Recombinant Biomineralizing Polypeptides P_xS_y:

Structure and Dynamics upon the Interaction with Silica

Dissertation

zur Erlangung des Akademischen Grades
Doktor der Naturwissenschaften,

am Fachbereich 10, Biologie
der Johannes Gutenberg-Universität
Mainz

von

Christian Zerfaß

geb. am 7. März 1988 in Bad Kreuznach

Mainz
2015

Dekan: Prof. [redacted]
Erster Berichterstatter: [redacted]
Zweiter Berichterstatter: [redacted]
Dritter Berichterstatter: [redacted]
Tag des Prüfungskolloquiums: 4. August 2015

“Non quia difficilia sunt, non audemus, sed quia non audemus, difficilia sunt”

”It is not because things are difficult that we do not dare; it is because we do not dare that things are difficult.”

Lucius Annaeus Seneca, the Younger

Letter CIV, Section 26, Line 5

“Ad Lucilium epistulae morales Seneca; with an English translation by Richard M. Gummere.” *(Published 1917)*
(openlibrary.org; ID OL23268774M)

Table of Content

Declaration

This dissertation was written based on laboratory data which were produced between summer 2011 and spring 2015. The laboratory work was conducted in the working group of Professor [redacted], the principal advisor, within the Institute of General Botany, Faculty of Biology, of the Johannes Gutenberg University Mainz.

During the work on the PhD project, three Bachelor final theses were co-supervised by the author (the work of students [redacted], [redacted], [redacted]). In part, data from their theses (unpublished) were reproduced within this dissertation and labelled accordingly.

In addition, cooperation projects were conducted with the University of Crete (UoC, Greece, Professor [redacted], *Crystal Engineering, Growth & Design Lab*), the Pacific Northwestern National Laboratories (PNNL, Richland, WA, United States of America; Professor [redacted], PhD, *Physical Sciences Division*), within the Johannes Gutenberg University (working group of Professor [redacted], Mainz, [redacted], *Institute of Zoology*) as well as the University of Jena ([redacted], *Center for Electron Microscopy, University Hospital*). In cases data were derived from these cooperations, these are either marked in particular, or by an author list in case they resulted in a publication manuscript.

Summary

Diatoms deposit silica to form their mineralized cell walls and achieve structural control of the morphology of silica on the nano- to microscale. They employ organic compounds such as peptides, carbohydrate-based polymers and poly(alkylamines) to deliberately guide the formation of silica from soluble silicic acid, under chemically benign, i.e. physiological conditions. These features have raised interest in the implementation of biomineralization techniques in technical applications. Polypeptides in particular are interesting for controlling the silica-formation due to their variability in sequence and structure, hence being a versatile class of biomineralizing agents. However, polypeptides which facilitate the formation of silica usually bear high net cationic charges and are hence difficult to synthesize by recombinant expression in bacteria, whereas chemical syntheses fail for longer polypeptides because of incomplete coupling reactions. For rational design of biomineralizing peptides, it is furthermore necessary to understand their dynamics and structure during the peptide-guided silica formation.

In this work, designed polycationic polypeptides termed P_xS_y were produced in bacteria as fusions with detoxifying pre-sequences, and their impact on as well as behavior during silicic acid condensation and silica-dissolution was assessed. The amino acid composition of the P_xS_y polypeptides is similar to that of silaffins, which are polycationic, phosphorylated peptides used by diatoms to guide the formation of silica. Furthermore, P_xS_y can be phosphorylated *in vitro* by using a commercially available kinase.

The recombinant expression of P_xS_y was achieved by either fusing them to a bacterial export leader in combination with a charge-compensating polyanionic pro-sequence, or alternatively to a hydrophobic fusion which mediates the deposition in insoluble inclusion bodies. With both fusions, expression of P_xS_y was at highest level when the expression was induced at high cell densities (optical density at 600 nm above 2). Depending on the fusion used, different purification methods were established for removal of the fusions by chemical or proteolytic cleavage and purification of the P_xS_y targets. With the hydrophobic fusion, which supported high overexpression levels, the longest peptide from the P_xS_y-series, P_5S_3, yielded up to 15 µmol per liter bacterial culture. Expression of this polypeptide as charge-compensated fusion instead yielded only around 0.5 µmol per liter. P_xS_y preparations were fairly pure irrespective of which method was applied, so that the fusions provide two alternative approaches for the bacterial expression of the polypeptides.

The P_xS_y polypeptides adopted extended secondary structures with a considerable fraction of the polyproline II conformation, which is typical for polypeptides with high net cationic charge. They unexpectedly self-assembled in solution, even though electrostatic repulsion between the cationic side chains was expected. The self-assembly mechanism remains to be elucidated, but it is possible that interactions occur between charged side chains and the peptide backbone amides which are exposed in the extended conformation. This is a known interaction in the folding of

proteins, but also in the binding of peptide ligands, and hence can be presumed to be sufficiently strong to mediate polypeptide assembly.

The longest P_xS_y version P_5S_3 facilitated the formation of silica from supersaturated 30 mmol/L silicic acid solutions above pH 6.5 within 30 minutes. At neutral pH, silica nano-spheres were formed with diameters around 50 - 200 nm. P_5S_3 was occluded in the silica-matrix, even at high salt concentrations of 1 mol/L sodium chloride which sufficed to desorb P5S3 from pre-formed silica. The phosphorylated P_5S_3 also facilitated the formation of silica, though with slightly reduced yields.

At lower, but still supersaturated silicic acid concentrations of 8.3 mmol/L, P_5S_3 at low micromolar concentrations conversely inhibited the polycondensation of silicic acid in an 8 hour term. However, after this time a co-precipitate formed as well and the polypeptide guided the formation of silica nano-spheres. After dilution of pre-condensed silica to globally undersaturated silicic acid concentrations, P_5S_3 furthermore retarded the dissolution of silica.

The interaction with silica caused a decrease of the extended polyproline II conformation, as revealed by circular dichroism (CD) spectroscopy. This conformational change progressed concomitantly with the polycondensation of silicic acid in supersaturated solutions above pH 6.5. At pH 4.5, where silicic acid polycondensation is very slow, no changes in CD were observed, revealing that the conformational change only occurs in the presence of silica, but not silicic acid.

During dissolution of silica, two distinct phases of structure changes were observed for P_5S_3. In the first stage, which lasts around 150 minutes at pH 7.5, the CD changes indicated an increasingly intimate binding between P_5S_3 and silica. In the second stage, the interactions became successively weaker, ending up in the eventual release of the polypeptide after 450 minutes where silica was completely dissolved. Tracing the liberation of silicic acid revealed that the dissolution progresses during this complete period.

These findings in conclusion offer new opportunities for the design of recombinant, biomineralizing peptides which either facilitate, or mitigate the formation of silica from soluble silicic acid. Circular dichroism spectroscopy thereby offers a facile approach to probe polypeptides during the mineralization and dissolution reaction.

Zusammenfassung

Diatomeen bilden mineralische Zellhüllen aus Kieselsäure, wobei sie die Morphologie des abgelagerten Silikates auf der Nano- bis Mikroebene strukturell kontrollieren können. Sie nutzen diverse organische Verbindungen, wie Peptide, Kohlenhydrat-basierte Polymere oder poly(alkylamin)e, um die Bildung von Silikaten aus löslicher Kieselsäure zu steuern. Dadurch erzielen sie eine kontrollierte Silikat-Bildung unter chemisch milden, physiologischen Umgebungsbedingungen. Aufgrund der hohen strukturellen Kontrolle und der Silikat-Bildung unter milden Bedingungen ist die Biomineralisations-Strategie der Diatomeen von großem Interesse für Umsetzungen in technischen Applikationen. Insbesondere Polypeptide bieten breite Anwendungsmöglichkeiten bei der biomimetischen Erzeugung von nanostrukturierten Silikaten, da sie in ihrer Aminosäure-Sequenz variiert werden können und somit eine vielseitige Klasse von Biomineralisations-Agenzien darstellen. Die Herstellung Silikat-biomineralisierender Peptide und Polypeptide ist allerdings mit Schwierigkeiten verbunden. Da sie oft eine hohe kationische Netto-Ladung tragen, wirken sie in rekombinanten, bakteriellen Expressionen toxisch, während chemische Synthesen für die Produktion längerer Sequenzen aufgrund von unvollständigen Kopplungs-Reaktionen ungeeignet ist. Für die Entwicklung von neuen, biomineralisierenden Peptid-Sequenzen ist weiterhin ein tiefergehendes Verständnis der Peptid-Struktur und -Dynamik während der Peptid-gesteuerten Silikat-Bildung notwendig, um über die Sequenz hinaus die Konditionen für eine effiziente Interaktion zwischen den Peptiden und Silikat zu gewährleisten, wodurch in der Folge die Silikat-Bildung gesteuert werden soll.

In dieser Arbeit gelang die bakterielle Expression synthetischer Polypeptide mit hoher kationischer Nettoladung als Fusion mit Präsequenzen, die die schädlichen Effekte der Polypeptide auf den bakteriellen Expressionswirt lindern. Die Polypeptide wurden mit der Bezeichnung P_xS_y benannt und waren in ihrer Aminosäure-Zusammensetzung ähnlich den Silaffinen, polykationischen und phosphorylierten Peptiden welche Diatomeen zur Silikatbildung einsetzen. Weiterhin konnten die P_xS_y Polypeptide *in vitro* enzymatisch durch Einsatz einer Kinase phosphoryliert werden. Die rekombinante Expression von P_xS_y gelang sowohl durch Fusion mit einer bakteriellen Leitsequenz für den periplasmatischen Export in Kombination mit einer Ladungs-kompensierenden anionischen Prosequenz, als auch in Fusion mit einer hydrophoben Sequenz welche die Ablagerung in unlöslichen Einschlusskörpern („inclusion bodies“) innerhalb des bakteriellen Wirtes bewirkt. Mit beiden Fusionen wurden die besten Expressionsraten erzielt, wenn die Expression des Zielproteins bei hohen Zelldichten (optische Dichte der Kulturen bei 600 nm größer 2) induziert wurde. Abhängig von der verwendeten Fusion wurden unterschiedliche Reinigungsprozeduren etabliert, innerhalb derer auch die Fusionen per proteolytischer oder chemischer Spaltung entfernt wurden. Mit der hydrophoben Fusion, die hohe Expressionsraten ermöglichte, wurden bis zu 15 µmol pro Liter Expressionskultur des Polypeptids P_5S_3, die längste Version der P_xS_y-Serie, gewonnen. Demgegenüber befand sich die Ausbeute nach Expression als Ladungs-kompensierte Fusion im Bereich von 0,5 µmol pro Liter Expressionskultur. Mit beiden Fusionsstrategien wurden indes

recht reine P_xS_y Präparationen erzielt, sodass sie als zueinander komplementäre Methoden für die rekombinante Expression von Polypeptiden mit hoher kationischer Nettoladung dienen können.

In Lösung nehmen die P_xS_y Polypeptide eine gestreckte Sekundärstruktur ein, wobei insbesondere das Polyprolin II Strukturmotif zugeordnet werden konnte. Diese strukturellen Eigenschaften sind typisch für Polypeptide mit hoher Nettoladung. Unerwarteterweise war P_xS_y in der Lage, in Lösung zu assemblieren, obgleich eine elektrostatische Abstoßung der kationisch geladenen Seitenketten zu erwarten gewesen wäre. Der Mechanismus der Assemblierung verbleibt im Detail ungeklärt. Es ist indes möglich, dass geladene Seitenketten mit den Amiden der Peptid-Bindung, welche in gestreckten Konformationen dem Lösungsmittel exponiert sind, interagierten und dadurch die Peptid-Ketten miteinander banden. Diese Interaktion ist bekannt innerhalb gefalteter Proteine und dient auch der Bindung von Liganden, sodass davon auszugehen ist, dass sie insgesamt ausreichend stark ist, eine Assemblierung zu vermitteln.

Die längste P_xS_y-Version P_5S_3 unterstützte die Silikat-Bildung oberhalb von pH 6,5 in übersättigten 30 mmol/L Kieselsäure-Lösungen innerhalb einer 30-minütigen Reaktion. In neutraler pH-Umgebung bildeten sich in Anwesenheit von P_5S_3 Silikat nano-Sphären mit Durchmessern zwischen 50 und 200 nm. Das Polypeptid wurde in die Silikat-Matrix eingeschlossen, und dies sogar bei hohen Salzkonzentrationen von 1 mol/L Natriumchlorid, welche indes ausreichten, um P_5S_3 von präformiertem Silikat abzulösen. Auch das phosphorylierte P_5S_3 Polypeptid unterstützte die Silikat-Bildung, jedoch war die Menge ausgefällten Silikates herabgesetzt.

Bei erniedrigter, obgleich nach wie vor übersättigter Kieselsäure-Konzentration von 8,3 mmol/L verursachte P_5S_3 in niedriger mikromolarer Konzentration derweil eine Verlangsamung der Kieselsäure-Kondensation innerhalb eines acht-stündigen Zeitraumes. Nichtsdestotrotz bildete sich am Ende dieser Zeit ein Präzipitat aus Silikat und P_5S_3, in dem das Polypeptid wiederum eingeschlossen war und die Ausbildung einer nano-sphärischen Strukturierung hervorrief.

Nach Verdünnung präformierten Silikates zu global ungesättigter Kieselsäure-Konzentration verlangsamte P_5S_3 zudem die Auflösung der Silikate.

Während der Interaktion mit Silikat verlor P_5S_3 zum Teil seine gestreckte Polyprolin II Konformation, was in veränderten Zirkulardichroismus (CD)-Spektren ersichtlich wurde. Die Konformationsänderung trat parallel mit der Polykondensation von Kieselsäure in übersättigten Lösungen oberhalb von pH 6,5 auf. Demgegenüber wurde keine Konformationsänderung bei pH 4,5 beobachtet. Bei diesem pH ist die Kieselsäure-Kondensation sehr langsam. Folglich ist hier nahezu ausschließlich Mono-Kieselsäure im Reaktionsansatz zu erwarten, welche offenbar keine strukturelle Änderung in P_5S_3 hervorrief.

Währen der Auflösung von Silikaten in Anwesenheit von P_5S_3 traten zwei distinkte Phasen mit unterschiedlichen Strukturänderungen in P_5S_3 auf. In einer ersten Phase, die bei pH 7,5 für 150 Minuten andauerte, deuteten die Änderungen im CD auf eine sukzessive stärker werdende

Interaktion des Polypeptides mit den Silikaten hin. In der zweiten Phase wurde diese wieder zunehmend schwächer bis hin zur vollständigen Auflösung des Silikates und damit einhergehend der Freisetzung des Polypeptides nach ungefähr 450 Minuten. Die Freisetzung von Kieselsäure erfolgte während der kompletten Reaktionsphase, wie aus einer parallelen Quantifizierung der gelösten Kieselsäure hervorging.

Diese Resultate eröffnen neue Möglichkeiten für die Gestaltung rekombinanter, biomineralisierender Polypeptide welche entweder die Bildung von Silikat aus löslicher Kieselsäure erleichtern, oder vermindern. Die Untersuchung der Polypeptid-Silikat Interaktion per Zirkulardichroismus-Spektroskopie bietet hierbei einen einfachen Zugang zur Untersuchung des Polypeptid-Verhaltens während Mineralisations- und Auflösungs-Reaktionen.

Introduction - Biomineralization: Applying the strategies of diatoms for the *in vitro* production of nano-structured silica

Mineralized tissues play important roles in many organisms. They are often used to gain stability (skeletons), may serve to perceive the earth's magnetic field (magnetotactic bacteria), and protect against predators (particularly exoskeletons) [1]. The process of biological deposition of minerals, which is referred to as biomineralization, is commonly controlled by the formation of organic scaffolds as sites for the specific deposition of a particular mineral [2]. After the mineralization of a tissue, the organic matrix is either retained in the mineral phase (e.g. nacre [3]), or removed (e.g. enamel [4]), factors which tune the hardness and fracture resistance of the resulting materials.

Among biominerals formed by organisms, one of the most prominent is silica (SiO_2), the polymer of silicic acid ($Si(OH)_4$). Silicic acid is very abundant in the earth's crust [5] and constantly delivered to surface waters by weathering processes [6]. It is used by several organisms to build silicified exoskeletons, such as diatoms, sponges or radiolaria [1]. These organisms are capable to deliberately form silica with precise control on different length scales. While sponges commonly form needle-like scaffold structures (spiculae) [7], diatoms usually assemble their epi- and hypotheca starting from nano-particulate silica [8-10]. In higher plants, which take up silicic acid with the ground water, cell-wall stabilizing sheet-, fiber- and particulate deposits are found [11]. This morphological control is promising for the use of biomineralization techniques to achieve structural specificity in the production of advanced nano-materials, by applying biomimetic strategies which work under benign, often physiological conditions [12-15]. The *in vitro* biomimetic silica formation can be accomplished for example by employing enzymes isolated from sponges, for instance silicatein [7,16,17], or peptides from diatoms which are often highly post-translationally modified in their native state [18-21]. Because of the versatility of peptides and polypeptides in sequence and structure, and the access by synthetic [22,23] and recombinant [24-27] production routes, the diatom's strategy of silica formation in particular is promising in technical applications for biomimetic production of nanomaterials with specific properties [13,12].

Silicic acid transport and deposition in diatoms

Diatoms [28], unicellular algae with silicified cell walls, are abundant in the world's aquatic habitats and contribute to a considerable fraction to the photosynthetic carbon fixation in the oceans [29]. They build a variety of silicified shells under physiological conditions, in a hierarchical manner, starting with nanoscale silica spheres which are further assembled into specific meso- and microscale structures [10,30]. Both in fresh- and saltwater, diatoms face a low

silicic acid concentration (on average estimated to 150 µmol/L and 70 µmol/L in rivers and oceans, respectively [6,31]), which is below the autopolycondensation threshold of silicic acid (1 - 2 mmol/L [32,33]). During production of their cell wall, diatoms efficiently accumulate silicic acid as a soluble silicon pool, and achieve intracellular concentrations of up to 100 mmol/L, the value varying among different species [34]. The enrichment of silica is presumably achieved by silicon transport proteins (SIT) [35-37], transmembrane proteins which serve as symporters for silicic acid and sodium ions [38].

In *Thallasiosira pseudonana*, Si-NMR-studies revealed that silicic acid within the cell is rapidly condensed and virtually completely stored as polymeric, but soluble silica species [39,40]. Silica-nanostructures are then delivered to the silica deposition vesicle (SDV) [10,41,42]. The intracellular transport of silica particles is presumably achieved by silicon transport vesicles (STV) [9,43]. In the SDV itself, silica structures are assembled under acidic conditions [44], before silica is released by fusion of the SDV with the plasma membrane [42]. The silicified cell wall contains, besides silica, proteins, polyamines and polysaccharides [45,46] and is additionally protected by a mucilaginous coating [8,47].

Organic compounds from diatoms which facilitate the formation of silica

Within the silicified cell wall of diatoms, organic compounds are tightly associated with the silica matrix. These compounds in majority belong to three substance categories: peptides, acidic sugar-based compounds as well as poly(alkylamine)s (particularly poly(propylamine)s) [18-21,48-53]. Considering the acidic nature of silica, polycationic compounds are of particular importance, since these can electrostatically interact with the inorganic matrix [18,52].

In *Cylindrotheca fusiformis*, the most tightly associated peptide of the cell wall was termed silaffin [18,50], acknowledging the high silica affinity. *cf*-Silaffins are polycationic peptides rich in the amino acids lysine and serine. The lysines are mostly post-translationally modified by hydroxylations, side-chain amine methylations, and polycationic poly(propylamine) tails. Serines are virtually fully phosphorylated. The silaffins induce the *in vitro* formation of silica from supersaturated solutions of silicic acid above pH 4 and with a maximum at pH 5 [18] which suits well with the acidic conditions expected for the silica deposition vesicle [44]. To acquire silica formation activity, they depend on the presence of phosphorylations, or alternatively polyvalent anions in the medium to mediate their self-assembly [54]. Depending on their particular composition, silaffins guide the formation of silica with different morphologies. Pure preparations of silaffin-1A, a 15 amino acid peptide bearing the above mentioned post-translational modifications [18,50], mediated the formation of defined silica-spheres 500 - 700 nm in diameter, whereas a mixture of native silaffins, which differ in peptide length and modification patterns, induced the formation of 50 nm particles, respectively [18]. The unmodified peptide moiety of silaffin-1A itself, referred to as R5, guided the formation of silica-particles with diameters around 150 - 300 nm [55], but was only active at neutral pH and above

[18], in contrast to the native silaffins which retain activity at lower pH. Analysis of R5-derivatives with scrambled sequences revealed that the primary sequence of the peptide has a strong influence on the morphology of silica formed, but not on the specific activity of the peptides [56].

Silaffin peptides occur as well in the diatom *Thalassiosira pseudonana* [19]. As their counterparts in *Cylindrotheca fusiformis, tp*-silaffins are rich in lysine and serine. However, modifications are far more variable. Beside phosphorylations and polyamine tails, sulfonations and glycosylations are known. Alternative processing leads to an ensemble of different silaffin peptides from one precursor protein, the peptides differing in length and hence molecular weight. Interestingly, although bearing numerous lysines and polyamine tails, some *tp*-silaffins become hyperphosphorylated and do not support silica formation individually in their native state, though assist in the formation of silica induced by polyamines in mediating their self-assembly. However, the lysines in *tp*-silaffins play a role in the intracellular transport of silaffins. In *Thalassiosira pseudonana*, analysis of silaffin mutants revealed that five non-consecutive lysines within stretches of 12 - 14 amino acids promote the targeting to the biosilica formation sites, eventually leading to the incorporation of these peptides into silica [57].

Recently, silaffins have been predicted computationally from the genome of *Fistulifera sp.* [58], but have not been isolated for *in vitro* analyzes yet.

Cingulins, microring forming proteins found in *Thalassiosira pseudonana* and *Coscinodiscus wailesii,* share the high content in lysine and serine with silaffins [21]. Fusion with green fluorescent protein (GFP) revealed that they are associated with the girdle band region of *Thalassiosira pseudonana*, most likely templating the specific girdle band structure. The *in vitro* mineralization was accomplished in supersaturated solution of silicic acid at slightly acidic pH, conditions expected for the *in vivo* mineralization within the silicon deposition vesicles [44].

Besides proteins and peptides with numerous amines, poly(alkylamine)s are abundant in the silica cell walls of diatoms [52,59]. Polyamines induce silica formation from silicic acid *in vitro* in the presence of polyvalent anions mediating their self-assembly [60-62] and are furthermore flocculants of colloidal silica [63]. In diatoms, poly(propylamine)s are prominent [64-66]. These are usually attached to putrescine (1,4-diaminobutane), spermidine (N-(3-aminopropyl)butane-1,4-diamine) or a silaffin peptide. Nitrogen atoms in the polyamines are often additionally methylated, rendering the amines more likely to be protonated (tertiary amines) or introducing permanent charges (quartenary amines). In part, polyamines may be sulfonated. Among different diatom species, polyamines differ in the degree of propylamine-polymerization and methylation pattern of amines [59]. An increased methylation pattern leads to an increased yield in polyamine-induced *in vitro* silica formation [67]. The morphology of silica formed in the presence of polyamines is commonly spherical, and sizes of the spheres differ depending on the degree of polymerization of the polyamines [52] and the pH of the environment [67].

As noted, silaffins and polyamines require counter-charged additives, e.g. phosphate, for electrostatically mediated self-assembly, a prerequisite for their *in vitro* silica precipitation activity [54,62,68]. Diatoms use a variety of anionically charged counterions for silaffin- or polyamine-directed silica formation, with different morphological outcomes for the resulting silica. Both in *Cylindrotheca fusiformis*, and in *Thalassiosira pseudonana*, polyanionic polypeptides are associated with the silica cell wall [20,53]. In *Cylindrotheca fusiformis* [20], silaffin-2 is a polypeptide rich in hydroxyl-group bearing amino acids serine, threonine and hydroxyproline, but also contains hydroxylated and methylated lysines. Though the peptide sequence is cationic, the native natSil-2 is phosphorylated, sulfonated and glycosylated, rendering it overall anionically charged. It does not induce the formation of silica itself, but instead acts as an additive for activation of polyamines *in vitro*. However, this is only the case at low concentrations of natSil-2. At increased concentration, it even hindered the polyamine-induced silica formation, and did so as well when added to native silaffin-1A, the latter being an active inducer of silica formation individually. In *Thalassiosira pseudonana*, acidic, hyperphosphorylated polypeptides termed silacidins occur [53] which are increasingly expressed under silicon starvation conditions [69]. The polypeptide chain is rich in serine and acidic glutamate, and, as natSil-2, exerts a supporting effect on polyamine-induced silica formation in its native state [53]. However, when dephosphorylated, the silacidins lose their ability to influence the silica formation induced by poly(allylamine) [69], demonstrating the necessity of phosphorylations. Compared to orthophosphate, phosphorylated silacidins are active at far lower concentrations.

Beyond acidic polypeptides, phosphorylated sugar compounds were isolated from the cell walls of *Stephanopyxis turris* [70]. Mannose-6-phosphate was found to be tightly embedded within the silica matrix and mediated the polyamine-induced silica precipitation *in vitro*. Comparison with orthophosphate revealed a structure-guiding effect of the phosphorylated compound, not exerted by the unphosphorylated carbohydrate.

Further polypeptides [71] were isolated from diatoms which were associated with the silica cell walls, though did not exert *in vitro* silica precipitation activity. Among these were frustulins [48,72] and pleuralins [49,73]. The function of these polypeptides is not yet elucidated, but they may be involved in the control of cell wall formation, or increase the cell wall integrity [71].

Biomimetic in vitro silica formation guided by polypeptides

The peptide part of silaffin-1A, usually referred to as R5, has been subject of elaborate investigations on the mineralizing action of peptide compounds. In detail, R5 is a sequence repeat of the silaffin precursor gene *Sil1* [18]. It contains predominantly lysine and serine (as the native silaffin-1A), but in addition bears a carboxy-terminal RRIL motif which during maturation of silaffins becomes removed *in vivo*. R5 induces the formation of silica from supersaturated silicic acid solutions at neutral pH and above [18,74], whereas native silaffin-1A (bearing polyamine

tails and lysine methylations) is still active in more acidic environments [18,50], as long as phosphate is present in solution or as phosphorylation on the silaffin [18,54]. Generally, the R5 activity towards silica formation and flocculation is attributed to its polycationic charge [18]. R5 is believed to form micellar assemblies in the presence of phosphate [55], and silica is deposited on the micelle surface, mediated by cationic amino acid side-chains exposed to the bulk solution [75]. However, the extent of R5 self-assembly is debated [76].

Although the double-arginine motif is removed in native silaffins, it strongly enhances the *in vitro* silica formation activity of R5 with respect to derivatives without this extension, but it is not a prerequisite [55]. Furthermore, its location within the peptide sequence plays only a minor role for the activity, although it has an impact on the resulting silica morphology [55,56]. Generally, silica formed under the influence of R5 is spherical, and changes in the order of the amino acids can tune the size and the texture of the silica particles [56]. However, environmental conditions were found which strongly affect the outcome of R5-induced silica precipitation [77]. Under static conditions, spherical silica was produced. By contrast, bubbling in nitrogen gas led to bent silica chains, and shear stress allowed the deposition of fibrillary silica.

The impact of native-like modifications of silaffins was also studied on R5 in biomimetic environments [78]. Interestingly, when examined in phosphate buffer (i.e. under conditions where R5 is active), phosphorylation had a detrimental effect on the amount of silica precipitated. Increasing the number of phosphoryl groups successively reduced the capacity of silica-precipitation, indicating that electrostatic repulsion from the silica-surface occurred. To mimic the polyamine tails found in native silaffin-1A [18], spermidine, an oligo-alkylamine was attached to lysine side-chains and compared with lysine-methylations which introduced permanent charges [78]. Both modifications unexpectedly reduced the silica precipitation activity at neutral pH, but had an effect on silica morphology.

