACCTTCCAGCGTGACCCCCTGCAGGACTACCCTGAGGCTGTGGACT
GGCGTACCAAGGGTGCTGTGACCGCTGTGAAGGACCAGGGCGATTGCGGTGCTGCTTAC
GCTTTCTCCGCTATGGGTGCTCTCGAGGGTGCTAACGCTCTGGCTAAGGGCAACGCTGTG
TCCCTGTCTGAGCAGAACATCATCGACTGCTCCATCCCTTACGGCAACCACGGTTGCCAC
GGTGGAAATATGTACGACGCTTTCCTGTACGTGATCGCTAACGAGGGTGTCGACCAGGAC
TCCGCTTACCCCTTCGTGGGCAAGCAGTCCTCCTGCAACTACAACTCCAAGTACAAGGGC
ACCTCCATGTCCGGCATGGTGTCCATCAAGTCCGGTTCCGAGTCCGACCTGCAGGCTGCT
GTGTCCAACGTCGGTCCTGTGTCCGTGGCTATCGACGGCGCTAACTCCGCTTTCCGCTTCT
ACTACTCCGGTGTCTACGACTCCTCCAGGTGCTCCTCCTCCTCTCTGAACCACGCTATGGT
GGTCACCGGTTACGGTTCCTACAACGGAAAGAAGTACTGGCTGGCTAAGAACTCCTGGG
GCACCAACTGGGGCAACTCCGGTTACGTGATGATGGCTCGTAACAAGTACAACCAGTGC
GGTATCGCTACCGACGCTTCCTACCCCACCCTGCTGGTGCCTCGTGGTTCTCATCATCATC
ACCATCACCACCACTAA

SIL (used in § 4)

747 coding nucleotides + stop codon

synthetic DNA

ATGGTAGTGACCGTGGTGGTGCTGGGCCTGCTGGGTTTCGCTTCCGCTGCTGACTACCCC
GAGGCTGTGGACTGGCGTACCAAGGGTGCTGTGACCGCTGTGAAGGACCAGGGCGACTG
CGGTGCTTCCTACGCTTTCTCCGCTATGGGTGCTCTCGAGGGTGCTAACGCTCTGGCTAAG
GGCAACGCTGTGTCCCTGTCCGAGCAGAACATCATCGACTGCTCCATCCCTTACGGCAAC
CACGGTTGCCACGGTGGAAATATGTACGACGCTTTCCTGTACGTGATCGCTAACGAGGGT
GTCGACCAGGACTCCGCTTACCCCTTCGTGGGCAAGCAGTCCTCCTGCAACTACAACTCC
AAGTACAAGGGCACCTCCATGTCCGGCATGGTGTCCATCAAGTCCGGTTCCGAGTCCGAC
CTGCAGGCTGCTGTGTCCAACGTCGGTCCTGTGTCCGTGGCTATCGACGGCGCTAACTCC
GCTTTCCGCTTCTACTACTCCGGTGTCTACGACTCCTCCAGGTGCTCCTCCTCCTCTCTGAA
CCACGCTATGGTGGTCACCGGTTACGGTTCCTACAACGGCAAGAAGTACTGGCTGGCTAA
GAACTCCTGGGGCACCAACTGGGGCAACTCCGGTTACGTGATGATGGCTCGTAACAAGT
ACAACCAGTGCGGTATCGCTACCGACGCTTCCTACCCCACCCTGCTGGTGCCTCGTGGTT
CTCATCATCATCACCATCACCACCACTAA

8.5 PROTEIN SEQUENCES

TF-P1 (expressed in §3)