Since lysine-clusters are prominent in R5 and silaffin-1A [18] as well as *tp*-silaffins [19], short peptides with sequence GK*AXKG (G: glycine, K: lysine, A: alanine, X: alanine, serine or glutamate), in which the first lysine (*asterisk*) was modified with an oligoamine (w,w-dimethyl-dipropyleneaminolysine), were studied and found to suffice to initiate the formation of silica in phosphate buffer [79]. Depending on the central amino acid, resulting silica was spherical (alanine), grainy (glutamate) or an ensemble of both (serine), and total precipitation activity was highest in the last version. To further reduce the complexity of the peptide, the biomineralizing activity of individual amino acids and their homopolymers was evaluated. Both poly-L-lysine, and poly-L-arginine induced the formation of silica from silicic acid solutions and mediated the rapid deposition of suspended silica [80]. Additionally, they accelerated silicic acid polycondensation, which their respective amino acid monomers did not. Poly-L-serine accelerated the condensation of silicic acid, but did not eventually precipitate with silica. By contrast, poly-L-proline precipitated immediately with silica when added to an orthosilicic acid solution, which was primarily because of the low solubility of the polypeptide, but silicic acid was clearly retrieved from the precipitate. The interaction between poly-L-proline and silica was

further supported by analyzes using proline-rich minicollagens (oligomers of the sequence motif proline-hydroxyproline-glycine). These collagens were soluble in water and formed co-precipitates with silica when silicic acid was offered [81], demonstrating that the interaction between silica and uncharged polypeptides is sufficient for the latter to guide the silica formation.

The poly-L-lysine (PLL) induced silica formation has been extensively studied because of the abundance of lysine in silaffins [18] and its *in vitro* capability to accelerate silicic acid condensation and silica flocculation [60,80]. Since PLL gets entrapped into the silica matrix during polypeptide-induced precipitation, it was successfully used as template and to control the pore structure of silica [82,83]. Depending on the degree of polymerization and buffer environment, silica with a variety of morphologies and textures can be achieved [74,84]. PLL can be induced to self-assemble with phosphate prior to silicification, which results in the production of hollow silica spheres [85]. However, at sufficient degree of PLL polymerization, phosphate induces the assembly of α-helical PLL into hexagonal single-crystalline platelets [86-88] which in turn template the formation of hexagonal silica platelets [86]. Shorter versions of PLL fail to assemble into platelets and instead mediate the formation of spherical silica particles [82,85,86,88-90].

Instead of merely taking advantage of the intrinsic capability of PLL to assemble in the presence of polyvalent anions such as phosphate, block-copolypeptides with lysine and leucine have been employed to generate a variety of silica structures such as spheres, individual platelets or fused platelets, fine-tuned by the presence of different counter-ions as phosphate, sulfate or carbonate [91]. Alternating LK-peptides (L: leucine, K: lysine) were designed which deliberately adopt different secondary structures, depending on their periodicity. Depending on their conformation, these peptides mediated the formation of silica spheres, elongated rods or wires [92]. The importance of peptide secondary structures for templating defined silica-morphologies was further supported by comparison of repetitive LKLL (α-helices) and VKVV (β-sheets) seequences (V: valine), fused to polyethylene glycol to increase their solubility. These polypeptides guided the formation of silica with paper-like or tube-like morphologies, respectively [93]. Short amphiphilic peptides A_6K and V_6K (A: alanine) self-assembled into nanotubes and lamellar stacks, respectively, and these structures guided the formation of flower-like or lamellar layered silica, their dimensions affected by the presence of different counter-ions (phosphate, sulfate, carbonate) [94,95]. By varying the length of alanine stretches in A_mK peptides, tailored formation of either spherical or rod-shaped silica was achieved [96]. Similarly, by tuning the hydrophobic domains of the related peptides I_mK and L_mK (I: isoleucin, m = 3 - 5), the self-assembly properties were tightly controlled [96-98], and the peptides were capable to template the formation of twisted silica nanotubes or spheres, respectively [96,99,100]. If desired, polypetides can be removed after templating the silica formation by e.g. calcination or, if accessible, by enzymatic digestion, which was succesfully applied in the production of pure silica nanotubes templated by self-assembling α-helical peptides [101].

Further insight into the principles of peptide-induced silica formation was derived from investigations of the peptide KSL [102]. The peptide (sequence KKVVFKVKFK, K: lysine, V: valine, F: phenylalanine) is an antimicrobial peptide that is able to induce silica formation at near-neutral pH. KSL was predicted by computational modelling to self-assemble in the presence of phosphate [103], a behavior similar to the silaffin self-assembly [54]. Analyses of KSL within the silica-matrix, after peptide induced silica formation, revealed a remarkable interaction between the peptide bond amides and the silica matrix, whereas contacts between charged lysine side-chains and silica were negligible [104,105]. It was hence concluded that the peptide bonds play an important role in the acceleration of silica formation guided by peptides.

A comparison of the pH-dependence of R5 with PRP1, a proline-rich cucumber-peptide with silica-biomineralizing activity, revealed that an increased activity at lower pH can be achieved by increasing the peptide charge density [106]. The PRP1 sequence contains 7 lysines and 2 arginines in 22 amino acids, whereas R5 contains 4 lysine and 2 arginine in 19 amino acids. PRP1 induced the formation of silica above pH 6, with highest yields around pH 6.5, both with its native primary structure and as a scrambled derivative. In comparison, R5 was only active above pH 6.5 with highest silica yields above pH 7.0.

The peptide TBP-1, which was identified as titania binding peptide (TBP) from a phage-display screening, is active in mineralization of a variety of different materials, including silica [107]. Activity in silicification was found both in Tris, and phosphate-buffer, pointing out that no electrostatic assembly mediated by multivalent anions is required. Mutating the peptide by single amino acid exchanges additionally revealed that an arginine, proline and aspartate were necessary for gaining mineralization activity. While the first one is anticipated because of the cationic charge (hence facilitating interaction with acidic silica), the other two amino acids may point out the importance of polypeptide conformations and hydrogen bonds with silica, respectively. Hydrophobic amino acids are presumed from computational modelling to be as well involved in the interaction of peptides with quartz surfaces [108], and the binding strength is increased for peptides with extended conformation in solution, allowing a high degree of flexibility for adaptation to a surface upon binding [109].

For the *in vitro* analysis of biomineralizing peptides, native isolates (bearing posttranslational modificiations) are usually compared to synthetic and recombinant peptide-versions (e.g. in the analysis of *cf*-silaffins [18,110]). The chemical synthesis by solid phase peptide synthesis [22,23,111] provides large amounts of peptides and derivatives with altered primary structures which facilitates extensive analysis of their biomineralization activity. Additionally, the introduction of side-chain modifications is possible by deliberate usage of pre-modified amino acids (e.g. ref's [56,78,79]). For the production of longer polypeptides, this approach yields successively reduced amounts of peptide because of incomplete coupling reactions [22], so that recombinant expression in bacteria is favorable [24-26]. However, the high cationic net charge of silica-biomineralizing peptides as *cf*-silaffin may exert detrimental effects on the expression rate, e.g. because of undesired and harmful interactions with anionic cellular compounds as DNA or

biomembranes [112,113], leading only to polypeptide yields in the low milligram range per liter of bacterial culture [110]. Strategies to overcome this issue were evaluated for antimicrobial peptides [114,115], peptides which most often are cationic and amphiphilic and perforate bacterial membranes. In particular, protein fusions were identified which detoxify the antimicrobial peptides, facilitate high-level over-expression and can be cleaved of afterwards [116,117]. A detailed overview on the different fusion proteins investigated so far is given in the introductions of sections I and III of this thesis.

Biomimetic in vitro silica formation guided by poly(alkylamine)s

The identification of long chain polyamines in diatoms [52] raised interest in the utilization of polycationic polymeric compounds in biomimetic silica formation procedures. Generally, polyamines are thought to accelerate the polycondensation of silicic acid by an interplay of charged and uncharged nitrogens, resulting in an acid/base-catalysis by supporting de-protonation of silicic acid and protonation of the leaving group (-OH) during condensation [60,61]. Additionally, they serve as flocculating agents for as formed silica [63]. When added to supersaturated solutions of silicic acid, polyamines and silica form stabilized sols, though when phosphate is present to mediate the assembly of polyamines this results in rapid precipitation of silica [118]. This was demonstrated with poly(vinylamine) which led to either the precipitation of a silica-polymer composite, or a stabilization of silica nanoparticles, depending on its degree of vinylamine polymerization and hence its capability to form networks with silica [119]. The driving force for the precipitation process is a self-assembly of the polyamine, which can undergo phase-separation in the presence of polyvalent anions like phosphate, which more efficiently mediates electrostatic bridging than (acidic) silicates alone [62,120-123]. The effectiveness of polyamines to precipitate silica depends on the charge state (e.g. methylation of amines enhances this capability) [124,125], which in turn is affected by the choice of alkyl-spacers separating the amines e.g. in poly(alkylamines) [67,123,125]. Activity was found for different short chain (diamines to hexamines) [61,126] and long chain polyamines (including primary and secondary amines), but also for cationic sulphur-containing polymers, the resulting silica differing in its morphology and texture [123,127].

Besides the actual composition of the polymer, the interplay with phosphate to induce phase-separation was used to control the silica morphology [62,123]. Usually, silica produced in the polyamine-phosphate phase-separated system appeares as hollow silica spheres [128]. However, poly(ethyleneimine) in the presence of different organic polyvalent acids (e.g. poly(ethylene glycol)bis(carboxymethyl)ethers or tetra(sulfophenyl)porphyrin) instead templated fiber- and flower-like silica structures [129].

Organic compounds that stabilize supersaturated silicic acid solutions

As noted above, polycationic poly(vinylamine) stabilizes colloidal sols of silica-nanoparticles at pH 10 in a soluble form, but accelerates the condensation of silicic acid [119]. Poly(vinylimidazole) however retarded silicic acid condensation below pH 9, though it is as well a polycationic polymer [130].

Cationically charged poly(allylamine hydrochloride) (PAH) [131], poly(aminoamide) dendrimers (PAMAM) [132-134] and poly(ethyleneimine) (PEI) [131] at low dosages retarded the polycondensation of silicic acid when the latter was present at a supersaturated concentration around 8 mmol/L. At near-equimolar concentration of nitrogen-groups per silicon, they conversely accelerated the polycondensation of silicic acid [135]. Polycationic chitosane, a polymer based on (D)-glucosamine which was examined in a partially phosphorylated state, also stabilized silicic acid at low concentrations [136,137]. Combining the polycationic agents with anionic polymers as carboxymethylinulin increased their effectiveness at low dosages of the polyanion, though the effect was detrimental at higher concentrations [132,134,136,137]. Poly(vinylpyrrolidone), an uncharged polymer bearing amide groups, stabilized silicic acid solutions by retarding the condensation of silicic acid at neutral pH, proven by nuclear magnetic resonance spectroscopy of the silica condensation state [135]. The underlying mechanism was proposed to be a binding of silicic acid monomers and dimers, shielding them from further polymerization. A similar stabilizing effect was found for poly(ethyleneglycol)s (PEG), which exerted increased effectiveness concomitant with an elevated degree of polymerization [138]. However, after extended time periods, silica-precipitates were usually found in the presence of all stabilizing agents, the polymer exerting a structure guiding effect on the morphologies of the formed silica [134,135,137]. The stabilizing effect was not limited to amine polymers, since phosphonium-grafted PEG as well impeded the condensation of silicic acid, eventually co-precipitating as silica structure-directing additive [139].

In the opposite direction, it is interesting to note that polyamines (poly-L-lysine, poly(allylamine hydrochloride), poly(ethyleneamine) and oligomeric alkylamines) accelerated the de-condensation of silica after dilution to globally undersaturated concentrations of silcic acid [140].

Application of biomimetic silica formation techniques for enzyme entrapment

The biomimetic formation and precipitation of silica, as described in the previous sections, is usually achieved under physiological, i.e. chemically benign, conditions. This offers the possibility to include biomaterials as enzymes into the mineralization assays to immobilize these proteins, making them easier to handle and potentially stabilize them by confinement. This was achieved with a variety of proteins, from which some examples are presented in the following. Glucose oxidase was encapsulated in silica by layer-by-layer deposition on silica nanoparticles [141]. As biosilicifying agent, protamine, a polycationic (arginine-rich) peptide, was either used

individually or after covalent cross-linking to the enzyme. The oxidase retained activity in the matrix and had increased resistance against thermally induced denaturation. The increased thermostability of proteins in silica was also demonstrated with a light-harvesting chlorophyll a/b complex (LHC II) [142], a membrane protein which carries several non-covalently attached pigments (chlorophylls, xanthophylls). Silica formation was induced by spermine in the presence of recombinant LHC II versions with either the native primary sequence, or as chimera with an amino-terminal insertion of the silaffin-peptide R5, leading to an entrapment of LHC II in both cases. Since in the properly folded LHC II all excitation energy from chlorophyll *b* is transferred to chlorophyll *a*, an increase in chlorophyll *b* fluorescence served as indicator for denaturation. It was shown that LHC II is better stabilized against heat denaturation in silica-gels composed of smaller particles (20 nm compared to 200 nm) because of a higher level of confinement. Carboanhydrase was also entrapped in silica either by co-precipitation, induced by spermine [143], or by autoencapsulation of a chimera with the R5-peptide [144]. Both ways of encapsulation increased the thermostability and the immobilization allowed the usage of the enzyme for CO_2-sequestration in subsequent reaction cycles, with buffer exchange in between. However, the autoencapsulated R5-chimera was significantly more thermostable, showed a better catalytic performance at increased temperatures, and less leakage from the co-precipitate with time was observed. By entrapping NADPH-dependent nitrobenzene nitroreductase (NBNR) together with glucose-6-phosphate dehydrogenase (G6PDH) in silica precipitated by poly(ethyleneimine), a multienzyme system was created, the first enzyme reducing nitrobenzene to hydroxylaminobenzene under consumption of NADPH (electron donor), the latter regenerating NADPH by oxidizing glucose-6-phosphate to 6-phosphoglucono-δ-lactone [145].

A flow-through reactor for conversion of hydroxylaminobenzene (HAB) to *ortho*-aminophenol was built by immobilizing HAB mutase in silica [146]. Beyond, silica-immobilized glucose oxidase was implemented in a biosensor for glucose detection [147]. Further reports on the successful encapsulation of enzymes have been published, principally pointing out the thermostabilizing and immobilizing advantages for biotechnological applications [148-165].

Besides biomimetic silica formation in the presence of proteins, diatoms can be genetically modified to entrap proteins in their silica cell wall matrix. A methodology for "living diatom silica immobilization" (LiDSI) was established by fusing the silaffin tpSil3 to different proteins and express the constructs in *Thalassiosira pseudonana* [166,167]. This allowed the incorporation of a hydroxylaminobenzene mutase, β-Glucoronidase, glucose oxidase, galactose oxidase, and horseradish peroxidase into the biosilica cell wall, demonstrating the versatility of this strategy.

Aim of this investigation

Biomimetic mineralization is a promising approach for the production of nanomaterials under benign conditions. Polypeptides are of particular interest because of their high variability in the

amino acid sequence, but the knowledge about the interaction with mineral surfaces and the dynamics during mineralization reactions is limited. The objective of this work is the analysis of a series of designed polypeptides with high cationic net charge, termed P_xS_y, in terms of their solution behavior and mineral interaction. These polypeptides have a similar amino acid composition as silaffins, silica-biomineralizing peptides from diatoms, and shall serve as model peptides to understand the impact of peptides on the biomimetic mineralization of silica. The P_xS_y polypeptides are repetitive and contain different numbers of silaffin-like partial sequences ("S_y", y = 0 - 3). They are equipped with kinase target sites ("P_x", x = 2 - 5) so that they can be phosphorylated *in vitro*. Lysine, arginine and serine are the most abundant residues in the polypeptide sequences. Because of these features, P_xS_y polypeptides may represent an *in vitro* model for silaffins, which are polycationic and phosphorylated in their native state.

For the synthesis of the high-charged P_xS_y polypeptides, recombinant expression in *Escherichia coli* will be evaluated. Fusions shall be examined which are known to facilitate the overexpression of toxic polypeptides and protect them from degradation within the bacterial host. As one strategy, a polyanionic pre-sequence will be fused to the cationic targets, in order to serve as an intramolecular binding partner, and the interaction shall prevent adverse interactions of the polypeptides with cellular compounds. Alternatively, a fusion will be employed which mediates the rapid deposition of the polypeptides in inclusion bodies, which shall as well prevent disfavorable interactions with cellular compounds. For removal of the fusion, both proteolytic and chemical cleavage methods will be applied.

The silica formation will be studied in the presence of unmodified, and phosphorylated P_xS_y to investigate their impact on silicic acid solubility as well as silica morphology and stability. The effect of P_xS_y on silicic acid condensation and silica dissolution will be examined by colorimetric quantification of silicic acid in solution, whereas the effects on silica morphology will be analyzed by electron microscopy. Furthermore, the polypeptide localization after silica formation will be examined to test whether it co-precipitates with silica, or exerts an effect on the formation of silica from solution.

Of particular interest are the structure and dynamics of P_xS_y before, during and after silica-formation, as well as during silica dissolution. By using circular dichroism the structure of P_xS_y under different solution conditions will be investigated. Since self-assembly is a known pre-requisite for silaffins to gain biomineralizing activity, the self-assembly properties of P_xS_y will be analyzed by dynamic light scattering and electron microscopy. A circular dichroism spectroscopic approach shall be established to monitor changes in the polypeptide conformation during the association with silica, in order to achieve insight into the interaction under different conditions and trace the formation and dissolution of silica by probing the behavior of the polypeptides.

References

[1] H. Lowenstam, Science 211 (1981) 1126.

[2] E. Bonucci, J. Bone Miner. Metab. 27 (2009) 255.

[3] J. Sun, B. Bhushan, RSC Adv. 2 (2012) 7617.

[4] J. Moradian-Oldak, Front. Biosci. 17 (2012) 1996.

[5] K.H. Wedepohl, Geochim. Cosmochim. Acta 59 (1995) 1217.

[6] R. Siever, Am. Mineral. 42 (1957) 821.

[7] Müller, Werner E G, H.C. Schröder, Z. Burghard, D. Pisignano, X. Wang, Chemistry 19 (2013) 5790.

[8] S.A. Crawford, M.J. Higgins, P. Mulvaney, R. Wetherbee, J. Phycol. 37 (2001) 543.

[9] A.-M.M. Schmid, D. Schulz, Protoplasma 100 (1979) 267.

[10] M. Hildebrand, Chem. Rev. 108 (2008) 4855.

[11] C.C. Perry, T. Keeling-Tucker, J. Biol. Inorg. Chem. 5 (2000) 537.

[12] C.L. Chen, N.L. Rosi, Angew. Chemie - Int. Ed. 49 (2010) 1924.

[13] J.M. Galloway, S.S. Staniland, J. Mater. Chem. 22 (2012) 12423.

[14] Y. Ding, L. Shi, H. Wei, J. Mater. Chem. B 2 (2014) 8268.

[15] J.M. Galloway, J.P. Bramble, S.S. Staniland, Chemistry 19 (2013) 8710.

[16] J.N. Cha, K. Shimizu, Y. Zhou, S.C. Christiansen, B.F. Chmelka, G.D. Stucky, D.E. Morse, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 361.

[17] H.C. Schröder, D. Brandt, U. Schloßmacher, X. Wang, M.N. Tahir, W. Tremel, S.I. Belikov, W.E.G. Müller, Naturwissenschaften 94 (2007) 339.

[18] N. Kröger, R. Deutzmann, M. Sumper, Science 286 (1999) 1129.

[19] N. Poulsen, N. Kröger, J. Biol. Chem. 279 (2004) 42993.

[20] N. Poulsen, M. Sumper, N. Kröger, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 12075.

[21] A. Scheffel, N. Poulsen, S. Shian, N. Kröger, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3175.

[22] L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, M. Verlander, Biopolymers 55 (2000) 227.

[23] S.B.H. Kent, Chem. Soc. Rev. 38 (2009) 338.

[24] G.L. Rosano, E.A. Ceccarelli, Front. Microbiol. 5 (2014) 172.

[25] A.L. Demain, P. Vaishnav, Biotechnol. Adv. 27 (2009) 297.

[26] J.W. Cuozzo, H.H. Soutter, J. Biomol. Screen. 19 (2014) 1000.

[27] S. Gräslund, P. Nordlund, J. Weigelt, B.M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. Dhe-Paganon, H. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T.C. Terwilliger, C.-Y. Kim, L.-W. Hung, G.S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J.L. Sussman, R.C. Stevens, S.A. Lesley, I.A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M.I. Donnelly, W.H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S.K. Burley, J.S. Emtage, J.M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S.C. Almo, J.B. Bonanno, A. Fiser, S. Swaminathan, F.W. Studier, M.R. Chance, A. Sali, T.B. Acton, R. Xiao, L. Zhao, L.C. Ma, J.F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C.K. Ho, D. Wang, H. Wang, M. Jiang, G.T. Montelione, D.I. Stuart, R.J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K.C. Gunsalus, Nat. Methods 5 (2008) 135.

[28] E.V. Armbrust, Nature 459 (2009) 185.

[29] V. Smetacek, Protist 150 (1999) 25.

[30] M. Hildebrand, M.J. Doktycz, D.P. Allison, Pflügers Arch. - Eur. J. Physiol. 456 (2008) 127.

[31] P. Tréguer, D.M. Nelson, A.J. Van Bennekom, D.J. Demaster, A. Leynaert, B. Quéguiner, Science 268 (1995) 375.

[32] R.K. Iler, The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry, Wiley, New York, 1979.

[33] S. Sjöberg, J. Non. Cryst. Solids 196 (1996) 51.

[34] V. Martin-Jezequel, M. Hildebrand, M.A. Brzezinski, J. Phycol. 36 (2000) 821.

[35] M. Hildebrand, B.E. Volcani, W. Gassmann, J.I. Schroeder, Nature 385 (1997) 688.

[36] M. Hildebrand, K. Dahlin, B.E. Volcani, Mol. Gen. Genet. 260 (1998) 480.

[37] K. Thamatrakoln, A.J. Alverson, M. Hildebrand, J. Phycol. 42 (2006) 822.

[38] P. Curnow, L. Senior, M.J. Knight, K. Thamatrakoln, M. Hildebrand, P.J. Booth, Biochemistry 51 (2012) 3776.

[39] C. Gröger, M. Sumper, E. Brunner, J. Struct. Biol. 161 (2008) 55.

[40] E. Brunner, C. Gröger, K. Lutz, P. Richthammer, K. Spinde, M. Sumper, Appl. Microbiol. Biotechnol. 84 (2009) 607.

[41] R.W. Drum, H.S. Pankratz, J. Ultrastruct. Res. 10 (1964) 217.

[42] E.F. Stoermer, H.S. Pankratz, C.C. Bowen, Am. J. Bot. 52 (1965) 1067.

[43] V. V Annenkov, T.N. Basharina, E.N. Danilovtseva, M.A. Grachev, Protoplasma 250 (2013) 1147.

[44] E.G. Vrieling, W.W.C. Gieskes, T.P.M. Beelen, J. Phycol. 35 (1999) 548.

[45] B. Tesson, S. Masse, G. Laurent, J. Maquet, J. Livage, V. Martin-Jézéquel, T. Coradin, Anal. Bioanal. Chem. 390 (2008) 1889.

[46] B. Tesson, M.J. Genet, V. Fernandez, S. Degand, P.G. Rouxhet, V. Martin-Jézéquel, ChemBioChem 10 (2009) 2011.

[47] B. Tesson, M. Hildebrand, PLoS One 8 (2013) e61675.

[48] N. Kröger, C. Bergsdorf, M. Sumper, EMBO J. 13 (1994) 4676.

[49] N. Kröger, G. Lehmann, R. Rachel, M. Sumper, Eur. J. Biochem. 250 (1997) 99.

[50] N. Kröger, R. Deutzmann, M. Sumper, J. Biol. Chem. 276 (2001) 26066.

[51] A. Chiovitti, R.E. Harper, A. Willis, A. Bacic, P. Mulvaney, R. Wetherbee, J. Phycol. 41 (2005) 1154.

[52] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14133.

[53] S. Wenzl, R. Hett, P. Richthammer, M. Sumper, Angew. Chemie 120 (2008) 1753.

[54] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, Science 298 (2002) 584.

[55] M.R. Knecht, D.W. Wright, Chem. Commun. (2003) 3038.

[56] C.C. Lechner, C.F.W. Becker, J. Pept. Sci. 20 (2014) 152.

[57] N. Poulsen, A. Scheffel, V.C. Sheppard, P.M. Chesley, N. Kröger, J. Biol. Chem. 288 (2013) 20100.

[58] M. Nemoto, Y. Maeda, M. Muto, M. Tanaka, T. Yoshino, S. Mayama, T. Tanaka, Mar. Genomics 16 (2014) 39.

[59] M. Sumper, G. Lehmann, ChemBioChem 7 (2006) 1419.

[60] T. Mizutani, H. Nagase, N. Fujiwara, H. Ogoshi, Bull. Chem. Soc. Jpn. 71 (1998) 2017.

[61] D.J. Belton, S. V Patwardhan, C.C. Perry, J. Mater. Chem. 15 (2005) 4629.

[62] E. Brunner, K. Lutz, M. Sumper, Phys. Chem. Chem. Phys. 6 (2004) 854.

[63] G.M. Lindquist, R.A. Stratton, J. Colloid Interface Sci. 55 (1976) 45.

[64] M. Sumper, E. Brunner, G. Lehmann, FEBS Lett. 579 (2005) 3765.

[65] M. Sumper, E. Brunner, Adv. Funct. Mater. 16 (2006) 17.

[66] M.C. Bridoux, R.G. Keil, A.E. Ingalls, Org. Geochem. 47 (2012) 9.

[67] D.J. Belton, S. V Patwardhan, V. V Annenkov, E.N. Danilovtseva, C.C. Perry, Proc. Natl. Acad. Sci. 105 (2008) 5963.

[68] M. Sumper, S. Lorenz, E. Brunner, Angew. Chemie Int. Ed. 42 (2003) 5192.

[69] P. Richthammer, M. Börmel, E. Brunner, K.-H. van Pée, ChemBioChem 12 (2011) 1362.

[70] R. Hedrich, S. Machill, E. Brunner, Carbohydr. Res. 365 (2013) 52.

[71] N. Kröger, M. Sumper, Protist 149 (1998) 213.

[72] N. Kröger, C. Bergsdorf, M. Sumper, Eur. J. Biochem. 239 (1996) 259.

[73] N. Kröger, R. Wetherbee, Protist 151 (2000) 263.

[74] F. Rodriguez, D.D. Glawe, R.R. Naik, K.P. Hallinan, M.O. Stone, Biomacromolecules 5 (2004) 261.

[75] A. Roehrich, G. Drobny, Acc. Chem. Res. 46 (2013) 2136.

[76] L. Senior, M.P. Crump, C. Williams, P.J. Booth, S. Mann, A. Periman, P. Curnow, J. Mater. Chem. B 3 (2015) 2607.

[77] R.R. Naik, P.W. Whitlock, F. Rodriguez, L.L. Brott, D.D. Glawe, S.J. Clarson, M.O. Stone, Chem. Commun. (2003) 238.

[78] C.C. Lechner, C.F.W. Becker, Chem. Sci. 3 (2012) 3500.

[79] R. Wieneke, A. Bernecker, R. Riedel, M. Sumper, C. Steinem, A. Geyer, Org. Biomol. Chem. 9 (2011) 5482.

[80] T. Coradin, J. Livage, Colloids Surf. B. Biointerfaces 21 (2001) 329.

[81] F. Weiher, M. Schatz, C. Steinem, A. Geyer, Biomacromolecules 14 (2013) 683.

[82] K.M. Hawkins, S.S.-S. Wang, D.M. Ford, D.F. Shantz, J. Am. Chem. Soc. 126 (2004) 9112.

[83] J. Pires, A.C. Fernandes, R. Avó, J. Mater. Sci. 49 (2014) 6087.

[84] S. V Patwardhan, S.J. Clarson, Mater. Sci. Eng. C 23 (2003) 495.

[85] N. Li, X. Zhang, Q. Wang, F. Wang, P. Shen, RSC Adv. 2 (2012) 3288.

[86] M.M. Tomczak, D.D. Glawe, L.F. Drummy, C.G. Lawrence, M.O. Stone, C.C. Perry, D.J. Pochan, T.J. Deming, R.R. Naik, J. Am. Chem. Soc. 127 (2005) 12577.

[87] L. Xia, Z. Li, Langmuir 27 (2011) 1116.

[88] E.G. Bellomo, T.J. Deming, J. Am. Chem. Soc. 128 (2006) 2276.

[89] S. V Patwardhan, R. Maheshwari, N. Mukherjee, K.L. Kiick, S.J. Clarson, Biomacromolecules 7 (2006) 491.

[90] P.A. Mirau, J.L. Serres, M. Lyons, Chem. Mater. 20 (2008) 2218.

[91] L. Xia, Y. Liu, Z. Li, Macromol. Biosci. 10 (2010) 1566.

[92] J.E. Baio, A. Zane, V. Jaeger, A.M. Roehrich, H. Lutz, J. Pfaendtner, G.P. Drobny, T. Weidner, J. Am. Chem. Soc. 136 (2014) 15134.

[93] T. Kuno, T. Nonoyama, K. Hirao, K. Kato, Chem. Lett. 41 (2012) 1547.

[94] Q. Wang, J. Yu, X. Zhang, D. Liu, J. Zheng, Y. Pan, Y. Lin, RSC Adv. 3 (2013) 2784.

[95] J. Yu, Q. Wang, X. Zhang, Appl. Surf. Sci. 311 (2014) 799.

[96] Q. Wang, J. Yu, J. Zheng, D. Liu, F. Jiang, X. Zhang, W. Li, RSC Adv. 3 (2013) 15955.

[97] S. Han, S. Cao, Y. Wang, J. Wang, D. Xia, H. Xu, X. Zhao, J.R. Lu, Chem. - A Eur. J. 17 (2011) 13095.

[98] Q. Meng, Y. Kou, X. Ma, Y. Liang, L. Guo, C. Ni, K. Liu, Langmuir 28 (2012) 5017.

[99] H. Xu, Y. Wang, X. Ge, S. Han, S. Wang, P. Zhou, H. Shan, X. Zhao, J.R. Lu, Chem. Mater. 22 (2010) 5165.

[100] S. Wang, X. Ge, J. Xue, H. Fan, L. Mu, Y. Li, H. Xu, J.R. Lu, Chem. Mater. 23 (2011) 2466.

[101] S.C. Holmström, P.J.S. King, M.G. Ryadnov, M.F. Butler, S. Mann, D.N. Woolfson, Langmuir 24 (2008) 11778.

[102] D.M. Eby, K.E. Farrington, G.R. Johnson, Biomacromolecules 9 (2008) 2487.

[103] D.M. Eby, G.R. Johnson, B.L. Farmer, R.B. Pandey, Phys. Chem. Chem. Phys. 13 (2010) 1123.

[104] D.M. Eby, K. Artyushkova, A.K. Paravastu, G.R. Johnson, J. Mater. Chem. 22 (2012) 9875.

[105] K. Artyushkova, P. Atanassov, ChemPhysChem 14 (2013) 2071.

[106] H. Kauss, K. Seehaus, R. Franke, S. Gilbert, R.A. Dietrich, N. Kröger, Plant J. 33 (2003) 87.

[107] K.-I. Sano, H. Sasaki, K. Shiba, Langmuir 21 (2005) 3090.

[108] L.B. Wright, T.R. Walsh, Phys. Chem. Chem. Phys. 15 (2013) 4715.

[109] E.E. Oren, R. Notman, I.W. Kim, J.S. Evans, T.R. Walsh, R. Samudrala, C. Tamerler, M. Sarikaya, Langmuir 26 (2010) 11003.

[110] N. Kröger, M.B. Dickerson, G. Ahmad, Y. Cai, M.S. Haluska, K.H. Sandhage, N. Poulsen, V.C. Sheppard, Angew. Chem. Int. Ed. Engl. 45 (2006) 7239.

[111] R.B. Merrifield, J. Am. Chem. Soc. 85 (1963) 2149.

[112] W. Hartmann, H.J. Galla, Biochim. Biophys. Acta 509 (1978) 474.

[113] D.E. Olins, A.L. Olins, Von Hippel, P H, J. Mol. Biol. 24 (1967) 157.

[114] N.Y. Yount, M.R. Yeaman, Ann. N. Y. Acad. Sci. 1277 (2013) 127.

[115] J. Jarczak, E.M. Kościuczuk, P. Lisowski, N. Strzalkowska, A. Jóźwik, J. Horbanczuk, J. Krzyzewski, L. Zwierzchowski, E. Bagnicka, Hum. Immunol. 74 (2013) 1069.

[116] Y. Li, Protein Expr. Purif. 80 (2011) 260.

[117] N.S. Parachin, K.C. Mulder, A.A.B. Viana, S.C. Dias, O.L. Franco, Peptides 38 (2012) 446.

[118] M. Sumper, Angew. Chemie Int. Ed. 43 (2004) 2251.

[119] V. V Annenkov, E.N. Danilovtseva, V.A. Pal'shin, V.O. Aseyev, A.K. Petrov, A.S. Kozlov, S. V Patwardhan, C.C. Perry, Biomacromolecules 12 (2011) 1772.

[120] C. Gröger, K. Lutz, E. Brunner, Cell Biochem. Biophys. 50 (2008) 23.

[121] K. Lutz, C. Gröger, M. Sumper, E. Brunner, Phys. Chem. Chem. Phys. 7 (2005) 2812.

[122] A. Jantschke, K. Spinde, E. Brunner, Beilstein J. Nanotechnol. 5 (2014) 2026.

[123] A. Bernecker, R. Wieneke, R. Riedel, M. Seibt, A. Geyer, C. Steinem, J. Am. Chem. Soc. 132 (2010) 1023.

[124] D.B. Robinson, J.L. Rognlien, C.A. Bauer, B.A. Simmons, J. Mater. Chem. 17 (2007) 2113.

[125] H. Menzel, S. Horstmann, P. Behrens, P. Bärnreuther, I. Krueger, M. Jahns, Chem. Commun. (2003) 2994.

[126] D. Belton, S. V Patwardhan, C.C. Perry, Chem. Commun. (2005) 3475.

[127] S. V Patwardhan, S.J. Clarson, Silicon Chem. 1 (2002) 207.

[128] R. Zhao, B.-L. Su, Mater. Lett. 74 (2012) 163.

[129] R.-H. Jin, J.-J. Yuan, Polym. J. 39 (2007) 464.

[130] V. V Annenkov, E.N. Danilovtseva, Y. V Likhoshway, S. V Patwardhan, C.C. Perry, J. Mater. Chem. 18 (2008) 553.

[131] K.D. Demadis, A. Stathoulopoulou, Ind. Eng. Chem. Res. 45 (2006) 4436.

[132] K.D. Demadis, J. Chem. Technol. Biotechnol. 80 (2005) 630.

[133] E. Neofotistou, K.D. Demadis, Colloids Surfaces A Physicochem. Eng. Asp. 242 (2004) 213.

[134] K.D. Demadis, E. Neofotistou, Chem. Mater. 19 (2007) 581.

[135] K. Spinde, K. Pachis, I. Antonakaki, S. Paasch, E. Brunner, K.D. Demadis, Chem. Mater. 23 (2011) 4676.

[136] K.D. Demadis, A. Ketsetzi, K. Pachis, V.M. Ramos, Biomacromolecules 9 (2008) 3288.

[137] K.D. Demadis, K. Pachis, A. Ketsetzi, A. Stathoulopoulou, Adv. Colloid Interface Sci. 151 (2009) 33.

[138] M. Preari, K. Spinde, J. Lazic, E. Brunner, K.D. Demadis, J. Am. Chem. Soc. 136 (2014) 4236.

[139] K.D. Demadis, A. Tsistraki, A. Popa, G. Ilia, A. Visa, RSC Adv. 2 (2011) 631.

[140] S. V Patwardhan, G.E. Tilburey, C.C. Perry, Langmuir 27 (2011) 15135.

[141] N.R. Haase, S. Shian, K.H. Sandhage, N. Kröger, Adv. Funct. Mater. 21 (2011) 4243.

[142] S. Roeder, S. Hobe, H. Paulsen, Langmuir 30 (2014) 14234.

[143] C. Forsyth, T.W.S. Yip, S. V Patwardhan, Chem. Commun. 49 (2013) 3191.

[144] B.H. Jo, J.H. Seo, Y.J. Yang, K. Baek, Y.S. Choi, S.P. Pack, S.H. Oh, H.J. Cha, ACS Catal. 4 (2014) 4332.

[145] L. Betancor, C. Berne, H.R. Luckarift, J.C. Spain, Chem. Commun. (2006) 3640.

[146] H.R. Luckarift, L.J. Nadeau, J.C. Spain, Chem. Commun. (2005) 383.

[147] A.Y. Khan, S.B. Noronha, R. Bandyopadhyaya, Biochem. Eng. J. 91 (2014) 78.

[148] R.A. Sheldon, S. van Pelt, Chem. Soc. Rev. 42 (2013) 6223.

[149] R.R. Naik, M.M. Tomczak, H.R. Luckarift, J.C. Spain, M.O. Stone, Chem. Commun. (2004) 1684.

[150] C. Forsyth, S. V Patwardhan, J. Mater. Chem. B 1 (2013) 1164.

[151] Q. Chen, G.L. Kenausis, A. Heller, J. Am. Chem. Soc. 120 (1998) 4582.

[152] J.-K. Lai, T.-H. Chuang, J.-S. Jan, S.S.-S. Wang, Colloids Surfaces B Biointerfaces 80 (2010) 51.

[153] W.D. Marner, A.S. Shaikh, S.J. Muller, J.D. Keasling, Biotechnol. Prog. 25 (2009) 417.

[154] S.A. Miller, E.D. Hong, D. Wright, Macromol. Biosci. 6 (2006) 839.

[155] Y. Zhang, H. Wu, J. Li, L. Li, Y. Jiang, Y. Jiang, Z. Jiang, Chem. Mater. 20 (2008) 1041.

[156] S. Emond, D. Guieysse, S. Lechevallier, J. Dexpert-Ghys, P. Monsan, M. Remaud-Siméon, Chem. Commun. 48 (2012) 1314.

[157] L. Zhou, C. Wang, Y. Jiang, J. Gao, Chinese J. Chem. Eng. 21 (2013) 670.

[158] T. Moreau, C. Depagne, G. Suissa, H. Gouzi, T. Coradin, J. Mater. Chem. B 1 (2013) 1235.

[159] I.-C. Kuan, C.-A. Chuang, S.-L. Lee, C.-Y. Yu, Biotechnol. Lett. 34 (2012) 1493.

[160] A. Rai, A. Prabhune, C.C. Perry, Mater. Sci. Eng. C 32 (2012) 785.

[161] Y. Kawachi, S. Kugimiya, H. Nakamura, K. Kato, Appl. Surf. Sci. 314 (2014) 64.

[162] J. Li, H. Wu, Y. Liang, Z. Jiang, Y. Juang, L. Zhang, J. Biomater. Sci. 24 (2013) 119.

[163] L. Zhang, J. Shi, Z. Jiang, Y. Jiang, R. Meng, Y. Zhu, Y. Liang, Y. Zheng, ACS Appl. Mater. Interfaces 3 (2011) 597.

[164] J. Shi, Z. Jiang, J. Mater. Chem. B 2 (2014) 4435.

[165] P.J. Baker, S. V Patwardhan, K. Numata, Macromol. Biosci. 14 (2014) 1619.

[166] N. Poulsen, C. Berne, J. Spain, N. Kröger, Angew. Chemie Int. Ed. 46 (2007) 1843.

[167] V.C. Sheppard, A. Scheffel, N. Poulsen, N. Kröger, Appl. Environ. Microbiol. 78 (2012) 211.

Section I - Expression and purification of designed polypeptides with high cationic net charge facilitated by a charge-compensating or a hydrophobic protein fusion

Introduction

Many proteins can be efficiently expressed in recombinant *Escherichia coli*, however, some are challenging to express, so gene sequences or expression conditions have to be modified to suite with the bacterial host [1-3]. One important issue for heterologous expression of eukarya-derived genes is the codon usage. The genetic code for translation of genes into peptides and proteins is considered basically universal, however, prokarya favor different codons than eukarya within a degenerated genetic code [4]. In effect, expression of some heterologous proteins may be stalled because of insufficient supply of certain aminoacyl-tRNAs during translation, which can lead to evasion reactions, causing frameshifts and subsequently mis-translations [5]. This can be overcome either by changing the gene sequence to more abundant codons, or by using bacterial strains with additional, e.g. plasmid-encoded tRNAs increasing their abundance [6].

Further problems can be due to toxicity of the recombinant protein after expression. A special case where this issue becomes important is the bacterial expression of antimicrobial peptides [7,8], which are intended to harm bacteria. Detrimental effects to the host can be circumvented by using eukaryotic hosts, however, bacterial hosts have advantages in terms of growth rate, easy genetic modification techniques and facile culture conditions [1,9].

To allow recombinant expression of toxic peptides and proteins in *Escherichia coli*, fusion systems have been developed which detoxify proteins of interest, and can be cleaved off after expression and isolation of the construct. For cationically charged peptides (i.e. antimicrobial peptides), several fusions have been evaluated. Anionic, charge-compensating fusions [10,11] were used for detoxification, presumably by reducing advert interactions with e.g. anionic bacterial membranes. Solubility enhacing fusions can also facilitate the overexpression, shown with maltose binding protein (MBP) [12-14], thioredoxin (trx) [15-41], self-cleaving intein [16,42-46], ubiquitin [47] and small ubiquitin-like modifiers (SUMO) [34,36,44,48-56]. As an alternative concept, fusions can be used to mediate rapid deposition of target genes into inclusion bodies [57]. This fusion strategy was successful in the production of effector peptides [58], antimicrobial peptides exerting activity in their extended conformation [59-62], self-assembling peptides [63-66], intrinsically disordered proteins [67] as well as protein domains [68]. All of these can help to avoid degradation of recombinant proteins, but this is not granted in general and the constructs may be not expressed at all. Another option is hence to fuse export leader peptides [69] which sequester the recombinant proteins into the periplasm or the extracellular medium, where protease degradation is less likely [70,71].

For removal of the fusion, chemical cleavage methods [57] or protease digestion at specific sequences [72,73] can be applied. Alternatively, self-cleaving, inducible fusions (e.g. intein) can be used for the expression of a target protein or polypeptide [74]. Protease digestion and self-cleavable fusions are advantageous and facile for soluble proteins which are are folded into their native conformation within the bacterial cell. Hydrophobic proteins, especially when deposited into inclusion bodies after expression, require harsh denaturing conditions for solubilization. These are usually not appropriate for protease digestions, but instead, chemical cleavage methods can be used. However, because of the denaturing conditions, this methodology is restricted to proteins which can be reconstituted *in vitro*.

In this work, different fusions for the expression of designed polypeptides with high cationic net charge were evaluated. The polypeptides were similar to the silaffin-peptides from *Cylindrotheca fusiformis* [75] in terms of amino acid composition and contained kinase target sites for *in vitro* enzymatic phosphorylation. The prototypes were established in the diploma thesis of Christian Roos [76] and termed P_xS_y to describe their composition of several phosphorylation sites "P" and silaffin-like domains "S", the indices "x" and "y" denoting their respective numbers. The general gene construction procedure is presented in section III and published elsewhere [77]. In the thesis of Christian Roos, the P_xS_y polypeptides were ligated in frame with the pelB-leader [78] sequence (in the plasmid pET 22b(+)), which proved to be indispensable for expression although did not lead to export into the periplasm [76,79,80]. Furthermore, the plasmid provided a hexahistidyl tag carboxy-terminal of the recombinant expression product. The pelB-leader was separated from P_xS_y by a sequence coding for a factor Xa protease cleavage site [76,80] for removal of the leader after expression. Because of unintended digestion of P_xS_y by factor Xa protease [76], the target site was later exchanged for a HRV3C site [79]. A cysteine which was in the N-terminal region of the initial P_xS_y genes was exchanged for serine by site directed mutagenesis [81].

Attempts to express P_xS_y-genes in *Escherichia coli* without any fusion failed, presumably because they either were not overexpressed or immediately digested by proteases [79]. pelB-fused P_xS_y-versions were well expressed, however, efforts to purify them by immobilized metal affinity chromatography followed by protease cleavage of the fusion and purification by cation exchange chromatography did not result in satisfactory yields and purities [76,79,81].

Starting from these pelB-preceded P_xS_y genes, fusion strategies from the recombinant expression of antimicrobial peptides were evaluated. Because of the high cationic charge of P_xS_y, a charge neutralizing anionic leader from the silaffin *sil1*-gene [75] was examined for detoxification. Additionally, ketosteroid isomerase (KSI) [58,82] was evaluated as hydrophobic fusion which promotes the deposition in inclusion bodies.

Abbreviations

IEC: ion exchange chromatography; IMAC: immobilized metal affinity chromatography; IPTG: isopropyl β-d-1-thiogalactopyranoside; KSI: ketosteroid isomerase; HRV3C: human rhinovirus 3C protease; PAGE: polyacrylamide gel electrophoresis; PCR: polymerase chain reaction; pelB: export leader sequence of the *Erwinia carotovora* pectate lyase gene; SDS: sodium dodecylsulfate; SilL: anionic leader domain of the *Sil1* gene of *Cylindrotheca fusiformis*; Tris: tris(hydroxymethyl)aminomethane.

Materials and Methods

General considerations

To distinguish the versions of P_xS_y, a systematic terminology is applied through this thesis. The overall sequence of all P_xS_y peptides is as follows:

*(GPAM)***GS***(Y)***SRRASLG**[KSKKLRRASLG]$_m$**KLRRASL***(EHHHHHH)*

(sequence written in the standard amino acid single letter code [83])

The bold letters occurred in all P_xS_y polypeptides. The sequence "LRRASLG" marks a target site for protein kinase A [84] and its occurrence is denoted in the systematic name by "P_x", where "x" gives the numbers of the site in the polypeptide. The core-sequence written in squared brackets was inserted in varying numbers, denoted by "m". Since this repetitive sequence shares the amino acid composition with the silaffin-peptide [75], it is termed silaffin-like and abbreviated in the systematic name by "S_y", the number of repeats denoted by "y". Sequences in italics and rounded brackets did not occur in all peptide versions. The amino-terminal extension "GPAM" only occurred in the peptide versions produced with the double-leader sequence pelB-SilL (see below) and was a remnant from the proteolytic removal of the leader by HRV3C protease. The tyrosine (Y) was inserted for UV/vis quantification. Since it is therefore of particular importance, sequences missing the tyrosine are termed "P_xS_y(-Y)". The carboxy-terminal hexahistidyl-tag, spaced by a glutamate, was only retained when necessary for purification by immobilized metal affinity chromatography (IMAC). Versions bearing the tag are named "P_xS_y(H)". An overview of all versions created and used in this study is given in the results section (Table 1) and all sequences of fusion polypeptides are presented in the appendix at the end of this thesis.

All cloning and protein expression procedures were done according to standard biological procedures [85] and are described in detail in the following paragraphs. Restriction enzymes used were from New England Biolabs GmbH (Frankfurt, Germany), except where noted otherwise, and were generally used as recommended by the suppliers. Plasmids used were in the well-

established pET-vector system initially created by Maffot and Studier [86] and meanwhile widely available commercially, the advantages and challenges being discussed elaborately elsewhere [1].

After creation of new plasmids, DNA-solutions were used for transformation of chemically competent *Escherichia coli* [87]. Different strains used are noted in the appropriate paragraphs. For long-term storage of the transformants, aliquots of overnight-cultures were supplemented with 50 % w/v glycerol and frozen at -80 °C.

The P_xS_y sequences were developed by Christian Roos [76] in his diploma thesis, the methodology is described in the manuscript in section III and published elsewhere [77]. For the generation of P_xS_y genes in first instance, primer hybrids PIN 1 - 3 [76] were assembled by using a ligase and inserted into the plasmid pET 22b(+) (Merck Chemicals GmbH, Schwalbach (Germany)). By this, the target sequences had an amino-terminal pelB leader [78] and a carboxy-terminal hexahistidine tag. The pelB and P_xS_y were separated later by a HRV3C protease cleavage site [88] for later removal of the leader, introduced by insertion of a primer hybrid [79].

Some of the P_xS_y sequences were transferred into plasmid pET 31b(+) (Merck Chemicals GmbH, Schwalbach (Germany)) in the Bachelor thesis of Janica Wiederstein [89]. This plasmid added an amino-terminal ketosteroid isomerase fusion [58,82], removable by cyanogen bromide chemical cleavage at a methionine located in the amino-terminal region of P_xS_y.

Cloning and mutagenesis procedures

The anionic leader sequence of the *sil1*-gene (Genbank **AF191634.1**, kindly provided by Nils Kröger), which was available on a pUC 18 plasmid, was amplified by polymerase chain reaction (PCR) and ligated into the pET 22b(+) plasmids which already contained the P_xS_y genes, preceded by the pelB leader sequence. PCR-amplification was accomplished with *Pwo*-polymerase (Peqlab Biotechnology GmbH, Erlangen (Germany)) by using the primers Sil1p-L_f (5'-AAACATATGGTGGCCAGCGACTCCTCGGATGACGC-3') and SilL-L_r (5'-AAAAAGAATTCTCGGGGATACGAAGTTCTTCCTCTTCGG-3'). PCR-conditions were: initial denaturation 2 min 95 °C, cycle (35x) [denaturation 1 min 95 °C, annealing 30 sec 52 °C, elongation 30 sec 72 °C], final elongation 5 min 72 °C. After the amplification, PCR-fragments were isolated with peqGOLD Cycle Pure kit (Peqlab Biotechnology GmbH, Erlangen (Germany)). Subsequently, a double-digest of the PCR amplificate and the desired target plasmids using *Nde*I and *Eco*RI was performed for 2 hours at 37 °C. Fragments were separated by agarose gel electrophoresis in 1.5 % w/v agarose gels, at 190 V, in TAE running buffer (20 mmol/L acetic acid, 1 mmol/L ethylenediaminetetraacetic acid, 40 mmol/L Tris, pH 8.4), and visualized by staining with GelRed Nucleic Acid Gel Stain (Biotium Hayward (USA)). After gel extraction using peqGOLD Gel Extraction kit (Peqlab Biotechnology GmbH, Erlangen (Germany)), a second, analytical, gel run was performed and the amount of DNA estimated by comparing the band intensity with a marker sample of known DNA amounts (Quick load 100 bp or 1 kbp ladder, New England Biolabs GmbH, Frankfurt (Germany)). For ligation, Quick

Ligation Kit (New England Biolabs GmbH, Frankfurt (Germany)) was used with 50 ng of target plasmid DNA and a three-fold molar excess of the insert (PCR) fragment.

For the insertion of the SilL leader sequence, the sequence coding for the pelB leader was excised. For re-insertion, i.e. creation of a pelB-SilL double leader, plasmids containing pelB-P_xS_y and SilL-P_xS_y were double-digested with *Eco*RV (cutting upstream of the multiple cloning site, within the plasmid's *lacI* gene) and *Msc*I (cutting in the 3' region of the pelB leader sequence as well as immediately in the 5' region of SilL, the latter site was introduced by primer Sil1p-L_f). By this, a fragment terminated 3' by the pelB coding region was created and inserted into the plasmid containing the SilL-leader.

Transfer of this new double-leader sequence to other pET 22b(+) plasmids containing P_xS_y genes was accomplished by double digestion with *Eco*RV (see above) and *Eco*RI (cutting immediately upstream of the HRV3C protease target site separating the leader from the target sequences). The resulting fragment containing the double leader was then inserted into the plasmid of interest.

Later transfer into pET 31b(+) was possible by digestion of both pET 22b(+) containing the pelB-SilL-P_xS_y fusion, and the target plasmid, using the restriction enzyme *Ava*I. Before digestion, the complete region was PCR-amplified (see above) to increase the amount of insert DNA. Primers used for this amplification were T7PrompET22b(+)fw (5'-TAATACGACTCACTATAGGG-3') and pET22b(+)rev (5'-GCTCAGCGGTGGCAGC-3'), the PCR conditions were: initial denaturation 1 min 95 °C, cycle (35x) [denaturation 30 s 95 °C, anneal 30 sec 50 °C, elongate 1 min 72 °C], final elongation 5 min 72 °C. After isolation of the PCR-product by using the Cycle Pure Kit, *Ava*I digestion was done for the amplificate and the target plasmid. Further treatments were as described above for subcloning procedures after double-digestion.

Plasmids ligated with respective inserts were transformed into *Escherichia coli* strain DH5-α (Life Technologies GmbH, Darmstadt (Germany)). Transformants were identified *via* growth on LB-media [90,91] solidified with 1.5 % w/v agar agar and supplemented with 100 µg/ml ampicillin, taking advantage of the resistance gene provided by the plasmids. Transformants were confirmed by isolation of the plasmids from liquid cultures (peqGOLD Miniprep Kit II, Peqlab Biotechnology GmbH, Erlangen (Germany)) and sequencing at Starseq GmbH (Mainz, Germany) by using plasmids T7PrompET22b(+)fw or pET22b(+)rev (see above).

Changes within the P_xS_y target sequences were achieved by site-directed mutagenesis using the QuikChange Lightning Site-Directed Mutagenesis kit or QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies GmbH, Waldbronn (Germany)). For insertion of a tyrosine codon in the amino-terminal region of the P_xS_y-genes (for UV/vis quantification), the primer combination InsY-s (5'-GGCCATGGGATCCTATTCTCGACGTGCTTCC-3') and InsY-as (5'-GGAAGCACGTCGAGAATAGGATCCCATGGCC-3') were used. For further insertion of a cysteine into the tyrosine-bearing polypeptide versions, the primers CtoYPxSy_fw (5'-GGCCATGGGATGCTATTCTCGACGTGCTTCC-3') and CtoYPxSy_r (5'-GGAAGCACGTCGAGAATAGCATCCCATGGCC-3') were used. Expression of the

hexahistidyl-tag was suppressed by changing a glutamate prior of the tag into a stop-codon, using primers EtoStopPxSy_s (5'-CGCGCTAGCCTCTAGCACCACCACC-3') and EtoStopPxSy_as (5'-GGTGGTGGTGCTAGAGGCTAGCGCG-3'). An additional mutagenesis was done in the pelB leader sequence, to change a lysine (second amino acid of the leader) into a glutamate, by using primers KtoEinpelB_f (5'-GGAGATATACATATGGAATACCTGCTGCCGACCG-3') and KtoEinpelB_r (5'-CGGTCGGCAGCAGGTATTCCATATGTATATCTCC-3'). New plasmids were used for transformation of chemically competent *Escherichia coli* strains Xl1-blue or Xl10-gold, which were supplied with the mutagenesis kits, and proven by sequencing as described above.

Expression of P_xS_y fusion polypeptides

For expression, plasmids pET 22b(+) or pET 31b(+) were transformed into *Escherichia coli* strain Rosetta (Merck4Biosciences GmbH, Schwalbach (Germany)). This strain contains an additional plasmid providing tRNAs of codons which are rarely used in *Escherichia coli*, and provides an additional antibiotic resistance gene against chloramphenicol. Transformants were screened on agar (1.5 % w/v) plates with LB-medium [90,91]. LB media for protein expression were supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and, where noted, with 0.1% (w/v) glucose.