88,56 kDa

MNHKVHHHHHHMQVSVETTQGLGRRVTITIAADSIETAVKSELVNVAKKVRIDGFRKGKVPM
NIVAQRYGASVRQDVLGDLMSRNFIDAIIKEKINPAGAPTYVPGEYKLGEDFTYSVEFEVYPEVE
LQGLEAIEVEKPIVEVTDADVDGMLDTLRKQQATWKEKDGAVEAEDRVTIDFTGSVDGEEFEG
GKASDFVLAMGQGRMIPGFEDGIKGHKAGEEFTIDVTFPEEYHAENLKGKAAKFAINLKKVEER
ELPELTAEFIKRFGVEDGSVEGLRAEVRKNMERELKSAIRNRVKSQAIEGLVKANDIDVPAALIDS
EIDVLRRQAAQRFGGNEKQALELPRELFEEQAKRRVVVGLLLGEVIRTNELKADEERVKGLIEEM
ASAYEDPKEVIEFYSKNKELMDNMRNVALEEQAVEAVLAKAKVTEKETTFNELMNQQASAGLE
VLFQGPSAGLVPRGSGGIEGRHMLVTVVVLGLLGFASAAQPKFEFVEEWQLWKSTHSKMYES
QLMELERHLTWLSNKKYIEQHNVNSHIFGFTLAMNQFGDLSELEYANYLGQYRIEDKKSGNYS
KTFQRDPLQDYPEAVDWRTKGAVTAVKDQGDCGASYAFSAMGALEGANALAKGNAVSLSE
QNIIDCSIPYGNHGCHGGNMYDAFLYVIANEGVDQDSAYPFVGKQSSCNYNSKYKGTSMSG
MVSIKSGSESDLQAAVSNVGPVSVAIDGANSAFRFYYSGVYDSSRCSSSSLNHAMVVTGYGSY
NGKKYWLAKNSWGTNWGNSGYVMMARNKYNQCGIATDASYPTLVDLQSR

TF-P2 (expressed in §3)

87,16 kDa

MNHKVHHHHHHMQVSVETTQGLGRRVTITIAADSIETAVKSELVNVAKKVRIDGFRKGKVPM
NIVAQRYGASVRQDVLGDLMSRNFIDAIIKEKINPAGAPTYVPGEYKLGEDFTYSVEFEVYPEVE
LQGLEAIEVEKPIVEVTDADVDGMLDTLRKQQATWKEKDGAVEAEDRVTIDFTGSVDGEEFEG
GKASDFVLAMGQGRMIPGFEDGIKGHKAGEEFTIDVTFPEEYHAENLKGKAAKFAINLKKVEER
ELPELTAEFIKRFGVEDGSVEGLRAEVRKNMERELKSAIRNRVKSQAIEGLVKANDIDVPAALIDS
EIDVLRRQAAQRFGGNEKQALELPRELFEEQAKRRVVVGLLLGEVIRTNELKADEERVKGLIEEM
ASAYEDPKEVIEFYSKNKELMDNMRNVALEEQAVEAVLAKAKVTEKETTFNELMNQQASAGLE
VLFQGPSAGLVPRGSGGIEGRHMAAQPKFEFVEEWQLWKSTHSKMYESQLMELERHLTWLS
NKKYIEQHNVNSHIFGFTLAMNQFGDLSELEYANYLGQYRIEDKKSGNYSKTFQRDPLQDYPE
AVDWRTKGAVTAVKDQGDCGASYAFSAMGALEGANALAKGNAVSLSEQNIIDCSIPYGNHG
CHGGNMYDAFLYVIANEGVDQDSAYPFVGKQSSCNYNSKYKGTSMSGMVSIKSGSESDLQA
AVSNVGPVSVAIDGANSAFRFYYSGVYDSSRCSSSSLNHAMVVTGYGSYNGKKYWLAKNSW
GTNWGNSGYVMMARNKYNQCGIATDASYPTLVDLQSR

PRE (expressed in §4)

38,0 kDa (36,4 kDa without signal sequence)

MVVTVVVLGLLGFASAAQPKFEFVEEWQLWKSTHSKMYESQLMELERHLTWLSNKKYIEQHN
VNSHIFGFTLAMNQFGDLSELEYANYLGQYRIEDKKSGNYSKTFQRDPLQDYPEAVDWRTKG
AVTAVKDQGDCGAAYAFSAMGALEGANALAKGNAVSLSEQNIIDCSIPYGNHGCHGGNMY
DAFLYVIANEGVDQDSAYPFVGKQSSCNYNSKYKGTSMSGMVSIKSGSESDLQAAVSNVGPV
SVAIDGANSAFRFYYSGVYDSSRCSSSSLNHAMVVTGYGSYNGKKYWLAKNSWGTNWGNSG
YVMMARNKYNQCGIATDASYPTLLVPRGSHHHHHHHH

SIL (expressed in §4)

26,5 kDa (24,9 kDa without signal sequence)