The expression of different versions of pelB-SilL-P_xS_y and KSI-P_xS_y was evaluated by diluting overnight pre-cultures (LB-medium) of recombinant bacteria 50-fold into fresh medium. The optical density at 600 nm (OD_{600}) was traced and the cultures induced at different culture stages (OD_{600} as stated in the particular results). Total protein samples were prepared by harvesting cells from 0.5 ml medium (5 min at 5,000 g) and storing the cell pellets at -20 °C. Pellets were usually resuspended in water according to a cell density OD_{600} of 10 (i.e. 50 µl water when cells were harvested from 500 µl at OD_{600} = 1). For SDS-PAGE analysis (see last section of materials and methods) of the total protein samples, 5 µl or 3 µl of this suspension were applied to a 10-well or 15 well PAGE, respectively (size of the separating gels 8.5 cm width and 6 cm height in both cases) to ensure compatibility between the samples.

Expression and lysis of bacteria

Expression and purification protocols were different for the pelB-SilL-P_xS_y (pET 22b(+)) and KSI-P_xS_y (pET 31b(+)) fusion proteins. The protocol for expression and purification of the latter, KSI-P_xS_y, is described in section III and published elsewhere [77]. Therefore, in the following, only the protocol for purification of pelB-SilL-P_xS_y is described.

For standard expression of pelB-SilL-P_xS_y (pET 22b(+)), an overnight pre-culture of 800 ml was grown at 37 °C and 185 rpm. Bacteria were harvested by centrifugation at 5,000 g for 5 minutes and resuspended in an equal volume of fresh culture media. After further 1.5 hours growth, the

culture was cooled at 4 to 8 °C for around one hour and subsequently protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mmol/L. Expression was allowed overnight at 20 °C and 185 rpm. Cells were harvested (5 min at 5,000 g), resuspended in 50 ml lysis buffer (50 mmol/L Tris-HCl pH 8.0, 1 mmol/L ethylenediamintetraacetic acid, 0.5 mg/ml lysozyme) and let stand for 20 min at 37 °C for lysozyme digestion. For DNA-digestion, 50 units Benzonase, 2 mmol/L magnesium chloride and 20 mmol/L dithiothreitole were added and the sample incubated for 20 min at 37 °C. Cell debris containing the target polypeptide was collected by centrifugation at 17,500 g for 5 min and the supernatant was discarded.

Further preparation of the pellet fraction differed for target polypeptide versions with and without a hexahistidine tag. Polypeptide fusions bearing a hexahistidine tag were purified by immobilized metal affinity chromatography (IMAC). The insoluble residue was therefore dissolved in 1 % w/v SDS (20 ml per 800 ml culture volume), buffered with 40 mmol/L Tris-HCl pH 7.5, by alternated heating (boiling water bath) and vortexing. The resulting cell extract was diluted ten-fold with 20 mmol/L Tris-HCl pH 7.5 for application to the IMAC resin. Tag-free fusions were instead isolated by ion exchange chromatography. In this case, the insoluble residue was dissolved in urea buffer (10 ml per 800 ml culture volume; 8 mol/L urea, 40 mmol/L Tris-HCl pH 7.5, 100 mmol/L sodium chloride) by the aid of an ultrasonic device (VC600 high intensity ultrasonic processor with V1A converter and equipped with an 1/8-th inch (3 mm) tapered microtip, Sonics and Materials Inc., Newton (CT/United States of America)). Conditions for ultrasonication were 5 minutes with 50 % power output and 40 % duty cycle. Afterwards, the solution was cleared by centrifuging 5 min at 17,500 g and the supernatant applied to the IEC-resin.

IMAC purification of fusions bearing a heahistidine tag

For IMAC isolation of the fusion polypeptide, 10 ml Chelating Sepharose (GE-Healthcare GmbH, München (Germany)) were loaded with nickel (applied as 0.3 M nickel chloride) and equilibrated with Washing buffer III (see below). Loading of the target polypeptide from the SDS-extract (see sub-section "Expression and lysis of bacteria") onto the resin was achieved by shaking for one hour at 225 rpm and room temperature. Subsequently, the sample was transferred to a Biorad-cartridge and the extract was allowed to run through. The Sepharose resin was washed with 20 ml Washing buffer I (0.1 % w/v SDS, 100 mmol/L Tris-HCl pH 7.5, 200 mmol/L sodium chloride, 20 mmol/L imidazole), 30 ml Washing buffer II (8 mol/L Urea, 20 mmol/L Tris-HCl pH 7.5; used to get rid of SDS) and 20 ml of Washing buffer III (as washing buffer I, though without SDS). Elution was achieved with 20 ml Elution buffer (1 mol/L sodium chloride, 0.5 mol/L imidazole, 20 mmol/L Tris-HCl pH 7.5). To release the target peptide, 20 units HRV3C-protease (Merck) were added per 10 ml elution fraction and digestion was performed over night at 4 °C, with constant rotation. The high salt concentration in the elution buffer (0.5 mol/L imidazole, 1 mol/L sodium chloride) was applied since HRV3C protease is known to be highly active in high-salt buffer [92]. In case aggregates were visible after the

digestion, sodium hydroxide solution (1 mol/L) was added stepwise until aggregates were dissolved and the sample was subsequently neutralized by addition of the same amount of hydrochloric acid (1 mol/L). The digestion assay was diluted 10-fold with 20 mmol/L Tris-HCl pH 7.5 for application on an ion exchange resin (see below) for purification of the liberated polycationic target peptide.

IEC purification of fusions without an hexahistidine tag

For IEC isolation of the fusion polypeptide, the bacterial extract in urea (see sub-section "Expression and lysis of bacteria") was applied to 4 ml equilibrated SP-Sepharose (GE-Healthcare GmbH, München (Germany)) rinsed in advance with equilibration buffer (100 mmol/L sodium chloride, 20 mmol/L Tris-HCl pH 7.5). The resin was suspended in the solution and binding was allowed for one hour at room temperature with shaking at 225 rpm, after which the sample was transferred into a Biorad-cartridge. The resin was washed with two-times each 8 ml urea buffer (8 mol/L urea, 100 mmol/L sodium chloride, 40 mmol/L Tris-HCl pH 7.5), and each 8 ml of two urea-free buffers containing 100 and 500 mmol/L sodium chloride, respectively, in 20 mmol/L Tris-HCl pH 7.5. Elution was achieved with 1 mol/L sodium chloride in 50 mmol/L sodium hydroxide, which was immediately neutralized by adding 400 µl of 1 mol/L hydrochloric acid and 420 µl of 1 mol/L Tris-HCl pH 7.5 (final concentration around 50 mmol/L). For proteolytic digest, 16 units HRV3C protease were added and cleavage proceeded overnight at 4 °C with constant rotation. Alternatively, cyanogen bromide [93-95] cleavage was applied. The elution from the IEC was in this case neutralized by adding 400 µl of 1 mol/L hydrochloric acid, but not buffered with Tris. The sample was transferred to a round-bottom flask (light-protected by aluminum foil) and five-fold diluted by adding 30 ml deionized water and 2 ml of 10 mol/L hydrochloric acid (final concentration 0.5 mol/L) for acidification. The solution was stirred by a magnetic stirrer and nitrogen was flushed to lower the amount of oxygen in the solution. Subsequently, 800 mg cyanogen bromide dissolved in 1 ml acetonitrile were added and the assay was stirred overnight. The solvent was then removed by rotary evaporation and the residue dissolved in 40 ml of 50 mmol/L sodium hydroxide. After centrifugation for 5 min at 10,000 g to clear the solution from the insoluble residue, the supernatant was neutralized by adding 1.85 ml 1 mol/L hydrochloric acid and 2 ml 1 mol/L Tris-HCl pH 7.5 (final concentration around 50 mmol/L). The pH was checked with a test strip and in case it substantially deviated from pH 7.5 was adjusted by adding sodium hydroxide and hydrochloric acid. This solution was diluted two-fold with 50 mmol/L Tris-HCl pH 7.5 in preparation of the ion exchange chromatographic isolation of the liberated polycationic target (see below).

IEC purification of the liberated, polycationic P_xS_y target polypeptides

The chemical (cyanogen bromide) or proteolytic (HRV3C) cleavage assays were applied to 4 ml SP-Sepharose (GE-Healthcare GmbH, München (Germany)) rinsed in advance with equilibration

buffer (100 mmol/L sodium chloride, 20 mmol/L Tris-HCl pH 7.5). After 1 hour at 225 rpm and room temperature to allow for polypeptide binding, the sample was transferred into a Biorad-cartridge and the solution was allowed to run through. The resin was washed with 0.2 mol/L ammonium carbonate in the standard assay (10 ml, in ultrapure water with electrical resitance R > 18.2 MΩ), though only 0.1 mol/L were used for washing when the shortest P_xS_y-version, P_2S_0, was purified. Elution was achieved with 0.5 mol/L ammonium carbonate solution (10 ml, in ultrapure water). Finally, the elution fraction was vacuum dried in a Speedvac or lyophilizator, thereby removing the volatile ammonium carbonate salt. The polypeptide was stored in the dried form and dissolved in 50 mmol/L sodium hydroxide solution immediately prior to being used for experiments. Long-term exposure of the polypeptide to high pH was avoided.

SDS-PAGE analysis of the purification process

Protein samples collected during expression pattern analyzes or purification procedures were analyzed by SDS-PAGE using the Tris-Tricine system established by Schägger and von Jagow [96-98] and coomassie staining or alternatively western blotting [99] with a primary antibody raised against the hexahistidyl-tag. Blotting of proteins onto a nitrocellulose membrane was performed in blotting buffer (20 mmol/L Tris-HCl, 150 mmol/L glycine, 20 % v/v methanol, 0.05 % w/v SDS) for 2 - 2.5 hours at 200 milliampere. Membranes were rinsed with deionized water and blocked in blocking buffer (3 % w/v BSA, 10 mmol/L Tris-HCl pH 7.5, 150 mmol/L sodium chloride) for at least 1 hour. Subsequently, the first antibody (anti penta-histidine antibody (Qiagen GmbH, Hilden (Germany), concentration as recommended by the supplier) was applied in a 1:1 dilution of blocking buffer with deionized water, overnight at 4 °C. Excess antibodies were washed away once with TBSTT (10 mmol/L Tris-HCl pH 7.5, 250 mmol/L sodium chloride, 0.025 % w/v Tween 20, 0.1 % w/v TritonX-100) and twice with TBS (10 mmol/L Tris-HCl pH 7.5, 150 mmol/L sodium chloride) buffer for 10 min each, under shaking. The second antibody (anti-mouse, alkaline phosphatase conjugate (Sigma-Aldrich GmbH, Taufkirchen (Germany), concentration as recommended by the supplier) was also applied in a 1:1 dilution of blocking buffer with deionized water, for 1 hour at room temperature. Again, rinsing of the membrane with TBSTT and TBS was performed. For polypeptide detection, 5-Bromo-4-chloro-3-indolylphsophate (16.5 µg/ml) and nitro blue tetrazolium chloride (33 µg/ml) were applied in a buffer suitable for the reaction of alkaline phosphatase (100 mmol/L Tris-HCl pH 9.5, 100 mmol/L sodium chloride, 5 mmol/L magnesium chloride). The reaction was allowed to proceed for 10 to 15 minutes, which was usually sufficient for clear polypeptide bands to appear.

Results

Within this work, several plasmid constructs for the expression of different P_xS_y-versions were generated. An overview of all constructs, their systematic names and some characteristic features

(including information on their generation) are summarized in Table 1. The detailed cloning procedure for the most important cloning steps is described in the following paragraphs.

Table 1. Overview of gene constructs and nomenclature. Names were given systematically. Abbreviations mean: **P_xS_y** - Basic core sequence with phosphorylation sites (P) and silaffin-like sequences (S); **(H)** - hexahistidyl tag following P_xS_y; **(-Y)** - P_xS_y-versions which do not have a tyrosine for UV/vis quantification; **Cys** - P_xS_y-versions bearing a cysteine for potential post-expression modifications; **SilL** - anionic leader sequence from *Cylindrotheca fusiformis* silaffin gene *Sil1p*; **pelB** - leader sequence from *Erwinia carotovora* pelB gene, potential export leader to mediate translocation of proteins into the intermembrane space in *Escherichia coli*; **pelB*(X)*** - modified pelB-leader with exchange of lysine (following methionine start codon) to glutamate; **KSI** - Ketosteroid isomerase gene from *Pseudomonas testosteroni*, mediating protein deposition into inclusion bodies during overexpression.

General Amino Acid Sequence of P_xS_y (see Materials and Methods for explanation)			
(GPAM)GS***(Y)***SRRASLG**[KSKKLRRASLG]$_m$**KLRRASL***(EHHHHHH)*			
Name of construct used in this thesis	Ancient name or alternative name	Production and function	Sequence part altered (if applicable)
<u>Part A:</u> P_xS_y with silaffin leader (SilL) fusion			
[SilL-(FXa)Sx3] *cloning intermediate only*		Cloning of SilL-sequence (PCR-amplified) to former pelB-Sx3, separated by FXa protease cleavage site [100]	Leader region, exchange of pelB for SilL-leader
SilL-P_5S_3(-Y,H)		Cloning of SilL-sequence (PCR-amplified) to former pelB-ΔCP_5S_3 [81]	Leader region, exchange of pelB for SilL-leader
<u>Part B:</u> P_xS_y with pelB-SilL-double leader fusion			
pelB-SilL-CysP_2S_0(-Y,H)	SPZ-01, pelB-SilL-P_2S_0	In thesis Janica Wiederstein [89]; Subcloning of double leader pelB-SilL from pelB-SilL-P_5S_3(-Y,H) (this table) into pelB-P2S0 [76] (containing Factor Xa cleavage site instead of HRV3C)	Double leader pelB and SilL; Sequence around N-terminal methionine of P_xS_y is …MG**<u>C</u>**SRR...
pelB-SilL-CysP_3S_1(-Y,H)	SPZ-04,	Subcloning of double	Double leader pelB

	Former: pelB-SilL-P_3S_1	leader pelB-SilL from pelB-SilL-P_5S_3(-Y,H) (this table) into pelB-CysP_3S_1 (see [79])	and SilL; Sequence around N-terminal methionine of P_xS_y is …MG**C**SRR...
pelB-SilL-CysP_5S_3(-Y,H)	SPZ-07, Former: pelB-SilL-ΔCP_5S_3	Subcloning of double leader pelB-SilL from pelB-SilL-P_5S_3(-Y,H) (this table) into pelB-P_5S_3 (see [79])	Sequence around N-terminal methionine of P_xS_y is …MG**C**SRR...
pelB-SilL-P_2S_0(-Y,H)	SPZ-02, Former: pelB-SilL-ΔCP_2S_0	Exchange Cysteine to Serine in N-terminal region of pelB-SilL-CysP_2S_0(-Y,H) (this table), single site mutagenesis with primer hybrid established [81]	Sequence around N-terminal methionine of P_xS_y is …MG**S**SRR...
pelB*(X)*-SilL-CysP_2S_0(-Y,H)	SPLK-05, Former: pelB(KtoE)-SilL- ΔCP_2S_0	In thesis Laura Krebs [101]; site directed mutagenesis to exchange lysine (second amino acid in pelB-leader) to glutamate in pelB-SilL-CysP_2S_0(-Y,H)	pelB-leader sequence begins with M**E**YLLP... **NOTE:** This version has a 5' duplication (including some mutations; see sequence in appendix). This leads to an extended leader sequence which may affect its function as an export leader.
pelB-SilL-CysP_2S_0(-Y)	SPLK-06, Former: pelB-SilL-ΔCP_2S_0-ΔH	In thesis Laura Krebs [101]; site directed mutagenesis to exchange glutamate (prior to hexahistidyl-tag) to a stop-codon in pelB-SilL-CysP_2S_0(-Y,H)	P_xS_y sequence is translated to …RRASLG*(no histidine)…
pelB-SilL-P_3S_1(-Y,H)	SPZ-05, Former: pelB-SilL-ΔCP_3S_1	Exchange Cysteine to Serine in N-terminal region of pelB-SilL-CysP_3S_1(-Y,H) (this table), single site mutagenesis with primer hybrid established in [81]	Sequence around N-terminal methionine of P_xS_y is …MG**S**SRR...
pelB*(X)*-SilL-P_3S_1(-Y,H)	SPLK-03, Former:	In thesis Laura Krebs [101]; site directed	pelB-leader sequence begins with

	pelB(KtoE)-SilL-ΔCP_3S_1	mutagenesis to exchange lysine (second amino acid in pelB-leader) to glutamate from pelB-SilL-P_3S_1(-Y,H)	M**E**YLLP...
pelB-SilL-P_3S_1(-Y)	SPLK-04, Former: pelB-SilL-ΔCP_3S_1-ΔH	In thesis Laura Krebs [101]; site directed mutagenesis to exchange glutamate (prior to hexahistidyl-tag) to a stop-codon in pelB-SilL-P_5S_3(-Y,H)	P_xS_y sequence is translated to …RRASLG*(no histidine)…
pelB-SilL-P_5S_3(-Y,H)	SPZ-08, Former: pelB-SilL-ΔCP_5S_3	Insertion of pelB-sequence into SilL-P_5S_3 (see part A)	Double leader pelB-SilL; Sequence around N-terminal methionine of P_xS_y is …MG**S**SRR...
pelB*(X)*-SilL-P_5S_3(-Y,H)	SPLK-01, Former: pelB(KtoE)-SilL-ΔCP_5S_3	In thesis Laura Krebs [101]; site directed mutagenesis to exchange lysine (second amino acid in pelB-leader) to glutamate from pelB-SilL-P_5S_3(-Y,H)	pelB-leader sequence begins with M**E**YLLP...
pelB-SilL-P_5S_3(-Y)	SPLK-02, Former: pelB-SilL-ΔCP_5S_3-ΔH	In thesis Laura Krebs [101]; site directed mutagenesis to exchange glutamate (prior to hexahistidyl-tag) to a stop-codon in pelB-SilL-P_5S_3(-Y,H)	P_xS_y sequence is translated to …RRASLG*(no histidine)…
pelB-SilL-P_2S_0(H)	SPZ-03, Former: pelB-SilL-ΔCYP_2S_0	Insertion of tyrosine into pelB-SilL-P_2S_0(-Y,H) (this table) sequence by single site mutagenesis (codon insertion)	Sequence around N-terminal methionine of P_xS_y is …MGS**Y**SRR...
pelB-SilL-P_3S_1(H)	SPZ-06, Former: pelB-SilL-ΔCYP_3S_1	Insertion of tyrosine into pelB-SilL-P_3S_1(-Y,H) (this table) sequence by single site mutagenesis (codon insertion)	Sequence around N-terminal methionine of P_xS_y is …MGS**Y**SRR...
pelB-SilL-P_5S_3(H)	SPZ-09, Former: pelB-SilL-ΔCYP_5S_3	Insertion of tyrosine-codon in pelB-SilL-P_5S_3(-Y,H) (this table) sequence by single site mutagenesis	Sequence around N-terminal methionine of P_xS_y is …MGS**Y**SRR...

Part C: P_xS_y with KSI-fusion			
KSI-P_3S_1(-Y,H)	SPJW-03, Former: KSI-ΔCP_3S_1	In Thesis Janica Wiederstein [89]; Subcloning P3S1(-Y) (donor plasmid in this table pelB-SilL-P_3S_1(-Y, H), but hexahistidyl-region not cloned) into pET 31b(+)	KSI fusion for P_3S_1, intended for cyanogen bromide cleavage of expressed protein (no methionine remains in P_3S_1) and therefore N-terminal P_3S_1-sequence GSSRR...
KSI-P_5S_3(-Y,H)	SPJW-01, Former: KSI-ΔCP_5S_3	In Thesis Janica Wiederstein [89]; Subcloning P_5S_3(-Y) (donor plasmid in this table pelB-SilL-P_5S_3(-Y,H), but hexahistidyl-region not cloned) into pET 31b(+)	KSI fusion for P_5S_3, intended for cyanogen bromide cleavage of expressed protein (no methionine remains in P_5S_3) and therefore N-terminal P_5S_3-sequence GSSRR...
KSI-P_3S_1(H)	SPJW-04, Former: KSI-ΔCYP_3S_1	In thesis Janica Wiederstein [89]; Insertion of tyrosine into KSI-P_3S_1(-Y,H) (this table) sequence by single site mutagenesis (codon insertion)	N-terminal sequence of P_xS_y is GS**Y**SRR...
KSI-P_5S_3(H)	SPJW-02, Former: KSI-ΔCYP_5S_3	In thesis Janica Wiederstein [89]; Insertion of tyrosine into KSI-P_5S_3(-Y,H) (this table) sequence by single site mutagenesis (codon insertion)	N-terminal sequence of P_xS_y is GS**Y**SRR...
KSI-CysP_5S_3(H)		Site directed mutagenesis to exchange serine to cysteine in N-terminal region of KSI-P_5S_3(H) (this table)	N-terminal sequence of P_xS_y is G**C**YSRR...
KSI-P_5S_3	SPZ-10	Site directed mutagenesis to exchange glutamate (prior to hexahistidyl-tag) to a stop-codon in KSI-P_5S_3(H)	Tag-free core-sequence of P_5S_3

Cloning of an anionic leader sequence from the Sil1-gene for charge-compensated expression of P_xS_y genes

Starting from the pET 22b(+) plasmid hosting the P_xS_y(-Y,H) genes, separated from the *pelB* leader sequence by the HRV3C protease cleavage site [79], it was aspired to create a fusion of the anionic prosequence from the *Sil1* gene [75] with the P_xS_y(-Y,H) genes.

The *Sil1* anionic leader, denoted as SilL, was amplified from pUC 18 hosting the complete *Sil1* gene. A primer combination (Sil1p-L_f/Sil1p-L_r) was used to specifically amplify the SilL-sequence and to introduce cleavage sites for restriction enzymes *Nde*I and *Eco*RI. The expected amplificate length of 223 basepairs was verified by agarose gel electrophoresis (Figure 1).

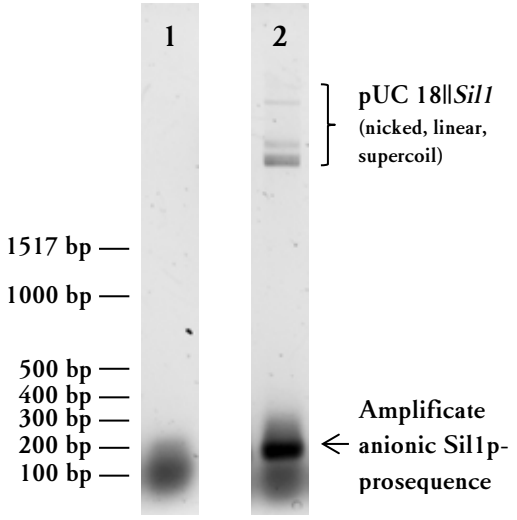

Figure 1. PCR-amplification of the *Sil1* anionic leader. The PCR-amplification was done with forward primer Sil1p_f and reverse primer Sil1p_r. The agarose gel electrophoretic analysis of a negative control without template plasmid (*lane 1*) and a successful PCR-assay (*lane 2*) are shown.

Given the high amount of template DNA seen in agarose gel analysis, the PCR-amplificate was electrophoretically separated, isolated from the gel, and subsequently used as template in a second PCR run. This procedure ensured the absence of template DNA and that a sufficiently large amount of PCR-amplificate for later cloning procedures was available. For ligation into the target plasmid, both the PCR-product and the plasmid were double digested with restriction enzymes *Nde*I and *Eco*RI. Ligation was initially successful in pET 22b(+), the desired plasmid hosting a recombinant gene termed Sx3 [100], which was a sequence for a polycationic peptide other than the desired P_xS_y (and was not considered in later studies). From this construct, the SilL leader sequence was excised with *Eco*RV and *Eco*RI (Figure 2). The first enzyme cutted the plasmid DNA far upstream of the recombinant gene, within the plasmid's *lacI*-gene, the latter enzyme between the leader sequence and the Sx3 or P_xS_y coding sequence, respectively. By this, a 1,346 bp fragment, bearing the pelB coding sequence, was excised from the target plasmids hosting the P_xS_y genes, leaving a 4204 and 4270 bp fragment for plasmids with P_3S_1(-Y,H) and P_5S_3(-Y,H), respectively. From the Sx3-bearing plasmid, a 1484 bp segment was excised bearing the SilL sequence. Fragments of the expected sizes were identified by agarose gel electrophoresis

(Figure 2), purified, and ligated to yield the intended versions SilL-P_xS_y(-Y,H). The result from sequencing of the new plasmid hosting SilL-P_5S_3(-Y,H) is shown in Figure 3.

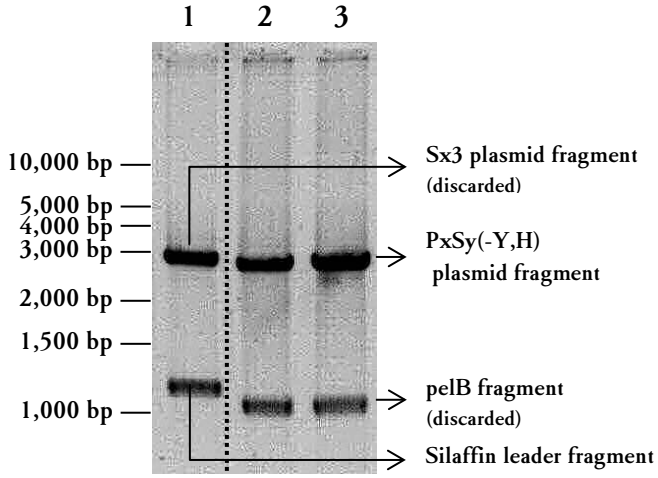

Figure 2. Double digestion with *Eco*RV/*Eco*RI for subcloning of the anionic silaffin leader sequence SilL. Double digestion was performed for the new plasmid construct hosting SilL-Sx3 *(lane 1)*, as well as target plasmids hosting pelB-P_xS_y(-Y,H), in particular P_3S_1(-Y,H) (*lane 2*) and P_5S_3(-Y,H) (*lane 3*), and the fragments were analyzed by agarose gel electrophoresis.

>anionic leader SilL from Sil1
ATGGTGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCC < 60
M V A S D S S D D A S D S S V E S V D A
10 20 30 40 50
GCCTCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTG < 120
A S S D V S G S S V E S V D V S G S S L
70 80 90 100 110
GAATCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCC < 180
E S V D V S G S S L E S V D D S S E D S
130 140 150 160 170
>HRV3C-Site
GAAGAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC < 240
E E E E L R I P E N S L E V L F Q G P A
190 200 210 220 230
>P_5S_3(-Y,H)
ATGGGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA < 300
M G S S R R A S L G K S K K L R R A S L
250 260 270 280 290
GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCT < 360
G K S K K L R R A S L G K S K K L R R A
310 320 330 340 350
>His_6
AGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA < 414
S L G K L R R A S L E H H H H H H *
370 380 390 400 410

Figure 3) Sequence of the SilL-P_5S_3(-Y,H) fusion. The anionic silaffin leader SilL replaced the former pelB leader sequence. The target gene of P_5S_3(-Y,H), bearing a hexahistidyl-tag, is seperated from its leader by a linker coding for a HRV3C protease cleavage site.

The new expression plasmid pET 22b(+) bearing SilL-P_5S_3(-Y,H) was used to create an expression plasmid with a double leader consisting of pelB and SilL. For this, SilL-P_5S_3(-Y,H) (anionic leader only) and pelB-P_5S_3(-Y,H) (pelB leader only) bearing plasmids were double-digested with *Eco*RV (see above for description of the cleavage site location) and *Msc*I (cutting in the 5'-region of SilL within the second translated codon as well as in the 3'-region of the pelB sequence). The double-digestion of the pelB-P_xS_y(-Y,H) bearing plasmid resulted in a 1.339 bp fragment containing the pelB coding sequence, which was ligated into the SilL-P_5S_3(-Y,H) bearing target vector. The success of the cloning procedure to generate the pelB-SilL double leader was proven by sequencing (Figure 4).

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from Sil1p
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P5S3(-Y,H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 420
K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         370       380       390       400       410
                              >His6
CTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 471
L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         430       440       450       460       470
```

Figure 4) Sequence of the pelB-SilL-P_5S_3(-Y,H) fusion. This construct contains a double-leader of pelB and anionic silaffin-leader SilL, followed by HRV3C protease site and hexahistidine-tagged P_5S_3(-Y,H).

This construct pelB-SilL-P_5S_3(-Y,H) served as a source of the double leader which was transferred to shorter versions of the P_xS_y(-Y,H) series, namely P_3S_1(-Y,H) and P_2S_0(-Y,H), which were both already available on pET 22b(+).

Transfer of the leader to P_3S_1(-Y,H) was achieved by excision with *Eco*RV (see above) and *Eco*RI (cutting immediately prior to the HRV3C site) (Figure 5). The initial cloning was achieved with a P_3S_1-version which contained a cysteine as second amino acid following the methionine of P_xS_y (name: CysP_3S_1(-Y,H)). This was later replaced for serine (see below).

P_2S_0(-Y,H) was only available in a pelB-fused version separated by a protease cleavage site for factor Xa protease. However, both factor Xa, and HRV3C site, were succeded by a restriction site for *Nco*I, so that the subcloning of the double leader was achieved by *Eco*RV/*Nco*I digestion and ligation of the appropriate fragments Figure 6).

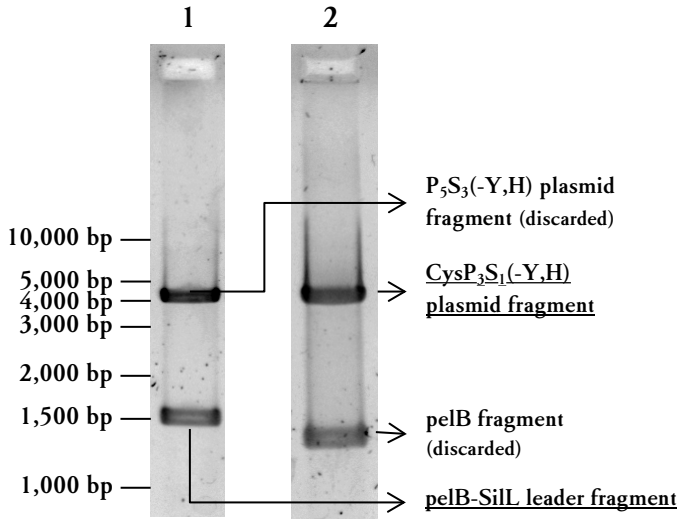

Figure 5. Double digestion with *Eco*RV/*Eco*RI for subcloning of the pelB-SilL double leader to CysP_3S_1(-Y,H). Digestion was performed for pelB-SilL-P_5S_3(-Y,H) (*lane 1*) and pelB-CysP_3S_1(-Y,H) (*lane 2*), both on pET 22b(+), and fragments were analyzed by agarose gel electrophoresis. The 1541 basepair fragment with the pelB-SilL double leader was later inserted into the CysP_3S_1(-Y,H) plasmid fragment (above 4000 basepairs).

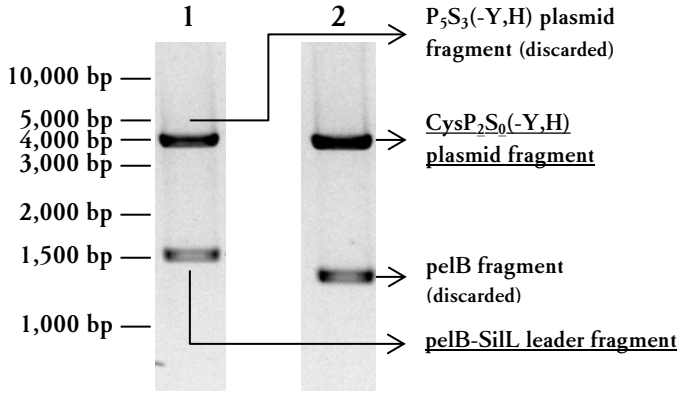

Figure 6. Double digestion with *Eco*RV/*Nco*I for subcloning of the pelB-SilL double leader to CysP_2S_0(-Y,H). Digestion was performed for pelB-SilL-P_5S_3(-Y,H) (*lane 1*) and pelB-CysP_2S_0(-Y,H) (*lane 2*), both on pET 22b(+), and fragments were analyzed by agarose gel electrophoresis. The 1541 basepair fragment with pelB-SilL double leader was later inserted into the CysP_2S_0(-Y,H) plasmid fragment (above 4000 basepair). This was done by J. Wiederstein and is an excerpt from her thesis [89].

Sub-cloning of P_xS_y into pET 31b(+) to create a fusion with ketosteroid isomerase

Besides the pelB-SilL double-leader versions of the P_xS_y-target genes on pET 22b(+), another plasmid, pET 31b(+), was intended as expression plasmid, hosting a ketosteroid isomerase (KSI) gene as fusion. This plasmid contains a single *Ava*I restriction site downstream of the KSI fusion sequence (5'-C|TCGAG-3'), which itself is followed by a His-tag coding sequence. *Ava*I has degenerate recognition sites (sequence 5'-C|YCGRG-3' where R is either A or G and Y is either T or C). In pelB-SilL-P_xS_y(-Y,H) fusions, two *Ava*I sites were available for excision of the P_xS_y region. One was located between the anionic SilL sequence and the HRV3C site (5'-C|CCGAG-3'), the second one between the actual P_xS_y sequence and the hexahistidyl-tag (5'-C|TCGAG-3'). This means that for ligation of the *Ava*I-excised P_xS_y sequence into the *Ava*I-digested pET 31b(+), the corresponding ends had a one-base mismatch in the 5' region of the P_xS_y gene. For subcloning of P_5S_3(-Y,H) into the pET 31b(+) plasmid, the region coding for the complete fusion gene was initially PCR-amplified to achieve a sample enriched with the gene, since large amounts of plasmid would have had to be digested for a sufficient amount of the insert segment (the *Ava*I excised sequence (186 bp) only makes up 3 % of the 5811 bp plasmid). PCR with primers T7PrompET22b(+)fw and pET22b(+)rev on plasmid pET 22b(+) hosting the pelB-SilL-P_5S_3(-Y,H) fusion produced a 617 bp PCR-amplificate (Figure 7). Upon *Ava*I digest, this amplificate resulted in 3 restriction fragments: the target fragment containing the P_5S_3(-Y) gene (186 bp) as well as two fragments which were cut off upstream (347 bp) and downstream (84 bp) of the P_5S_3 gene, respectively. Given the size differences, the target fragment bearing the P_5S_3(-Y) gene was easily identified in agarose gel electrophoresis. In P_3S_1(-Y), where the corresponding fragment length is only 120 bp, separation was still possible, though more difficult. The CysP_2S_0(-Y)-fragment is only 87 bp in size and can not be discriminated from the 84 bp cleavage fragment 3' of the fragment of interest.

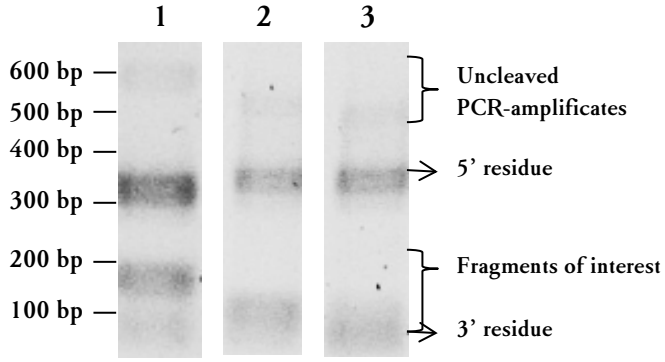

Figure 7. *Ava*I digestion of PCR-amplificates of different P_xS_y-version. The fragments were separated by agarose gel electrophoresis for insertion into plasmid pET 31b(+). Digestion with *Ava*I resulted in each three fragments. Desired fragments were 186 (P_5S_3(-Y), *lane 1*), 120 (P_3S_1(-Y), *lane 2*) and 87 (CysP_2S_0(-Y), *lane 3*) basepair in length. This work was done by J. Wiederstein and is an excerpt from her thesis [89].

The desired fragments were ligated with *Ava*I digested pET 31b(+). After transformation of bacteria and overnight growth, transformants bearing the new insert in right orientation were identified by colony PCR and restriction analysis of the amplificates (Figure 8). Colony PCR was accomplished using primers T7PrompET22b(+)fw and pET22b(+)rev, which amplify the complete fusion genes KSI-P_xS_y. Expected amplificate sizes were 740 bp (P_5S_3(-Y,H)), 674 bp (P_3S_1(-Y,H)) and 641 bp ($CysP_2S_0$(-Y,H)). Digestion with *Eco*RI, for which a cleavage site was located between the HRV3C-site and P_xS_y coding region, was used for analysis of the orientation of the inserts. In the correct orientation, fragments of 475 and 265 bp (P_5S_3(-Y,H)), 475 and 199 bp (P_3S_1(-Y,H)) or 475 and 166 bp ($CysP_2S_0$(-Y,H)), respectively, were expected. In case of opposite orientation, the *Eco*RI cleavage cut off only a 50 bp fragment. Positive transformants were identified for P_5S_3(-Y,H) and P_3S_1(-Y,H). To correct for the mismatch of the degenerated *Ava*I restriction sites (which would after the first replication cycle result in two plasmids differing in one basepair), the new plasmids were isolated from their host and subjected to a second transformation round to ensure a homogenous composition of the plasmid in the transformant. Afterwards, sequences were confirmed by sequencing (Figure 9 for P_5S_3(-Y,H)).

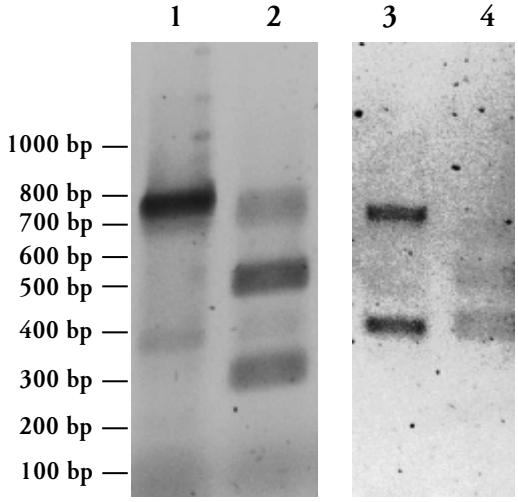

Figure 8. Colony PCR and *Eco*RI restriction analysis for P_xS_y(-Y,H) in pET 31b(+). Shown are positive results for P_5S_3(-Y,H) (*left*) and P_3S_1(-Y,H) (*right*). The PCR-amplifications (*lanes 1 and 3*) and respective *Eco*RI digestions (*lanes 2 and 4*) were analyzed by agarose gel electrophoresis. This analysis was done by J. Wiederstein and is an excerpt from her thesis [89].