MVVTVVVLGLLGFASAADYPEAVDWRTKGAVTAVKDQGDCGASYAFSAMGALEGANALAK
GNAVSLSEQNIIDCSIPYGNHGCHGGNMYDAFLYVIANEGVDQDSAYPFVGKQSSCNYNSKY
KGTSMSGMVSIKSGSESDLQAAVSNVGPVSVAIDGANSAFRFYYSGVYDSSRCSSSSLNHAMV
VTGYGSYNGKKYWLAKNSWGTNWGNSGYVMMARNKYNQCGIATDASYPTLLVPRGSHHH
HHHHH

9 ACKNOWLEDGEMENTS

Alberto Manfrin

Schwedenstrasse 5, 55128 Mainz (Germany)
alberto.manfrin@gmail.com
http://de.linkedin.com/in/albertomanfrin
+4915783805525

Work experience

- **University Medical Center of the Johannes Gutenberg-University Mainz – Institute for Physiological Chemistry (Germany)** — ***September 2012 - Present***

  Early Stage Researcher **ERC**

  Biomineralization in sponges: expression, refolding, purification, characterization and crystallization of recombinant proteins (BIOSILICA EU ERC project).

- **Shanghai Institute of Materia Medica (SIMM), Chinese Academy of Science (China)** — ***February 2014 – July 2014***

  Early Stage Researcher **Marie Curie** (Staff Exchange)

  Sponges as a source of pharmaceuticals: Purification and characterization of bioactive compounds (MarBioTec *EU-CN* Marie Curie EU project).

- **National Center For Scientific Research (NCSR) "Demokritos" (Greece)** — ***May 2010 – August 2012***

  Early Stage Researcher **Marie Curie**

  Biomineralization in sponges: crystallization, biophysics and optimization of the production of recombinant proteins in insect cells and yeast (collaboration with University of Mainz and Pasteur Institute Athens in the frame of BIOMINTEC Marie Curie EU project).

- **Istituto di Ricerche Biotecnologiche (IRB) (Italy)** — ***February 2007 – April 2010***

  R&D Researcher full-time

  Wide range of activities in plant biotechnology:
  (i) optimization of physical/chemical parameters ijn plant cell cultures fermentation both in small and large bioreactor scale:
  (ii) establishment of cell cryopreservation methods for plant cell lines;
  (iii) development of innovative products for the cosmetic/nutraceutic market.

Education

Ph.D Degree (Biology)

May 2010 - Present

Johannes Gutenberg University Mainz (Germany): Thesis: “Preparation and characterisation of recombinant silicatein α from the sponge *Suberites domuncula*

Master's Degree (Industrial Biotechnology)

October 2005 – December 2007

University of Padua (Italy): Thesis: "Implementation of plant cell lines for the large scale production of forsythoside B" (research project conducted at IRB – Istituto di Ricerche Biotecnologiche – Vicenza, Italy)

Bachelor's Degree (Biotechnology)

October 2002 – July 2005

University of Padua (Italy): Thesis: "Plant cell cultures for the production of the antitumoral camptothecin"

Secondary School (Science and Technology)

September 1997 – July 2002

Liceo Scientifico "G.B. Quadri" (Vicenza, Italy)

Summary

Scientific interests

Recombinant proteins production, Biopharmaceutical industry, Drug discovery, Natural products, Biosynthetic pathways, Protein engineering.

Hands-on experience in the following areas:

- Molecular biology (molecular cloning, PCR, electrophoresis)
- Microbiology techniques
- Cell culture (plant callus and cell suspensions, yeast, insect, bacteria)
- Protein expression (*E. coli*, Baculovirus, *P. pastoris*)
- Metabolites extraction from natural sources
- Purification techniques for small compounds and proteins (IMAC, HPLC)
- Biophysical methods (DLS, CD, spectrofluorometry)
- Protein crystallization screenings
- Bioreactors
- Modelling softwares (Chimera, PyMol)

Languages

ITALIAN - native speaker proficiency
ENGLISH - full professional proficiency
DEUTSCH - limited proficiency
ΕΛΛΗΝΙΚΆ- elementary knowledge

Computer Skills

- Office Suite Microsoft (Word, Excel, Powerpoint, Outlook)
- Operating Systems (Windows, Mac OS)
- Cloud Computing (Dropbox, Google Drive, iCloud)
- Graphic (Adobe Photoshop)
- Molecular modelling (Chimera, ChemDraw, PyMol)
- Data processing (Origin)

References

References available upon request.