```
>KSI
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10        20        30        40        50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70        80        90        100       110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130       140       150       160       170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190       200       210       220       230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250       260       270       280       290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310       320       330       340       350
                                   >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370       380       390       400       410
   >P5S3(-Y,H)
ATGGGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA  < 480
M  G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         430       440       450       460       470
GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCT  < 540
G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A
         490       500       510       520       530
                                   >His6
AGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 594
S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         550       560       570       580       590
```

Figure 9) Sequence of the KSI-P_5S_3(-Y,H) fusion. This construct contains a ketosteroid isomerase fusion, followed by HRV3C protease site and hexahistidyl-tagged P_5S_3(-Y,H). The cloning was done by J. Wiederstein and is an excerpt from her thesis [89].

Changing the P_xS_y sequences by single site mutagenesis

The P_xS_y expression plasmids were further subjected to different mutagenesis reactions, by single site mutagenesis with complementary primer hybrids bearing mismatches with respect to the parental genes, thereby inserting or exchanging single bases or complete codons. All mutageneses were verified by sequencing, and sequences are shown in the appendix. Five different sequence changes were introduced by single site mutagenesis (compare to the pelB-SilL-P_5S_3(-Y,H) sequence shown in Figure 4). The early versions of P_xS_y contained a cysteine as

second amino acid after the P_xS_y methionine, which was replaced by serine or conversely, cysteines were later re-introduced at that position. For UV/vis quantification, a tyrosine was introduced as third amino acid in P_xS_y (resulting in sequence "…MGS**Y**S…"). Another mutagenesis was done to create versions without the hexahistidyl-tag by exchanging a glutamate-codon prior to the tag for a stop-codon. In the pelB-SilL-P_xS_y, a mutagenesis in the pelB-leader exchanged the lysine immediately after the methionine start-codon for glutamate, thereby reversing the charge in this position in order to affect the expression of the fusion peptide (see below for expression analyzes).

Evaluation of culture conditions for expression of charge-compensated P_xS_y-fusions

To achieve expression of P_xS_y as a charge compensated fusion, both the pelB, and the SilL anionic leader were necessary. To optimize the expression conditions, analyzes were done with pelB-SilL-P_5S_3(-Y,H). Expression was evaluated for IPTG-induction at different cell densities (OD_{600} around 0.5 or 1.0 as well as overnight cultures after 18 hours) and for different temperatures in the expression phase (37 and 20 °C to affect the solubility of the recombinant proteins [102]). The expression success was analyzed by separating bacterial total cell proteins in polyacrylamide gels with subsequent coomassie-staining, or alternatively western blotting using a primary antibody raised against the hexahistidyl tag.

In coomassie stained gels (Figure 10), no clear proof for the expression of pelB-SilL-P_5S_3(-Y,H) was found. IPTG-induction did not lead to a specific increase of a particular protein band, except for a protein band around 24 kDa which occurred after prolonged expression periods at 37 °C. However, immunostaining in western blot (Figure 11) with an anti hexahistidine-tag antibody revealed a polypeptide with an apparent molecular weight around 35 kDa. This was higher than the molecular weight expected from the sequence of the fusion polypeptide, which is 16.5 kDa. However, aberrant molecular weights in SDS-PAGE are not uncommon. The migration of proteins in SDS-PAGE depends on the polypeptide charge, SDS-binding capability of particular amino acid side-chains and residual conformations in the presence of SDS, and single amino acid exchanges can have huge impact on the apparent molecular weight [103-107]. It is hence not unexpected that pelB-SilL-P_5S_3(-Y,H), being composed of two high-charged domains, migrates uncommon with respect to a vast amount of naturally occurring proteins.

The expression of pelB-SilL-P_5S_3(-Y,H) worked well at 20 °C for 5 to 8 hours (Figure 11). Interestingly, when the expression duration was extended to more than 20 hours, the polypeptide was only found in cultures which were induced at high cell density.

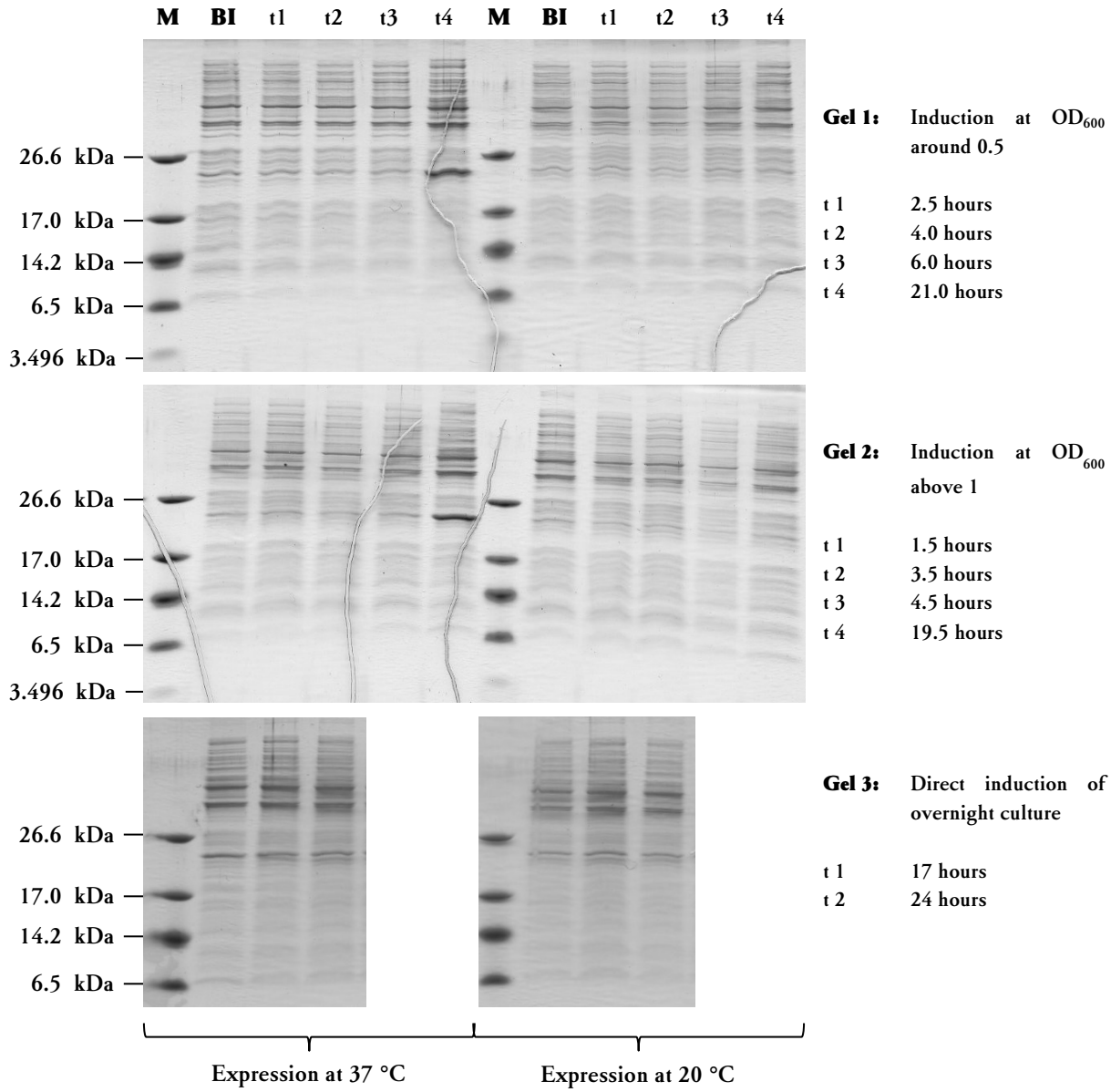

Figure 10. SDS-PAGE analysis of the pelB-SilL-P_5S_3(-Y,H) recombinant expression. Expression was induced by IPTG addition at different growth stages of the culture: OD_{600} around 0.5 (exponential growth phase), larger 1 (early stationary phase) or directly an overnight culture (late stationary phase). Both expression at 37 °C and 20 °C was evaluated. Total protein samples were collected before IPTG-induction (*lanes BI*) and after different periods of expression (*lanes t1 - t4*). The time points of collection are rounded to 0.5 hour steps. The marker (*lane M*) was ultra-low range molecular weight marker (Sigma Aldrich).

Since the expression yield was comparably low per cell (the fusion polypeptide was only revealed in western blot) the induction at high cell densities was applied as standard condition. Best yields were achieved by transferring the bacteria of an overnight culture into an equal volume of fresh medium and perform expression at 20 or 37 °C overnight (Figure 12). The expression at 20 °C was chosen as standard condition and worked in a reliable fashion.

Additionally, glucose (0.1 % w/v) was added to the media, as described in the original recipe for LB [90] medium, to further reduce the level of undesired expression before IPTG-induction [108]. With this protocol, the pelB-SilL-P_xS_y polypeptides were successfully expressed, as seen from the purification results in Figure 13 and later.

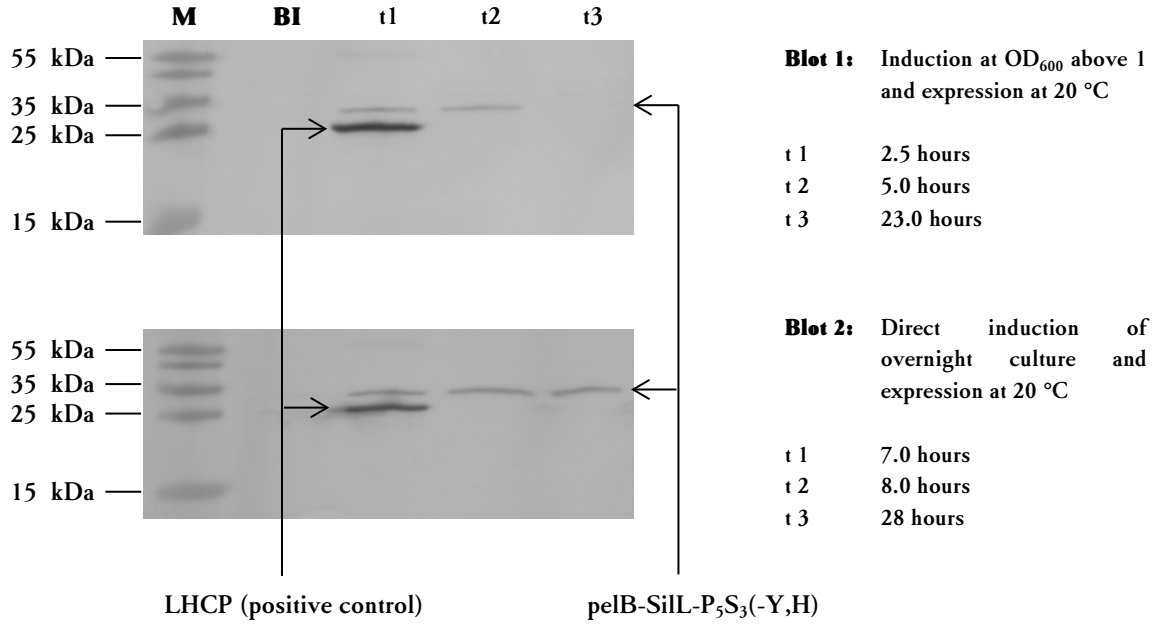

Figure 11. SDS-PAGE analysis with western blotting for the pelB-SilL-P_5S_3(-Y,H) recombinant expression. Expression was induced by IPTG at OD_{600} larger 1 (early stationary phase, *blot 1*) or directly in an overnight culture (late stationary phase, *blot 2*). Expression temperature was 20 °C. A purified, recombinant, his-tagged LHCP (light harvesting complex protein) was added to each t = 1 sample as positive control for the western blot. Total protein samples before induction with IPTG (*lane BI*) as well as at different time points after induction (*lanes t1 - t3*) were analyzed. PageRuler Prestained Protein ladder (Thermo Scientific, *lane M*) was added as molecular weight standard.

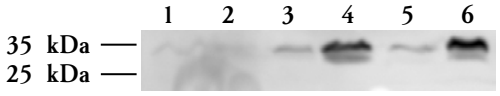

Figure 12. Expression of pelB-SilL-P_5S_3(-Y,H) after induction at high cell density. *Lanes 1 and 2*: An overnight culture at 37 °C was immediately induced and expression was at 37 °C. *Lanes 3 and 4*: Cells from an overnight culture were transferred into an equal volume of fresh medium and allowed to adapt for 1.5 h at 37 °C before induction, expression was at 37 °C. *Lanes 5 and 6*: Cells from an overnight culture were transferred into an equal volume of fresh medium and allowed to adapt for 1.5 h at 37 °C after which cultures were cooled below 20 °C before induction, expression was at 20 °C. For each condition, total protein samples before induction (*lanes 1, 3, 5*) and after an overnight expression phase (*lanes 2, 4, 6*) were compared by western blotting.

Purification of hexahistidine-tagged P_xS_y-versions by a two-step IMAC/SEC approach

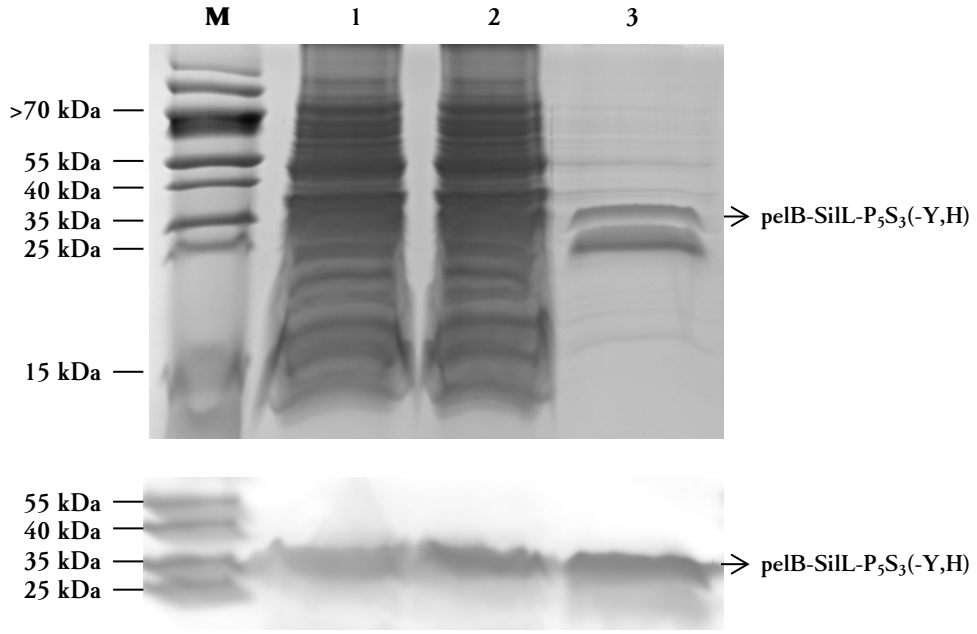

Figure 13. SDS-PAGE analysis of the immobilized metal affinity chromatoraphy (IMAC) purification of pelB-SilL-P_5S_3(-Y,H). Proteins were stained with coomassie (*upper figure*) or in western blot (*lower figure*). Analyzed were the SDS-dissolved protein extract (*lane 1*), the combined flow through and washing fraction (*lane 2*), as well as the elution fraction (*lane 3*). As molecular weight marker (*lane M*), PageRuler Prestained Protein Ladder (Thermo Scientific) was used.

Purification of the fusion protein pelB-SilL-P_5S_3(-Y,H) worked well by immobilized metal affinity chromatography (IMAC) (Figure 13). Cells were first lysed with lysozyme and benzonase nuclease. After centrifugation, the target fusion protein was found in the pelleted fraction, together with many bacterial impurities. This complex mixture was solubilized in SDS (1 % w/v) solution and, after tenfold dilution, subjected to metal affinity chromatography. To wash away unbound proteins, the resin was first washed with SDS (0.1 % w/v) containing buffer, and the SDS was subsequently removed by washing with 8 mol/L urea. In a third washing step, a salt buffer was applied to eventually remove urea. Elution was subsequently achieved with an elution buffer containing both high concentrations if imidazole (0.5 mol/L) and sodium chloride (1 mol/L). Although some target protein was not bound to the resin, a considerable amount was isolated, as revealed by SDS-PAGE analysis with both coomassie and immune staining. Besides the fusion protein (apparent molecular weight around 35 kDa), a protein around 25 kDa was co-purified in large amounts, whereas further impurities were strongly reduced after this chromatographic step. In a parallel assay with the pelB-free version SilL-P_5S_3(-Y,H) (Figure 14), no target protein was found expressed, as revealed by coomassie staining and western blot analysis. Interestingly, the protein with apparent molecular weight around 25 kDa, already seen in the elution fraction for pelB-SilL-P_5S_3(-Y,H), was enriched in the elution fraction again. A number of proteins from *Escherichia coli* with different molecular weights is known to become

bound to IMAC resins loaded with nickel [109]. These in majority are stress-response proteins. The interaction with the resin may be because of surface-exposed histidine-clusters (e.g. peptidoylproline *cis-trans* isomerase, SlyD, 20.8 kDa) or other interacting domains, such as specific binding motifs in natively metal-binding proteins (e.g. metal-binding lipocalin, YodA, 22.3 kDa). Their abundance as impurity in IMAC usually depends on the culture conditions (i.e. their synthesis rate and relative abundance in the bacterial cell) and they can be removed by washing the resin with increasing concentrations of imidazole, though some are not removed below 80 mmol/L imidazole. For the purification of pelB-SilL-P_xS_y(-Y,H) no further attempts to identify or remove this impurity during IMAC were made, since the impurity was removed in the following processing (see below).

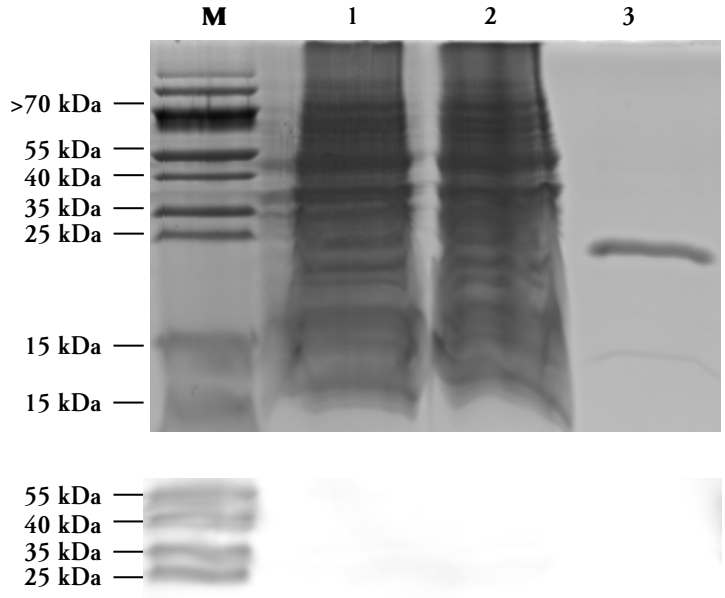

Figure 14. SDS-PAGE analysis of the immobilized metal affinity chromatoraphy (IMAC) purification attempt for SilL-P_5S_3(-Y,H). Proteins were stained with coomassie (*upper figure*) or in western blot (*lower figure*). Analyzed were the SDS-dissolved protein extract (*lane 1*), the combined flow through and washing fraction (*lane 2*), as well as the elution fraction (*lane 3*). As molecular weight marker (*lane M*), PageRuler Prestained Protein Ladder (Thermo Scientific) was used.

A SDS-PAGE analysis of the complete procedure of P_5S_3(-Y,H) purification from pelB-SilL-P_5S_3(-Y,H) is shown in Figure 15. The IMAC-purified fusion was efficiently cleaved by HRV3C protease in the IMAC elution buffer (0.5 mol/L imidazole, 1 mol/L sodium chloride, 20 mmol/L Tris-HCl, pH 7.5), in which the high salt-concentration is expected to be beneficial for enhancing the protease activity [92]. After proteolytic digest of the fusion, a protein band with apparent molecular weight of 15 kDa appeared (theoretical molecular weight of P_5S_3(-Y,H) is 6.7 kDa). This protein band was purified by ion exchange chromatography on sulphopropyl (SP) sepharose by washing the resin after polypeptide-binding with 200 mmol/L ammonium carbonate and eluting with 500 mmol/L ammonium carbonate. It is concluded that the polycationic P_5S_3(-Y,H) target has a notably decreased electrophoretic mobility because of its high net cationic charge,

which may additionally affect the SDS-binding capacity and hence net charge during electrophoretic separation [103-107]. The purification method was also applied for the tyrosine-bearing P_5S_3(H) and the cysteine-bearing CysP_5S_3(-Y,H). Analyzes by SDS-PAGE were similar to the purification of P_5S_3(-Y,H) shown in Figure 15 and are hence not shown in addition.

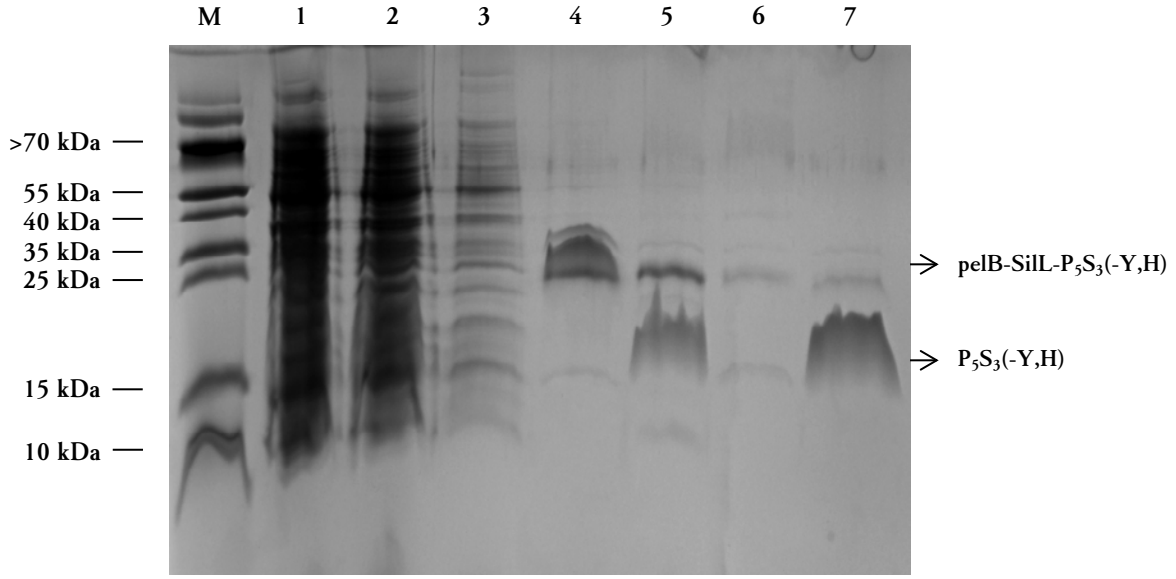

Figure 15. SDS-PAGE analysis of the purification of pelB-SilL-P_5S_3(-Y,H). The different lanes contained PageRuler Prestained Protein ladder (*lane M*), SDS-dissolved bacterial extract (*lane 1*), IMAC flow-through (*2*), IMAC washing fraction (*3*), IMAC elution fraction (*4*), HRV3C digestion (*5*), combined IEC flow through and washing fraction (*6*) and IEC elution fraction (*7*).

The purification procedure was also successful for isolation of the shorter versions of P_xS_y. The SDS-PAGE analysis of the P_3S_1 purification is shown for P_3S_1(-Y,H) (Figure 16) and was similar for the tyrosine-bearing P_3S_1(H) (not shown). The P_2S_0 purification was only assessed for the cysteine-bearing CysP_2S_0(-Y,H) (Figure 17). As with the P_5S_3 version, the fusion proteins were not visible in the total protein fraction after SDS-PAGE and coomassie staining, pointing to a low expression. However, specific bands appeared after IMAC purification. In case of the P_3S_1 fusion protein, again a protein band around 35 kDa (theoretical molecular weight is 14.0 kDa) was prominent and disappeared upon HRV3C-digest, by which a new band of P_3S_1(-Y,H) between 10 and 15 kDa (theoretical: 4.2 kDa) appeared. In the IMAC-elution containing the P_2S_0 fusion, two prominent bands at 25 kDa and slightly above 15 kDa (the fusion is theoretically 12.8 kDa) were visible, both again disappearing upon HRV3C cleavage, after which a new protein band below 10 kDa (CysP_2S_0(-Y,H) theoretically 3.0 kDa) appeared.

Different reasons for the running behavior of pelB-SilL-CysP_2S_0(-Y,H) are possible. It might be an indication that this fusion, still with an uncommon migration because of the charged domains, can migrate e.g. both as monomer, and as dimer in SDS-PAGE. This possibility was demonstrated for hydrophobic peptides [110,111] as well as an oligomeric β-glycosidase from which dimers resisted to up to 1 % w/v SDS, presumed to be mediated by hydrophilic, charged

patches which were surrounded by a domain with preference for anionic, SDS-repulsing amino acids [112]. The dimerization could also apply for the longer P_xS_y versions since their respective fusions have higher apparent molecular weight, though the dimerization may be reduced in P_2S_0 because of the shortened cationic domain. It would also be reasonable that the P_2S_0-fusion is in part exported to the periplasm, guided by the pelB-leader [78], where the leader may be cleaved off [113-116]. A series of silaffin-like polypeptides was expressed as fusion with pelB in an earlier diploma-thesis in this working group (Professor Harald Paulsen) and fractionated lysis of the bacteria did indicate that these were not exported [80], presumably because of their cationic domain which is expected to stall the membrane translocation [117]. However, it is possible that P_2S_0 is sufficiently short to not fully prevent the translocation. Although the cleavage of the leader would only decrease the molecular weight of the fusion by 2.2 kDa, the removal of this mostly hydrophobic sequence might have a strong impact on the apparent molecular weight in SDS-PAGE [103-107]. Despite these possible explanations, the electrophoretic mobility of P_2S_0-fusions was not further investigated since the double-band only appeared for the fusion polypeptide, and not after liberation of P_2S_0 by HRV3C proteolytic processing.

The HRC3C-liberated versions of P_3S_1 and P_2S_0 were both further purified by IEC. The method for P_5S_3 was successfully applied for P_3S_1, with the same concentrations of ammonium carbonate for washing of and elution from the resin (200 and 500 mmol/L, respectively). For P_2S_0 purification, the salt concentration of the washing fraction was lowered to 100 mmol/L. As seen in Figure 17 (lane 7), some P2S0 was already lost at this concentration, whereas neither for P_3S_1, nor P_5S_3 losses were substantial at up to 200 mmol/L (Figures 16 and 15).

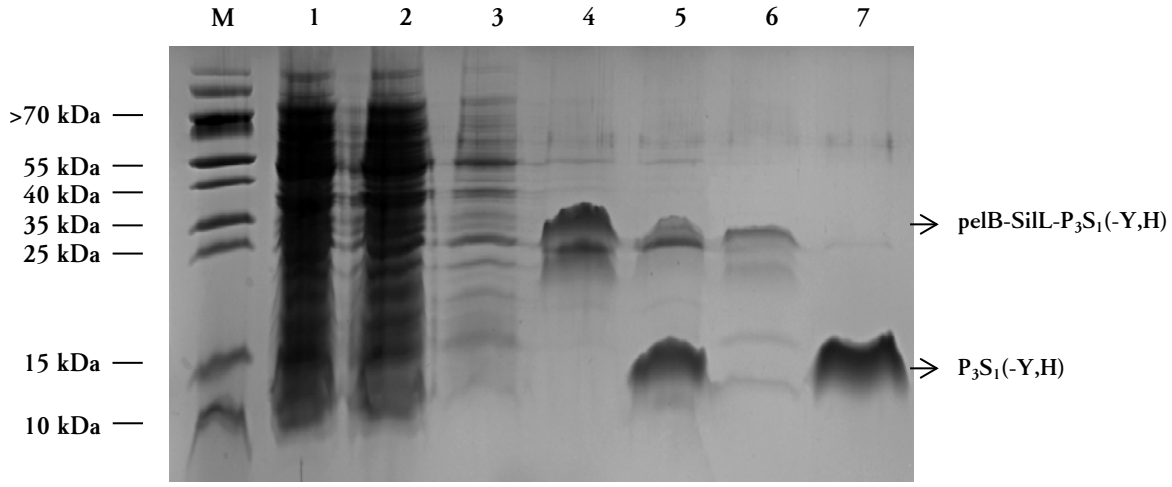

Figure 16. SDS-PAGE analysis of the purification of pelB-SilL-P_3S_1(-Y,H). The different lanes show PageRuler Prestained Protein ladder (*lane M*), SDS-dissolved bacterial extract (*lane 1*), IMAC flow-through (*2*), IMAC washing fraction (*3*), IMAC elution fraction (*4*), HRV3C digestion (*5*), combined IEC flow through and washing fraction (*6*) and IEC elution fraction (*7*). Results were principally similar for the tyrosine-containing version P_3S_1(H).

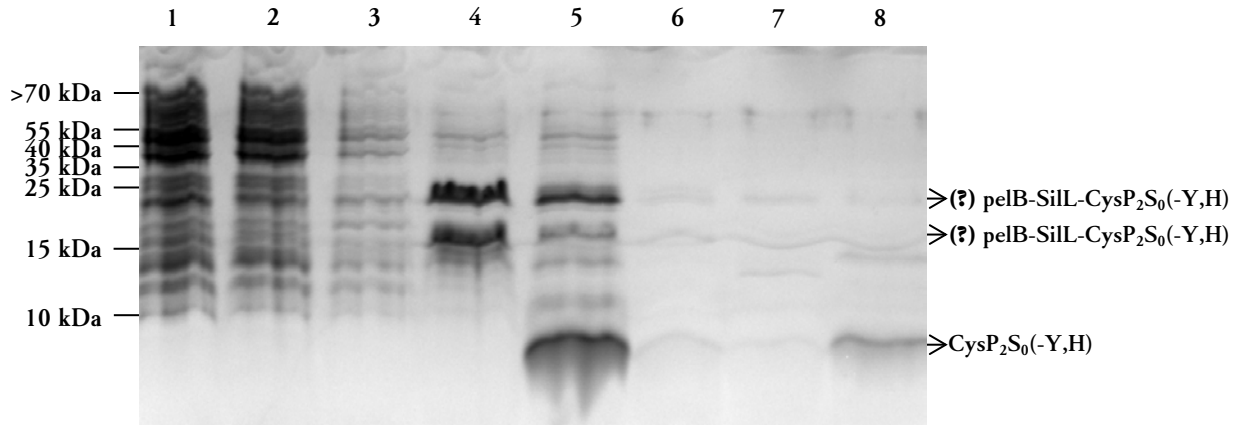

Figure 17. SDS-PAGE analysis of the purification of pelB-SilL-$CysP_2S_0$(-Y,H). The different lanes show SDS-dissolved bacterial extract (*lane 1*), IMAC flow-through (*2*), IMAC washing fraction (*3*), IMAC elution fraction (*4*), HRV3C digestion (*5*), IEC flow through (*6*) and washing fraction (*7*) as well as IEC elution fraction (*8*).Washing of the IEC resin was done with 100 mmol/L ammoniumcarbonate only, to avoid undesired elution of P_2S_0 polypeptide.

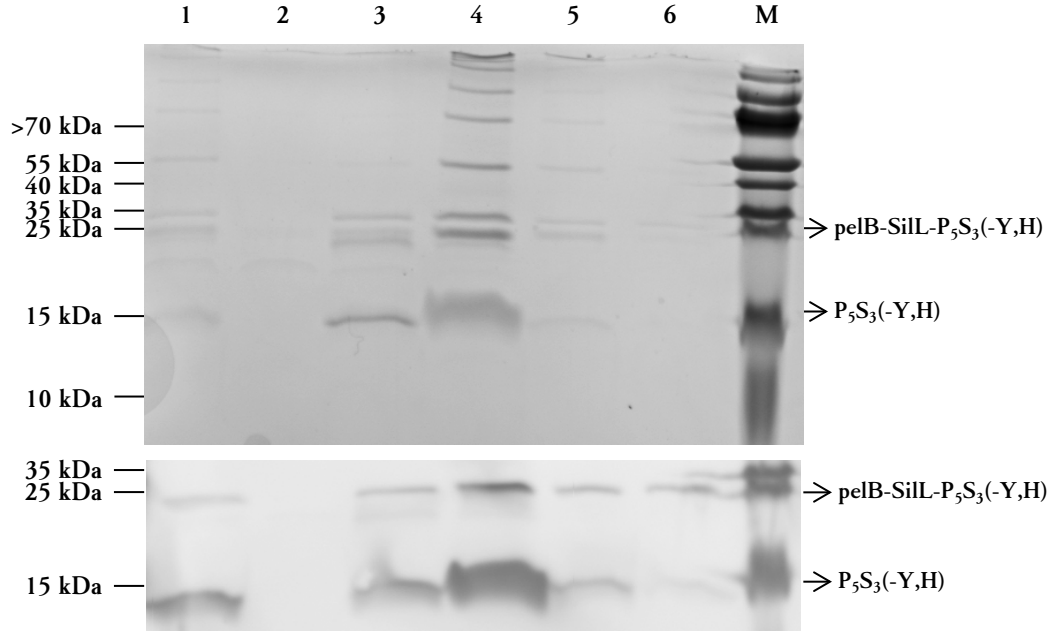

Figure 18. IEC elution profile of P_5S_3(-Y,H) and the corresponding fusion pelB-SilL-P_5S_3(-Y,H) in case of sodium hydroxide elution. Results were obtained in an assay where HRV3C-digest was incomplete. Fractions were both analyzed by SDS-PAGE with coomassie stain (*upper gel figure*) as well as after western blot and immune staining (*lower blot figure*). Lanes represent HRV3C digestion (*lane 1*), IEC flow through (*2*), washing fraction sodium chloride/Tris-HCl pH 7.5 (*3*), and three elution steps (4 - 6) at 20, 20 and 50 mmol/L sodium hydroxide, respctively, PageRuler Prestained Protein Ladder (Thermo Scientific, *M*).

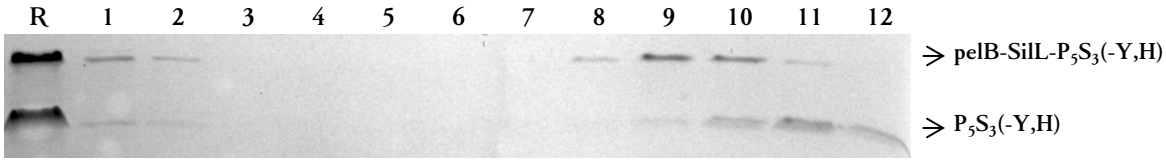

Figure 19. IEC elution of P_5S_3(-Y,H) and the corresponding fusion pelB-SilL-P_5S_3(-Y,H) in case of ammonium carbonate elution. Results were obtained in an assay where HRV3C-digest was incomplete, and fractions were analysed by SDS-PAGE with coomassie staining. Samples were: reference sample before IEC (*lane R*), flow through IEC (*1*), elution fraction 10 mmol/L ammonium carbonate (*2*) and elutions at 50 to 500 mmol/L ammonium carbonate, in steps of 50 mmol/L (*lanes 3 - 12*).

Interestingly, in case of incomplete protease digestion with HRV3C, both the liberated cationic target P_5S_3(-Y,H), and its corresponding fusion pelB-SilL-P_5S_3(-Y,H) bound to the cation exchange resin (Figures 18 and 19). Elution with either sodium hydroxide or ammonium carbonate did not allow complete discrimination in the elution fractions. This indicated that in the fusion, the cationic target P_xS_y(-Y,H) is sufficiently exposed to mediate the binding to the ion exchange resin in a similar, though reduced strength with respect to the target peptide after cleavage of the fusion.

Modifying the pelB-leader sequence in order to prevent periplasmic transport initiation

In order to enhance the expression rate for the charge-compensated fusion pelB-SilL-P_xS_y(-Y,H). the pelB-leader sequence was modified to disfavor the periplasmic transport initiation. In particular, a lysine (second amino acid in pelB) was exchanged for glutamate, to reverse the charge and inhibit the initiation of the pelB-guided translocation of the recombinant protein into the intermembrane space of *Escherichia coli* [118]. The expression of P_5S_3(-Y,H) as the modified fusion pelB*(X)*-SilL-P_5S_3(-Y,H) ("*(X)*" marks the manipulated leader) was evaluated. At an expression temperature of 37 °C, no remarkable differences in SDS-PAGE analysis with coomassie as well as immune staining were noticed (Figure 20 and 21) when compared to the version with the native leader (Figure 10 and 11), where similar samples with respect to the cell density were applied to SDS-PAGE (see materials and methods, equal volumes of total cell samples at OD600 = 10 were loaded). The western blot analysis (Figure 21) revealed that virtually no protein is expressed in the pre-cultures until OD_{600} up to 1. In the overnight uninduced pre-culture, some staining in the blot became visible, though it is not clear whether this indicates expression of the target protein or whether this is an artifact. A reliable signal of the target polypeptide was found for several hours post expression, but disappeared after overnight incubation at 37 °C, as already seen for the expression of pelB-SilL-P_5S_3(-Y,H) (Figure 11 and 12). The expression pattern was not further investigated since it is obvious that the mutation in the pelB-leader did not lead to a substantial increase of the P_5S_3(-Y,H) expression yield.

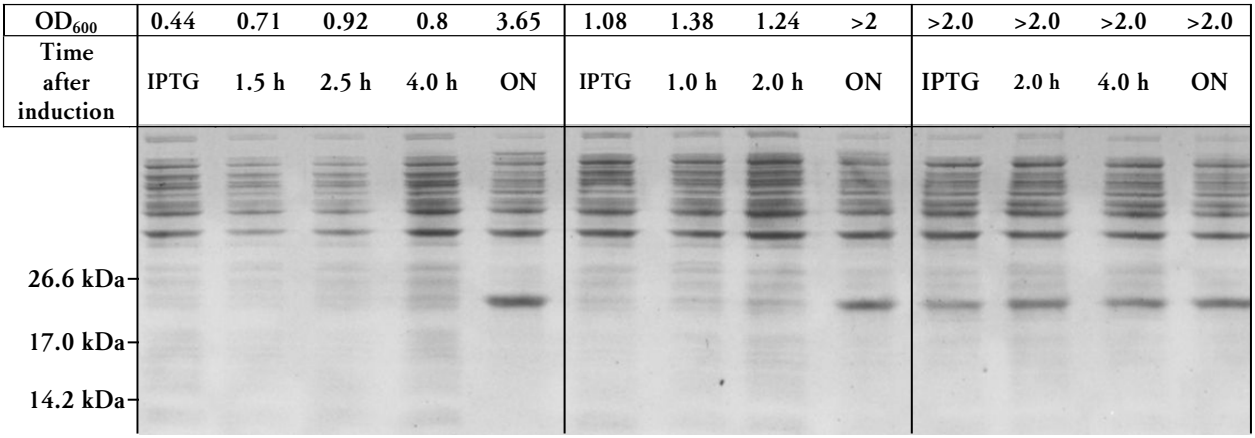

Figure 20. SDS-PAGE expression analysis for pelB*(X)*-SilL-P_5S_3(-Y,H) in LB medium. The analysis shows total protein samples collected during the course of culture growth and protein expression at 37 °C. To examine the effect of different cell densities at time of induction (marked in table as "IPTG"), cultures were either induced at controlled densities (*left and middle column*), or an overnight culture was directly induced (*right colum*n). For induction at controlled cell densities, bacteria from overnight LB pre-cultures were harvested by centrifugation and transferred to fresh LB-medium at 50-fold dilution with respect to the pre-culture volume. OD_{600} was used as an estimation for cell-density and induction was around OD_{600} of 0.5 (*left column*) or 1.0 (*middle column*). At different time-points after induction (hours as indicated, "ON" is overnight expression), samples were collected. This analysis was done by L. Krebs and is an excerpt from her thesis [101].

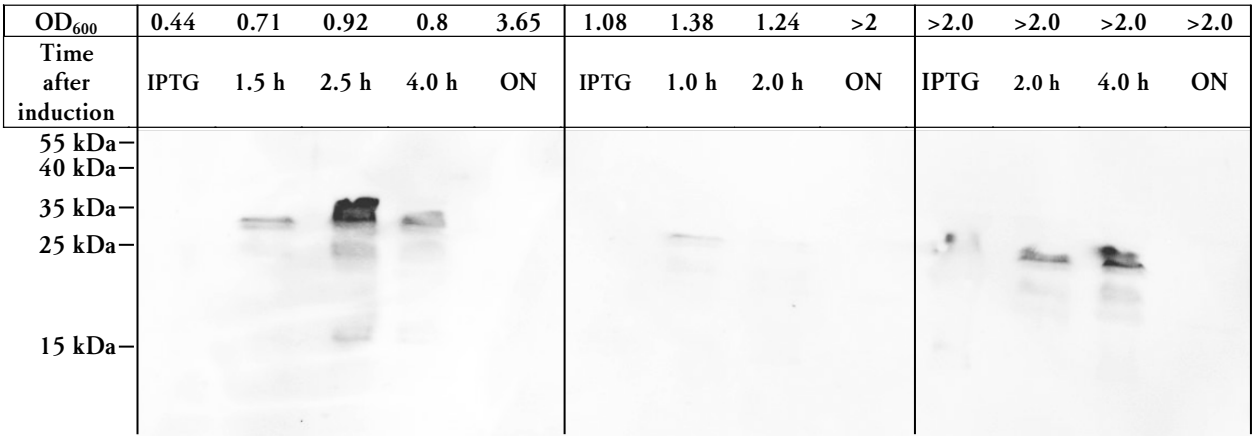

Figure 21. Western blot for the expression analysis of pelB*(X)*-SilL-P_5S_3(-Y,H). Samples are the same as in Figure 20 in a parallel SDS-PAGE analysis. This analysis was done by L. Krebs and is an excerpt from her thesis [101].

Evaluation of ketosteroid isomerase as hydrophobic fusion for P_xS_y expression

The highly insoluble ketosteroid isomerase (KSI) [58,82] was examined as fusion for the expression of P_xS_y in inclusion bodies. The fusion of KSI with P_5S_3(-Y,H) was expressed well when the bacteria were grown to an OD_{600} of 1 before induction of protein expression by IPTG-addition. A strong protein band around 25 kDa became visible after an overnight expression

phase in coomassie-stained SDS-PAGE (Figure 22, left). This protein was identified as the target protein by western blotting with an antibody raised against the hexahistidin tag (Figure 22, right), and this identification was later confirmed by purification of P_5S_3, expressed as KSI-fusion, and subsequent amino acid analysis (section III and published elsewhere [77]). The western blot also revealed that the protein was already expressed before induction to a considerable amount, though after the overnight expression phase it was more enriched with respect to the cellular proteins (coomassie-gel and western blot).

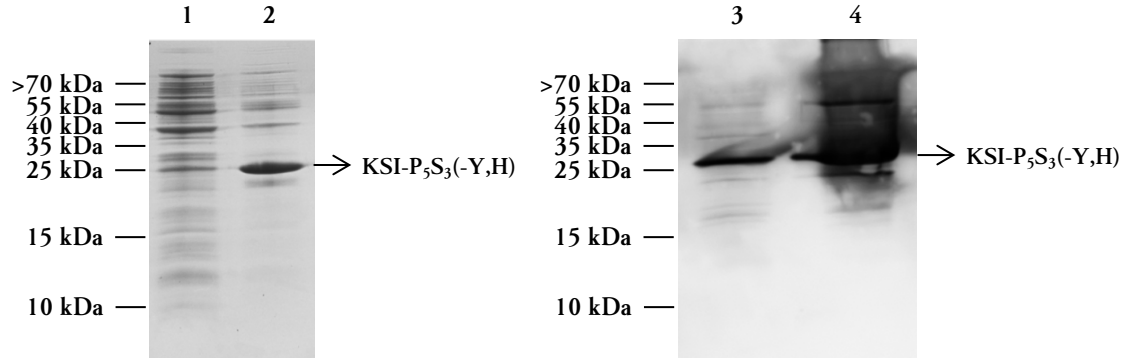

Figure 22. SDS-PAGE expression analysis for KSI-P_5S_3(-Y,H) in LB medium. Cells from overnight LB pre-cultures were harvested by centrifugation and transferred to fresh LB-medium at 50-fold dilution with respect to the pre-culture volume. OD_{600} was used as an estimation for cell-density and induction was around OD_{600} of 1.0. Expression was allowed overnight. Total protein samples of cells before induction (*lanes 1 and 3*) and after expression (*lanes 2 and 4*) were analyzed by coomassie staining (*left*) and western blot (*right*). This analysis was done by J. Wiederstein and is an excerpt from her thesis [89].

The course of the expression was again analyzed in terms of time of induction, optimal expression period length, and the effect of glucose in suppressing protein expression before deliberate IPTG-induction. Pre-cultures (overnight, 37 °C) were 50-fold diluted in fresh LB-medium and induced at different cell densities (OD_{600}). After induction, expression was done at 37 °C. In LB media without glucose (Figure 23), the protein band at 25 kDa of KSI-P_5S_3(-Y,H) was strong in coomassie-stained SDS-PAGE in all samples (induction at OD_{600} = 0.5, 1, and > 2). Adding glucose to the pre- and main-cultures (Figure 24) reduced the undesired expression of the target protein before IPTG-induction. This prevention of leaky gene expression by glucose is well-known in particular for plasmids from the pET series [108]. To hence control the expression of KSI-P_5S_3(-Y,H) and its mutagenized derivatives, glucose was added to all media to 0.1 % w/v, as described in the original recipe of the LB medium [90]. The overexpression of KSI-P_5S_3(-Y,H) was strong irrespective of the time point of induction (i.e. different growth phases) (Figure 23 and 24). However, taking into account the OD_{600} values (where determined), the total amount of target polypeptide was higher when cultures were induced at high cell densities. As seen in section III and reference [77], purification of P_5S_3 resulted in highest yields under these conditions.

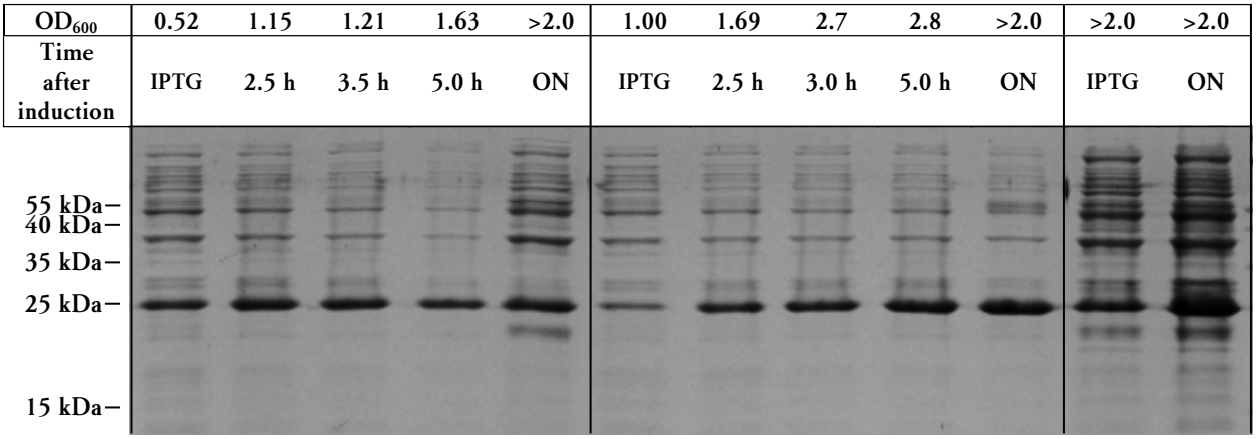

Figure 23. SDS-PAGE expression analysis for KSI-P_5S_3(-Y,H) in LB medium **without glucose**. The analysis shows total protein samples collected during the course of culture growth and protein expression. To examine the effect of different cell densities at time of induction (marked in table as "IPTG"), cultures were either induced at controlled densities (*left and middle column*), or an overnight culture was directly induced (*right column*). For induction at controlled cell densities, cells from overnight LB pre-cultures were harvested by centrifugation and transferred to fresh LB-medium at 50-fold dilution with respect to the pre-culture volume. OD_{600} was used as an estimation for cell-density and induction was around OD_{600} of 0.5 (*left column*) or 1.0 (*middle column*). At different time-points after induction (hours as indicated, "ON" is overnight expression), samples were collected. This analysis was done by J. Wiederstein and is an excerpt from her thesis [89].

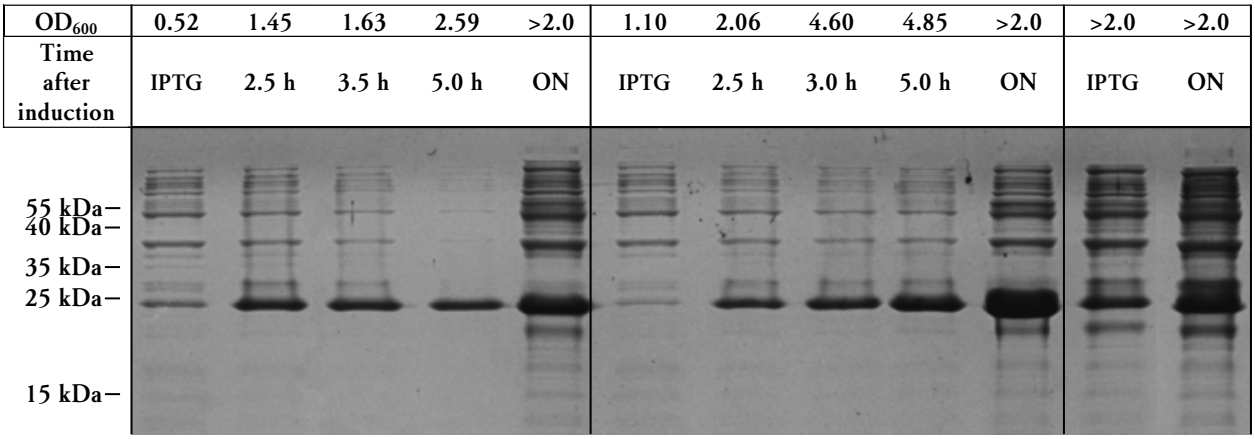

Figure 24. SDS-PAGE expression analysis for KSI-P_5S_3(-Y,H) in LB medium **with glucose**. Explanations for different samples and abbreviations are equal to those in Figure 23. This analysis was done by J. Wiederstein and is an excerpt from her thesis [89].

Preparation of charge-compensated P_xS_y without a hexahistidine tag

Purification of tag-free versions of P_xS_y was also evaluated for the charge-compensated fusion pelB-SilL-P_xS_y(-Y). In contrast to the earlier described purification with a two-step IMAC and IEC procedure, a double ion exchange chromatography was used for purification of first the fusion and second the liberated target polypeptide P_xS_y(-Y). The binding of both the fusion, and

the polycationic target to the IEC-resin was already proven from the analysis presented in Figures 18 and 19.

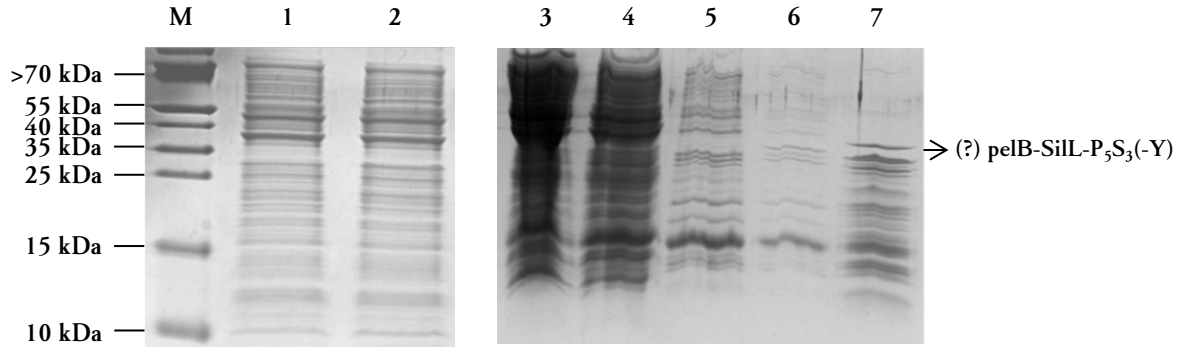

Figure 25. SDS-PAGE analysis for expression of pelB-SilL-P_5S_3(-Y) and purification from urea-dissolved bacterial extract. The lanes contained: PageRuler Prestained Protein Ladder (Thermo Scientific, *lane M*), total protein before induction (*lane 1*) and after expression (*2*), flow through IEC (*3*), IEC washing fractions at 8 mol/L urea (*4*) or 0.1 mol/L (*5*) and 0.5 mol/L (*6*) sodium chloride, elution fraction (*7*) at 1 mol/L sodium chloride. This analysis was done by L. Krebs and is an excerpt from her thesis [101].

In coomassie-stained SDS-PAGE, no overexpression of the hexahistidine-free pelB-SilL-P_5S_3(-Y) became obvious (Figure 25). The protein pattern was similar to that observed with the tag-bearing version (Figure 10) for which however the expression was proven by western blotting (Figure 11). Therefore, it was assumed that the tag-free version should be expressed at similar level, which is insufficient to be proven by coomassie-stained SDS-PAGE, and the purification was evaluated.

The expression of the tag-free polypeptide and lysis of bacteria was similar to the preparation of the tag-bearing version, however, changes had to be introduced. After lysis of the bacteria, the insoluble residue, containing the target poypeptide, was dissolved in 8 mol/L urea by the aid of ultrasonication. This solution, was applied to the ion echange resin and led to the binding of several proteins which co-eluted at 1 mol/L sodium chloride (Figure 25). Trials were made to bind the polypeptide from 0.1 % w/v SDS (standard purification protocol for the charge-compensated fusion) and 6 mol/L guanidium chloride, but no protein was bound to the resin at all (not shown). The binding of SDS to the polycationic P_5S_3(-Y) part presumably led to charge-compensation, because of which no interaction with the cation exchange resin was allowed. Guanidium chloride as a salt led to non-binding conditions in IEC.

In order to isolate the polycationic target P_5S_3(-Y), cleavage from the fusion and a second round of cation exchange chromatography was performed. HRV3C protease processing, as in the purification of pelB-SilL-P_xS_y(-Y,H), was not successful for isolation of the tag-free version. The impurities from the first chromatography (to isolate the fusion polypeptide) were again co-purified after the HRV3C digest (Figure 26). However, when cleavage was done with cyanogen bromide chemical cleavage, the impurities were fragmented. Analysis of the elution fractions in

this case revealed a polypeptide with apparent molecular weight around 15 kDa in SDS-PAGE, as expected for P_5S_3-versions (compare to Figure 15).

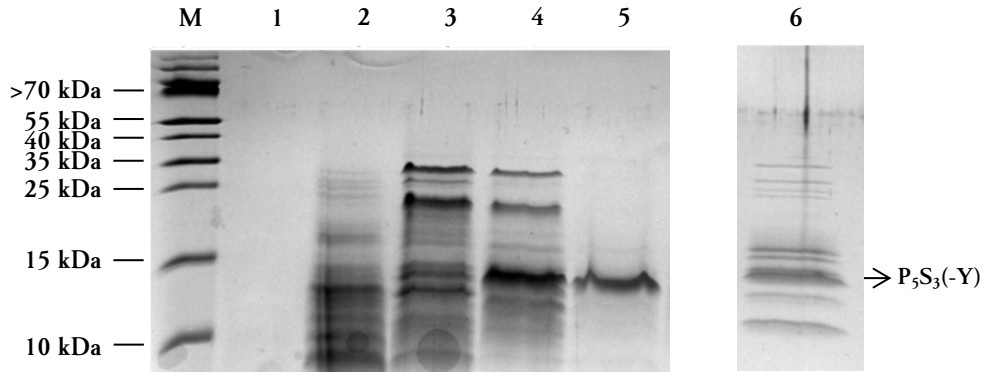

Figure 26. SDS-PAGE analysis for final purification of P_5S_3(-Y) by cation exchange chromatography. The analysis of IEC-purification of P_5S_3(-Y) liberated by cyanogen bromide cleavage (*left*) and HRV3C processing (*right*, only the last elution fraction) is shown. The different lanes contained: PageRuler Prestained Protein Ladder (Thermo Scientific, *lane M*), flow through IEC (*1*), elution fractions at 200 to 500 mmol/L ammonium carbonate in steps of 100 mmol/L (*2 - 5*), elution of the HRV3C-digested sample at 500 mmol/L ammonium carbonate (*6*). This analysis was done by L. Krebs and is an excerpt from her thesis [101].

Discussion

Three fusion technologies were evaluated for the expression of P_xS_y polypeptides in *Escherichia coli*: A charge-compensating fusion (SilL), a combination of the bacterial export leader pelB and the charge-compensating fusion (pelB-SilL), and ketosteroid isomerase (KSI) as inclusion body inducing fusion. P_xS_y was earlier examined for leader-free expression, which did not yield any target protein at all [79]. Fusion of only a bacterial export leader pelB allowed high overexpression rates, but the fusion protein was not recovered by metal affinity chromatography in satisfactory amounts and removal of the pelB fusion by protease processing was not proven.

For construction of the fusion proteins, plasmids providing the pelB leader sequence (pET 22b(+)) or the ketosteroid isomerase (KSI) fusion (pET 31b(+)) were used. The P_xS_y(-Y,H) sequences in pET 22b(+) were created earlier [76] and the pelB-leader was separated from the target gene by a HRV3C protease cleavage site [79]. The amplification of the anionic leader sequence from the *Sil1* gene [75], termed here SilL, was accomplished by standard PCR [85] with the gene provided on a pUC plasmid. Insertion into the pET 22b(+) plasmid hosting P_xS_y was successful with a standard cloning procedure (double restriction digestion, ligation). Later transfers of the new leader sequences between plasmids hosting different P_xS_y versions was facilitated by using enzyme combinations cutting in the expression region, and far upstream within other plasmid components, so that fragments above 1,000 bp instead of the 223 bp sequence of the anionic leader could be transferred, which was more convenient for sub-cloning.

Modifiations of the P_xS_y and leader sequences were possible by single-site mutagenesis using primers with central mismatches for full-plasmid PCR-amplification.

The transfer of the P_xS_y sequences into pET 31b(+) was achieved by *Ava*I digestion from the pelB-SilL fusion. *Ava*I cleaves at degenerate recognition sites, so that an an insert fragment with incompatible ends was generated. However, the insertion into the single-cleaved pET 31b(+) worked well. Since the opposite strands differed at the mismatch position in the cleavage site after ligation (until the first plasmid replication where the strands become separated), the plasmid was isolated and the different versions separated by a second transformation cycle. Transformants harboring the target plasmid with correct orientation of the insert were identified by colony-PCR with subsequent restriction analysis of the amplificates, and eventually by sequencing.

Analysis of the expression of the different P_xS_y fusions in the *Escherichia coli* strain Rosetta revealed that both the pelB-SilL charge-compensating double-leader, and the KSI-fusion supported the expression, whereas no expression was detected when only the SilL leader was present. In all cases, induction at high cell densities gave best overall expression yields, which is also known for other proteins [119]. Western blot analyzes revealed conditions under which overnight expression was supported. Extended expression phases were advantageous for expression of both the charge-compensated, and the KSI fusion (for the latter see also section III and ref [77]) To reduce the expression of the respective P_xS_y-fusions prior to induction, it was favorable to supplement the LB growth media with glucose as in the original LB recipe [90], which is well known to be advantageous for the suppression of leaky expression from pET plasmids [108].

Since the expression of the charge-compensated pelB-SilL-P_xS_y fusions was overall at a low level, attempts were made to increase the overexpression yield. It can in general be assumed that polypeptides with high cationic net charge are challenging in recombinant expression because they may bind and destroy anionic bacterial membranes [120] or interact with nucleic acids [121] and thereby affect the cellular metabolism. Though the charge-compensating fusion can be expected to reduce this adverse interactions [11], they are not necessarily prevented since e.g. the P_xS_y domain in the charge-compensated fusion was still sufficiently exposed to mediate the binding of the fusion to an anionic IEC resin. In pelB-SilL-P_xS_y, toxic effects during expression may also arise during translocation attempts to the periplasm, which can be inititated by the pelB-leader [78,115]. If this happens, the polycationic carboxy-terminal P_xS_y domain should, according to the "positive inside rule" [122], stall the translocation in a situation where the polyanionic SilL domain tries to pass through the hydrophobic membrane environment, presumably leading to membrane disturbances. Hence, the pelB-leader, which was found to be indispensable for the expression of P_xS_y with and without the SilL fusion as well as further silaffin-like polypeptides (results herein and ref's [79,80], diploma theses from the working group of Professor H. Paulsen), was modified in order to prevent the initiation of the polypeptide translocation. Export leader sequences typically contain a cationic amino-terminal region and a hydrophobic core, and their effectiveness for protein translocation can be tuned by varying these two characteristics

[118,123-125]. Amino-terminal charge reversal in the construction of pelB*(X)*-SilL-P_xS_y, i.e. exchanging a lysine for glutamate, did not substantially alter the expression pattern. In case of the P_xS_y expression, it might be that changing only the amino-terminal charge was insufficient for blocking the translocation initiation, which would be in accordance with earlier reports [118].

Successful increase of the overexpression of P_xS_y was achieved by changing the fusion to the inclusion body inducing ketosteroid isomerase (KSI). The purification of this version is presented and discussed in section III and published elsewhere [77].

The purification of the different P_xS_y versions bearing a hexahistidine-tag from the charge-compensated pelB-SilL fusion was accomplished by a two-step chromatographic approach and resulted in fairly pure polypeptide preparations. After lysis of the bacterial host, the insoluble cellular protein fraction was dissolved in SDS and a vast amount of bacterial impurities were removed by IMAC. One protein impurity (around 25 kDa) was usually co-purified with the charge-compensated fusion, though also was isolated in the absence of the fusion (i.e. in an attempt to isolate a fusion without the pelB-seuquence, namely SilL-P_5S_3(-Y,H) which however was not retrieved in the total cell protein fraction). Several proteins from *Escherichia coli* are well known impurities in IMAC [109]. They often belong to the class of stress-response proteins and bind to the resin either by exposed histidine-clusters, or they have particular metal-binding domains. However, since IMAC was only used for initial isolation of the fusion polypeptide, no additional efforts were made to reduce the amount of this impurity at this stage. The charge-compensated P_xS_y(H) versions were efficiently processed by HRV3C-protease, which worked well under the high-salt conditions in the IMAC elution fractions (0.5 mol/L imidazole and 1 mol/L sodium chloride), as expected from earlier reports describing the activation of the protease at high ionic strength [92]. The liberated polycationic P_xS_y(H) polypeptides were easily recovered by IEC. However, this required a complete digestion of the pre-cursor, since the fusion polypeptide was able to bind as well to the IEC-resin, and the fusion and polycationic target could not properly be separated by stepwise elution with ammonium carbonate solution.

An alternative purification approach was established using a two-step IEC for both the fusion and the target polypeptide purification, in order to produce P_xS_y versions without hexahistidine tag. The insoluble protein fraction from the lysed bacteria in this procedure had to be dissolved in urea, since no binding to the IEC resin was possible from SDS-solution as the polycationic target domain has a lowered cationic charge upon binding of the anionic detergent. Several bacterial proteins co-purified with the pelB-SilL-P_xS_y versions under the conditions applied. This is not unexpected since *Escherichia coli* has several proteins with basic isoelectric points, though the vast majority has an isoelectric point below 7 [126,127]. The unintended binding of bacterial proteins can be reduced by increasing the pH of the protein solutions which shall be applied on the IEC resin, but this does not generally prevent the occurrence of protein impurities [128]. For the P_xS_y purification, the IEC elution fraction containing the pelB-SilL-P_xS_y fusions was hence subjected to cyanogen bromide cleavage [93-95], which led to liberation of the P_xS_y polypeptides as well as partial fragmentation of the impurities. By this, additional impurities were removed in

a second IEC step for isolation of the P_xS_y target polypeptides. However, still there were impurities which resisted elution at considerably high salt cncentrations, so that the purification was only successful in case of P5S3-versions, which have the highest net charge in the P_xS_y series and hence eluted only at elevated salt concentrations.

In conclusion, two fusions (pelB-SilL and KSI) were identified to support the expression of P_xS_y polypeptides in *Escherichia coli*. The fusions can be removed by protease digest or chemical cleavage to liberate the target peptide. Here, purification procedures are described for the purification of P_xS_y expressed as a charge-compensated fusion, by applying a two-step chromatographic approach with IMAC purification of the fusion and IEC purification of the target polypeptides after proteolytic removal of the pre-sequence. For the longest version P_5S_3, an alternative purification method was established in which both the fusion, and the liberated target polypeptide were purified by IEC, and the pre-sequence was removed by cyanogen bromide cleavage. This procedure enabled the purification of P_5S_3 without an affinity tag. For shorter P_xS_y, the method failed since their discrimination from bacterial impurities during solution from the IEC-resin was insufficient.

References

[1] G.L. Rosano, E.A. Ceccarelli, Front. Microbiol. 5 (2014) 172.

[2] J.W. Cuozzo, H.H. Soutter, J. Biomol. Screen. 19 (2014) 1000.

[3] S. Gräslund, P. Nordlund, J. Weigelt, B.M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. Dhe-Paganon, H. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S.-H. Kim, Z. Rao, Y. Shi, T.C. Terwilliger, C.-Y. Kim, L.-W. Hung, G.S. Waldo, Y. Peleg, S. Albeck, T. Unger, O. Dym, J. Prilusky, J.L. Sussman, R.C. Stevens, S.A. Lesley, I.A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M.I. Donnelly, W.H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S.K. Burley, J.S. Emtage, J.M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S.C. Almo, J.B. Bonanno, A. Fiser, S. Swaminathan, F.W. Studier, M.R. Chance, A. Sali, T.B. Acton, R. Xiao, L. Zhao, L.C. Ma, J.F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C.K. Ho, D. Wang, H. Wang, M. Jiang, G.T. Montelione, D.I. Stuart, R.J. Owens, S. Daenke, A. Schütz, U. Heinemann, S. Yokoyama, K. Büssow, K.C. Gunsalus, Nat. Methods 5 (2008) 135.

[4] C. Gustafsson, S. Govindarajan, J. Minshull, Trends Biotechnol. 22 (2004) 346.

[5] J.F. Kane, Curr. Opin. Biotechnol. 6 (1995) 494.

[6] N.A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, O. Gileadi, Protein Expr. Purif. 59 (2008) 94.

[7] N.Y. Yount, M.R. Yeaman, Ann. N. Y. Acad. Sci. 1277 (2013) 127.

[8] J. Jarczak, E.M. Kościuczuk, P. Lisowski, N. Strzalkowska, A. Jóźwik, J. Horbanczuk, J. Krzyzewski, L. Zwierzchowski, E. Bagnicka, Hum. Immunol. 74 (2013) 1069.

[9] N.S. Parachin, K.C. Mulder, A.A.B. Viana, S.C. Dias, O.L. Franco, Peptides 38 (2012) 446.

[10] H.-K. Kim, D.-S. Chun, J.-S. Kim, C.-H. Yun, J.-H. Lee, S.-K. Hong, D.-K. Kang, Appl. Microbiol. Biotechnol. 72 (2006) 330.

[11] J.H. Lee, I. Minn, C.B. Park, S.C. Kim, Protein Expr. Purif. 12 (1998) 53.

[12] Y.S. Kim, M.J. Kim, P. Kim, J.H. Kim, Biotechnol. Lett. 28 (2006) 357.

[13] Y. Li, J. Wang, J. Yang, C. Wan, X. Wang, H. Sun, Protein Expr. Purif. 95 (2014) 182.

[14] L.E. Quadri, L.Z. Yan, M.E. Stiles, J.C. Vederas, J. Biol. Chem. 272 (1997) 3384.

[15] X. Xu, F. Jin, X. Yu, S. Ji, J. Wang, H. Cheng, C. Wang, W. Zhang, Protein Expr. Purif. 53 (2007) 293.

[16] Y.-G. Xie, F.-F. Han, C. Luan, H.-W. Zhang, J. Feng, Y.-J. Choi, D. Groleau, Y.-Z. Wang, Biomed Res. Int. 2013 (2013) 754319.

[17] Y. Li, X. Li, G. Wang, Protein Expr. Purif. 47 (2006) 498.

[18] P.J. Barrell, O.W. Liew, A.J. Conner, Protein Expr. Purif. 33 (2004) 153.

[19] G.M. Gibbs, B.E. Davidson, A.J. Hillier, Appl. Environ. Microbiol. 70 (2004) 3292.

[20] Y.-H. Yang, G.-G. Zheng, G. Li, X.-J. Zhang, Z.-Y. Cao, Q. Rao, K.-F. Wu, Protein Expr. Purif. 37 (2004) 229.

[21] K.M. Morin, S. Arcidiacono, R. Beckwitt, C.M. Mello, Appl. Microbiol. Biotechnol. 70 (2006) 698.

[22] C. Pi, J. Liu, L. Wang, X. Jiang, Y. Liu, C. Peng, S. Chen, A. Xu, J. Biotechnol. 128 (2007) 184.

[23] Z. Tian, D. Teng, Y. Yang, J. Luo, X. Feng, Y. Fan, F. Zhang, J. Wang, Appl. Microbiol. Biotechnol. 75 (2007) 117.

[24] Y. Li, X. Li, H. Li, O. Lockridge, G. Wang, Protein Expr. Purif. 54 (2007) 157.

[25] L. Huang, C.B. Ching, R. Jiang, Leong, Susanna Su Jan, Protein Expr. Purif. 61 (2008) 168.

[26] Y. Yanga, Z. Tiana, D. Teng, J. Zhang, J. Wang, J. Wang, J. Biotechnol. 139 (2009) 326.

[27] L. Zhou, Z. Zhao, B. Li, Y. Cai, S. Zhang, Protein Expr. Purif. 64 (2009) 225.

[28] L. Huang, Leong, Susanna Su Jan, R. Jiang, Appl. Microbiol. Biotechnol. 84 (2009) 301.

[29] X. Lu, X. Jin, J. Zhu, H. Mei, Y. Ma, F. Chu, Y. Wang, X. Li, Appl. Microbiol. Biotechnol. 87 (2010) 2169.

[30] X.-L. Jing, X.-G. Luo, W.-J. Tian, L.-H. Lv, Y. Jiang, N. Wang, T.-C. Zhang, Curr. Microbiol. 61 (2010) 197.

[31] L.L. Utkina, E.O. Zhabon, A.A. Slavokhotova, E.A. Rogozhin, A.N. Shiian, E. V Grishin, T.A. Egorov, T.I. Odintsova, V.A. Pukhal'skiĭ, Genetika 46 (2010) 1645.

[32] T. Gong, W. Li, Y. Wang, Y. Jiang, Q. Zhang, W. Feng, Z. Jiang, M. Li, Braz. J. Microbiol. 42 (2011) 1180.

[33] Y. Li, Protein Expr. Purif. 81 (2012) 201.

[34] Y. Li, Protein Expr. Purif. 87 (2013) 72.

[35] I. Garcia-Montoya, S.A. González-Chávez, J. Salazar-Martinez, S. Arévalo-Gallegos, S. Sinagawa-Garcia, Q. Rascón-Cruz, Biometals 26 (2013) 113.

[36] Y. Li, Appl. Biochem. Biotechnol. 169 (2013) 1847.

[37] R.A. Aleinein, R. Hamoud, H. Schäfer, M. Wink, Appl. Microbiol. Biotechnol. 97 (2013) 3535.

[38] L. Xia, F. Zhang, Z. Liu, J. Ma, J. Yang, Exp. Ther. Med. 5 (2013) 1745.

[39] G.S. Souza, V.V. do Nascimento, L.P. de Carvalho, E.J. de Melo, K.V. Fernandes, O.L. Machado, C.A. Retamal, V.M. Gomes, A. de O. Carvalho, Exp. Parasitol. 135 (2013) 116.

[40] A.A. Astafieva, E.A. Rogozhin, Y.A. Andreev, T.I. Odintsova, S.A. Kozlov, E. V Grishin, T.A. Egorov, Plant Physiol. Biochem. PPB 70 (2013) 93.

[41] Y. Tao, D.-M. Zhao, Y. Wen, Protein Expr. Purif. 94 (2014) 73.

[42] Y. qing Chen, S. quan Zhang, B.C. Li, W. Qiu, B. Jiao, J. Zhang, Z. yu Diao, Protein Expr. Purif. 57 (2008) 303.

[43] C. Morassutti, F. de Amicis, A. Bandiera, S. Marchetti, Protein Expr. Purif. 39 (2005) 160.

[44] Y.-G. Xie, C. Luan, H.-W. Zhang, F.-F. Han, J. Feng, Y.-J. Choi, D. Groleau, Y.-Z. Wang, Protein Pept. Lett. 20 (2013) 54.

[45] I. Hong, Y.-S. Kim, S.-G. Choi, J. Microbiol. Biotechnol. 20 (2010) 350.

[46] H. Diao, C. Guo, D. Lin, Y. Zhang, Biochem. Biophys. Res. Commun. 357 (2007) 840.

[47] X. Xu, F. Jin, X. Yu, S. Ren, J. Hu, W. Zhang, Protein Expr. Purif. 55 (2007) 175.

[48] R. Wu, Q. Wang, Z. Zheng, L. Zhao, Y. Shang, X. Wei, X. Liao, R. Zhang, Mol. Biol. Rep. 41 (2014) 4163.

[49] M. Zhang, Z. Qiu, Y. Li, Y. Yang, Q. Zhang, Q. Xiang, Z. Su, Y. Huang, Appl. Microbiol. Biotechnol. 97 (2013) 3913.

[50] H. Hou, W. Yan, K. Du, Y. Ye, Q. Cao, W. Ren, Protein Expr. Purif. 92 (2013) 230.

[51] C. Zhang, X. He, Y. Gu, H. Zhou, J. Cao, Q. Gao, PLoS One 9 (2014) e103456.

[52] J.F. Li, J. Zhang, R. Song, J.X. Zhang, Y. Shen, S. quan Zhang, Appl. Microbiol. Biotechnol. 84 (2009) 383.

[53] J.F. Li, J. Zhang, Z. Zhang, H.W. Ma, J.X. Zhang, S. quan Zhang, Protein J. 29 (2010) 314.

[54] B. Bommarius, H. Jenssen, M. Elliott, J. Kindrachuk, M. Pasupuleti, H. Gieren, K.-E. Jaeger, Hancock, R E W, D. Kalman, Peptides 31 (2010) 1957.

[55] J.F. Li, J. Zhang, Z. Zhang, C.T. Kang, S. quan Zhang, Curr. Microbiol. 62 (2011) 296.

[56] M. Satakarni, R. Curtis, Protein Expr. Purif. 78 (2011) 113.

[57] P.M. Hwang, J.S. Pan, B.D. Sykes, FEBS Lett. 588 (2014) 247.

[58] A. Kuliopulos, C.T. Walsh, J. Am. Chem. Soc. 116 (1994) 4599.

[59] T.-J. Park, J.-S. Kim, S.-S. Choi, Y. Kim, Protein Expr. Purif. 65 (2009) 23.

[60] I. Cipáková, J. Gasperik, E. Hostinová, Protein Expr. Purif. 45 (2006) 269.

[61] M. Zorko, B. Japelj, I. Hafner-Bratkovic, R. Jerala, Biochim. Biophys. Acta 1788 (2009) 314.

[62] T.-J. Park, J.-S. Kim, H.-C. Ahn, Y. Kim, Biophys. J. 101 (2011) 1193.

[63] S. Kyle, A. Aggeli, E. Ingham, M.J. McPherson, Biomaterials 31 (2010) 9395.

[64] J.M. Riley, A. Aggeli, R.J. Koopmans, M.J. McPherson, Biotechnol. Bioeng. 103 (2009) 241.

[65] S. Sharpe, W.-M. Yau, R. Tycko, Protein Expr. Purif. 42 (2005) 200.

[66] B.M. Hartmann, W. Kaar, R.J. Falconer, B. Zeng, A.P.J. Middelberg, Middelberg, A P J, J. Biotechnol. 135 (2008) 85.

[67] P.M. Hwang, J.S. Pan, B.D. Sykes, Protein Expr. Purif. 85 (2012) 148.

[68] D.-X. Zhao, Z.-C. Ding, Y.-Q. Liu, Z.-X. Huang, Protein Expr. Purif. 53 (2007) 232.

[69] Mergulhão, F J M, D.K. Summers, G.A. Monteiro, Biotechnol. Adv. 23 (2005) 177.

[70] K. Talmadge, W. Gilbert, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 1830.

[71] S.C. Makrides, Microbiol. Rev. 60 (1996) 512.

[72] H.P. Sørensen, K.K. Mortensen, J. Biotechnol. 115 (2005) 113.

[73] J. Arnau, C. Lauritzen, G.E. Petersen, J. Pedersen, Protein Expr. Purif. 48 (2006) 1.

[74] Y. Li, Biotechnol. Lett. 33 (2011) 869.

[75] N. Kröger, R. Deutzmann, M. Sumper, Science 286 (1999) 1129.

[76] Christian Roos, "Klonierung Und Charakterisierung von Silaffinaehnlichen Proteinen" - Diploma Thesis, 2008.

[77] C. Zerfaß, S. Braukmann, S. Nietzsche, S. Hobe, H. Paulsen, Protein Expr. Purif. 108 (2015) 1.

[78] S.P. Lei, H.C. Lin, S.S. Wang, J. Callaway, G. Wilcox, J. Bacteriol. 169 (1987) 4379.

[79] Konstantin Müller, "Klonierung Und Charakterisierung von Phosphorylierbaren Polykationischen Proteinen" - Bachelor Thesis, 2009.

[80] Anna-Angelika Hippmann, "Overexpression of Different Silaffin Fragments with Hexahistidyl-Tag" - Diploma Thesis, 2005.

[81] Andreas Richter, "Klonierung von Cysteinfreien Polykationischen Peptiden Und Click-Chemische Verknüpfung Eines Lichtsammelkomplexes Mit Einem Silaffinähnlichen Protein" - State Exam, 2010.

[82] A. Kuliopulos, Proc. Natl. Acad. Sci. 84 (1987) 8893.

[83] Biochem. J. 219 (1984) 345.

[84] B.E. Kemp, D.J. Graves, E. Benjamini, E.G. Krebs, J. Biol. Chem. 252 (1977) 4888.

[85] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 3. ed., 3., Wiley, New York, 1995.

[86] F.W. Studier, B.A. Moffatt, J. Mol. Biol. 189 (1986) 113.

[87] D. Hanahan, J. Mol. Biol. 166 (1983) 557.

[88] M.G. Cordingley, R.B. Register, P.L. Callahan, V.M. Garsky, R.J. Colonno, J. Virol 63 (1989) 5037.

[89] Janica Wiederstein, "Ketosteroid-Isomerase Als Hydrophobe Fusion Zur Bakteriellen Überexpression Polykationischer, Selbst-Assemblierender Peptide" - Bachelor Thesis, 2013.

[90] G. Bertani, J. Bacteriol. 62 (1951) 293.

[91] G. Bertani, J. Bacteriol. 186 (2004) 595.

[92] Q.M. Wang, R.B. Johnson, Virology 280 (2001) 80.

[93] E. Gross, B. Witkop, J. Am. Chem. Soc. 83 (1961) 1510.

[94] E. Gross, B. Witkop, J. Biol. Chem. 237 (1962) 1856.

[95] W.A. Schroeder, J.B. Shelton, J.R. Shelton, Arch. Biochem. Biophys. 130 (1969) 551.

[96] H. Schaegger, G. von Jagow, Anal. Biochem. 166 (1987) 368.

[97] G. von Jagow, H. Schaegger, A Practical Guide to Membrane Protein Purification, Academic Press, San Diego, 1994.

[98] H. Schägger, Nat. Protoc. 1 (2006) 16.

[99] W.N. Burnette, Anal. Biochem. 112 (1981) 195.

[100] Barbara Glöckle, "Zufallsmutagenese an Silaffinen" - Diploma Thesis, 2008.

[101] Laura Krebs, "Klonierung, Expression Und Isolierung Basischer, Selbst-Assemblierender Peptide Mit Und Ohne Histidin-Tag" - Bachelor Thesis, 2013.

[102] H.P. Sørensen, K.K. Mortensen, Microb. Cell Fact. 4 (2005) 1.

[103] Y. Shi, R.A. Mowery, J. Ashley, M. Hentz, A.J. Ramirez, B. Bilgicer, H. Slunt-Brown, D.R. Borchelt, B.F. Shaw, Protein Sci. 21 (2012) 1197.

[104] A. Rath, M. Glibowicka, V.G. Nadeau, G. Chen, C.M. Deber, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 1760.

[105] E.I. Georgieva, R. Sendra, Anal. Biochem. 269 (1999) 399.

[106] N. Panayotatos, E. Radziejewska, A. Acheson, D. Pearsall, A. Thadani, V. Wong, J. Biol. Chem. 268 (1993) 19000.

[107] W.W. de Jong, A. Zweers, L.H. Cohen, Biochem. Biophys. Res. Commun. 82 (1978) 532.

[108] T.H. Grossman, E.S. Kawasaki, S.R. Punreddy, M.S. Osburne, Gene 209 (1998) 95.

[109] V.M. Bolanos-Garcia, O.R. Davies, Biochim. Biophys. Acta - Gen. Subj. 1760 (2006) 1304.

[110] D. V. Tulumello, C.M. Deber, Biochemistry 48 (2009) 12096.

[111] H. Cho, R. LaMarca, C. Chan, Biochem. Biophys. Res. Commun. 427 (2012) 764.

[112] F. Gentile, P. Amodeo, F. Febbraio, F. Picaro, A. Motta, S. Formisano, R. Nucci, J. Biol. Chem. 277 (2002) 44050.

[113] G. von Heijne, Eur. J. Biochem. 133 (1983) 17.

[114] P.S. Chowdhury, A. Kushwaha, S. Abrol, V.K. Chaudhary, Protein Expr. Purif. 5 (1994) 89.

[115] H. Le Calvez, J.M.M. Green, D. Baty, Gene 170 (1996) 51.

[116] P.S. Hauser, R.O. Ryan, Protein Expr. Purif. 54 (2007) 227.

[117] G. Heijne, EMBO J. 5 (1986) 3021.

[118] J.W. Izard, S.L. Rusch, D.A. Kendall, J. Biol. Chem. 271 (1996) 21579.

[119] J. Ou, L. Wang, X. Ding, J. Du, Y. Zhang, H. Chen, A. Xu, Biochem. Biophys. Res. Commun. 314 (2004) 174.

[120] W. Hartmann, H.J. Galla, Biochim. Biophys. Acta 509 (1978) 474.

[121] D.E. Olins, A.L. Olins, Von Hippel, P H, J. Mol. Biol. 24 (1967) 157.

[122] G. von Heijne, J. Mol. Biol. 225 (1992) 487.

[123] J.W. Izard, M.B. Doughty, D.A. Kendall, Biochemistry 34 (1995) 9904.

[124] C. Hikita, S. Mizushima, J. Biol. Chem. 267 (1992) 12375.

[125] M. Akita, S. Sasaki, S. Matsuyama, S. Mizushima, J. Biol. Chem. 265 (1990) 8164.

[126] P.Z. O'Farrell, H.M. Goodman, P.H. O'Farrell, Cell 12 (1977) 1133.

[127] K.M. Champion, J.C. Nishihara, J.C. Joly, D. Arnott, Proteomics 1 (2001) 1133.

[128] S. Adhikari, P.V. Manthena, K. Sajwan, K.K. Kota, R. Roy, Anal. Biochem. 400 (2010) 203.

Section II - Analysis of the unusual self-assembly of a series of designed polycationic polypeptides that adopt extended conformations in solution

Introduction

Polypeptides with low hydrophobicity and high net charge are usually extended in solution and, due to the lack of rigid conformations as α-helices or β-sheets, are defined as "natively unfolded" or "intrinsically disordered" [1,2]. Intrinsically disordered proteins or protein domains fulfill a large number of biological functions [2]. As one example, they are abundant as biomineralizing proteins [3], for which low-order, flexible structures support the efficient binding to minerals concomitant with a structural adaptation [4].

The extended conformation of polypeptides with high net charge is considered to be a result from side-chain repulsion [5]. By studying the charged states of poly-L-glutamate and poly-L-lysine, it was found that this extended conformation is not fully unordered, but contains a large fraction of e.g. the extended polyproline II conformation, which is lost upon decharging by changing the environmental pH in favor to adopt more compact structures as α-helices [6-10].

In the presence of certain multivalent anions, polycationic polypeptides as poly-L-lysine can self-assemble [11], a similarity with biomineralization active silaffins [12] and polyamines [13]. This feature was used to guide their templating effect in biomimetic silica formation. For silaffins, the self-assembly was even demonstrated to be a strict prerequisite to acquire silica precipitation activity [12]. Poly-L-lysine of sufficient degree of polymerization can assemble into single-crystalline platelets in the presence of phosphate, by which an α-helical conformation of the polypeptide is induced [14]. These platelets can in turn be used to template the formation of hexagonal silica platelets by biomimetic mineralization techniques [15].

To understand the early process of peptide-mineral interaction, it is hence crucial to analyze the solution state behavior of biomineralizing peptides. Here, P_xS_y polypeptide versions were examined in terms of their solution conformation by circular dichroism spectroscopy as well as their self-assembling properties by dynamic light scattering and transmission electron microscopy.

Abbreviations

Bis-Tris: bis(2-hydroxyethyl)aminotris(hydroxymethyl)methane; CD: circular dichroism spectroscopy; DLS: dynamic light scattering; PPII: polyproline II secondary polypeptide structure; TEM: transmission electron microscopy; Tris: tris(hydroxymethyl)aminomethane

Materials and Methods

Preparation of recombinant polypeptides and electrophoretic analyzes

The peptide preparation was achieved as described in the sections I and III. Lyophilized polypeptides which originated from pelB-SilL-fusions were dissolved in 50 mmol/L sodium hydroxide solution. The amount of polypeptide was not in all cases determined since e.g. polypeptides without aromatic residues (for UV/vis spectroscopic quantification) were investigated and expression yields (see Section I of this thesis) were too low for gravimetric quantification. To hence compare different preparations, ten aliquots of purified polypeptide were prepared when the polypeptide was isolated from an 800 ml bacterial culture, and each aliquot was dissolved in a volume of 50 µl sodium hydroxide solution. For measurements at neutral pH, 1 mol/L hydrochloric acid was added for neutralization and the pH verified by pH indicator strips. Lyophilized polypeptides derived from KSI-fusions were dissolved in water.

UV/vis spectroscopy was done on a Jasco V550 UV/vis-Spectrophotometer (Jasco GmbH, Gross-Umstadt (Germany)) with the polypeptide samples in quartz cuvettes SUPRASIL 1000 QS (Hellma Analytics GmbH, Müllheim (Germany)). For quantification, the polypeptide was dissolved in 1 % w/v SDS solution. The extinction coefficient of tyrosine of 1,490 l mol^{-1} cm^{-1} was applied [16]. In most preparations of P_xS_y from the pelB-SilL fusion, concentrations were low and did not allow a proper quantification, but KSI-P_xS_y derived polypeptides were sufficiently enriched for reliable UV/vis measurements. Alternatively, SDS-PAGE in the Tris/Tricine system [17-19] was used for approximate estimation of the concentration.

Secondary structure determination by circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was done in mountable quartz cuvettes with 0.1 mm pathlength (Suprasil 106-QS, Hellma Analytics GmbH, Müllheim (Germany)). Spectra were recorded at 300 - 185 nm, with five- to ten-fold accumulation, and an appropriate peptide-free reference was subtracted. When the polypeptide concentration was known, the data were presented in mean residue ellipticity. If the quantification was unrealiable or impossible, spectra were only presented as raw data (millidegree) for qualitative comparison.

Self-assembly of P_xS_y studied by DLS and TEM

For dynamic light scattering (DLS) analysis, polypeptide samples derived from pelB-SilL fusions, redissolved in 50 mmol/L sodium hydroxide solution (see first section materials and methods) were analyzed ten-fold diluted. KSI-derived and quantified samples were analyzed at a polypeptide-concentration of 50 µmol/L. Samples were buffered with 50 mmol/L sodium acetate (pH 4.5), Bis-Tris-HCl (pH 5.5 and 6.5) or Tris-HCl (pH 7.5, 8.5 or 9.5). DLS was recorded on a

Zetasizer Nano-S Size (Malvern Instruments GmbH, Herrenberg (Germany)). Measurements were averages of 10 repetitions of each ten seconds measuring time, data were analyzed with "General Purpose Method" of the according software.

For transmission electron microscopy (TEM), polypeptide solutions were prepared as for DLS. From these samples, 10 µl were applied for 30 s to a glow-discharged carbon-coated copper-grid which had been pre-treated in plasma (K100X Glow Discharge, Emitech GmbH, Lohmar (Germany)) for hydrophilization. The grid was washed with water and subsequently negative-stained by applying 10 µl 1 % w/v uranyl acetate for 30 s. After each step, residual solution was blotted with tissue paper, and eventually the grid was allowed to air-dry. Micrographs were recorded on a Technai 12 transmission electron microscope (FEI, Gräfeling, Germany).

Results

Conformation of P_xS_y polypeptides examined by circular dichroism spectroscopy

Different P_xS_y versions were investigated by circular dichroism (CD) spectroscopy to examine the impact of different solution conditions and the polypeptide length on the polypeptide conformation. For those versions expressed as pelB-SilL fusions, exact concentration determination was impossible because of low expression and purification yields, so that the comparison was qualitative only.

CD-spectra of hexahistidine tagged P_5S_3(-Y,H), P_3S_1(-Y,H) and P_2S_0(-Y,H) as well as the respective tyrosine-containing versions P_5S_3(H) and P_3S_1(H) are presented in Figure 1. P_2S_0(H) was not purified and is missing in the analysis. All peptides were expressed as pelB-SilL fusions, and measurement was done at pH 7.5 and 23°C. Spectra of all peptide versions had a major negative peak between 195 and 200 nm. Additionally, they exhibited a local maximum between 215 and 220 nm, followed by a negative region around 230 nm. Some of the spectra were shifted from the baseline, indicated by a deviation from zero around 250 nm where polypeptide spectra do not show a circular dichroism [20]. This deviation was hence an indication of experimental bias which particularly occurred at low polypeptide concentrations where the intensity of specific signals was insufficient. One source of error which was repeatedly noted were deviations in the sample temperature, causing baseline shifts of up to 1 to 2 mdeg, as seen e.g. in Figure 1c. Despite this bias, the general spectral properties indicated a considerable fraction of polyproline II (PPII) secondary structure motif, which is characterized for non-proline polypeptides by a negative ellipticity around 195 nm and usually a positive peak around 220 nm [7]. This spectral assignement was further supported by the P_5S_3 (from KSI-fusion) spectra at different temperatures, shown in section III and ref [21]. Insertion of the tyrosine for UV/vis quantification did not cause significant changes in the CD spectra (Figure 1).

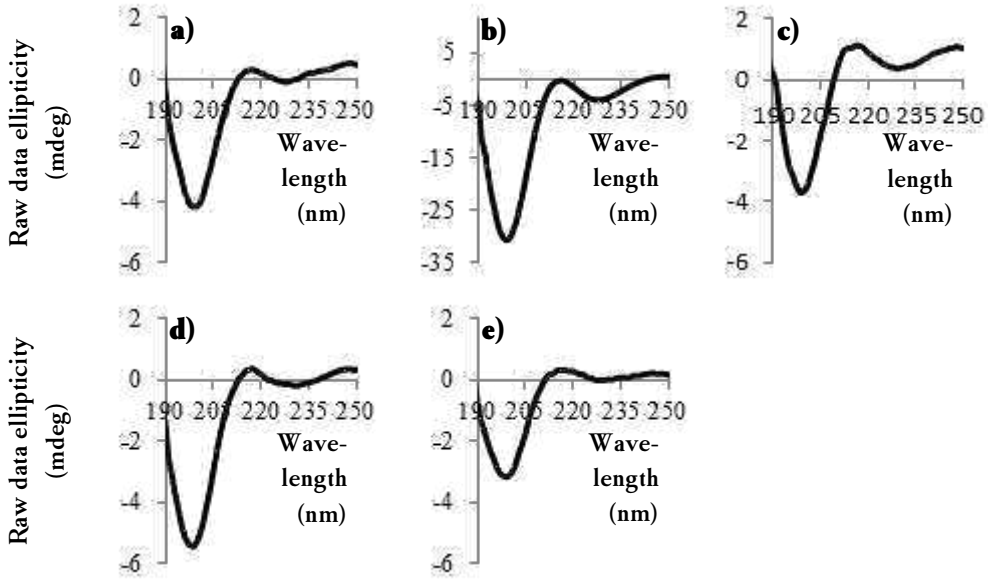

Figure 1. CD spectrum of P_xS_y at pH 7.5 and 23 °C. Polypetide versions were: (*a*) P_5S_3(-Y,H), (*b*) P_3S_1(-Y,H), (*c*) P_2S_0(-Y,H), (*d*) P_5S_3(H), (*e*) P_3S_1(H), all expressed as their respective pelB-SilL-P_xS_y fusions. For measurement, a lyophilized sample was dissolved in 50 mmol/l sodium hydroxide and subsequently neutralized with hydrochloric acid.

The effect of a hexahistidine tag on the polypeptide conformation was evaluated by comparison of tag-bearing and tag-free P_5S_3. Both versions were prepared as tyrosine-containing polypeptides from their KSI-fusions and exact determination of their concentration allowed the comparison of their mean residue ellipticity. Both polypeptides were characterized by virtually the same spectra (Figure 2). This shows that the hexahistidine tag does not have an impact on the P_5S_3 conformation in solution.

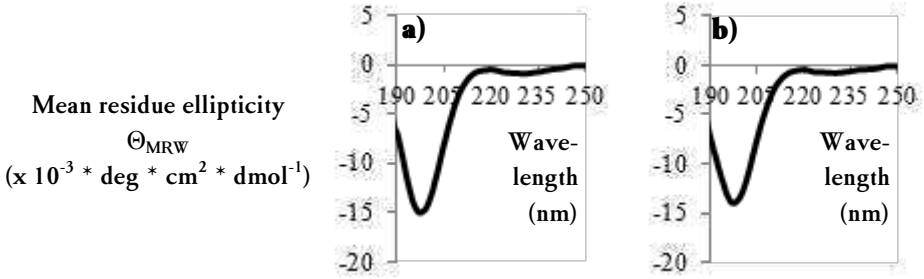

Figure 2. Influence of the hexahistidine tag on secondary structure of P_5S_3. CD spectra were recorded for (*a*) P_5S_3(H) and (*b*) P_5S_3. Both were expressed from their respective KSI fusion. Spectra were recorded at 23 °C.

In a basic environment (50 mmol/L sodium hydroxide, expected pH around 12.5), spectra were notably altered, as shown for P_5S_3(H) and cysteine-containing $CysP_3S_1$(-Y,H) (Figure 3). The strong negative ellipticity peak in the area 195 to 200 nm was still most prominent, however, around 220 nm the ellipticity was increasingly negative.

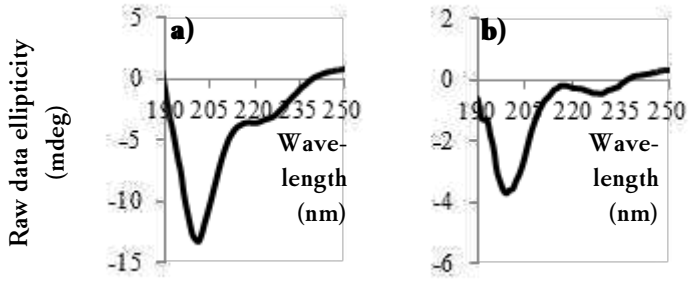

Figure 3. CD spectrum of P_xS_y in basic environment at 23 °C. Peptide versions were (*a*) P_5S_3(H) and (*b*) $CysP_3S_1$(-Y,H), expressed as their respective pelB-SilL fusions. For preparation, lyophilized samples were dissolved in 50 mmol/L sodium hydroxide.

The effect of the ionic strength was evaluated for P_5S_3(H) at 100 µmol/L in the presence of different cations and anions. The cations which were compared were monovalent sodium ions (Figure 4) and divalent calcium ions (Figure 5) as well as monovalent ammonium and guanidium ions (Figure 6) due to their similarity with lysine and arginine side-chains, respectively. In all assays of the cation series, the respective chloride salts were used. The effect of polyvalent phosphate anions (Figure 7) were examined by introducing phosphate buffer with sodium as counter-ion. As ion concentration, 50 mmol/L was used as standard, since this concentration was above the phosphate concentration necessary for silaffins to acquire silica formation activity [12]. Sodium chloride was additionally evaluated at 1,000 mmol/L to analyze the secondary structure in a high-salt environment. All salt effects were analyzed at near-neutral pH 7.5, sodium chloride and sodium phosphate buffer additionally at pH 9.5.

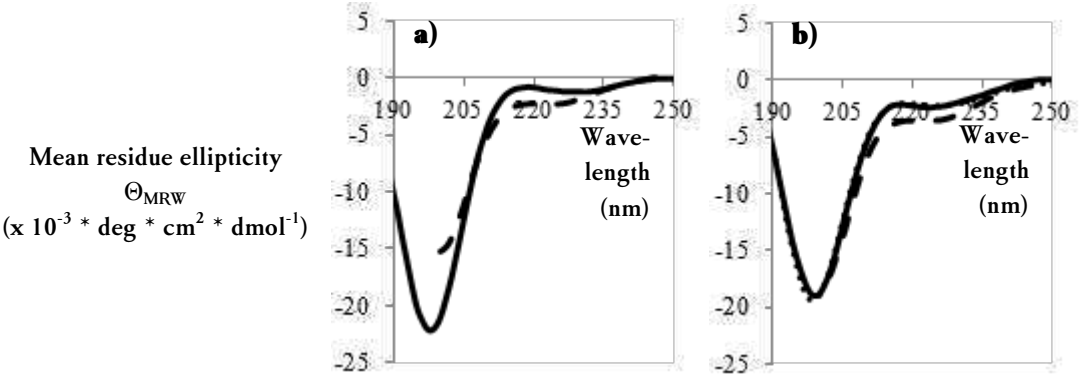

Figure 4. Effect of sodium chloride on CD spectra of P_5S_3(H). The polypeptide was expressed as KSI-P5S3(H). Lyophilized P5S3(H) was dissolved in water and buffered to 20 mmol/L Tris-HCl at (*a*) pH 7.5 or (*b*) pH 9.5. The polypeptide was examined at 100 µmol/L in the absence (*solid line*) or presence of 50 (*dotted line*, only at pH 9.5) or 1,000 mmol/L (*dashed line*) sodium chloride. Please note that below 200 nm, spectra at 1,000 mmol/L sodium chloride were unreliable because of a rapid increase in optical density, that is, spectra are not shown in this region.

At 50 mmol/L of the respective cations and anions, or 500 ions per polypeptide chain, no significant effects on the CD-spectra were recognized for any of the salts. However, at 1,000 mmol/L sodium chloride (10,000 of the respective ions per peptide chain), spectra were more negative in the 220 nm region, possibly indicating some appearance of α-helix or β-sheet conformation [20]. A more exact determination of this new spectral feature was impossible since

spectral information below 200 nm would have been required, but could not be measured because of the high optical density of the 1,000 mmol/L sodium chloride solution below this wavelength.

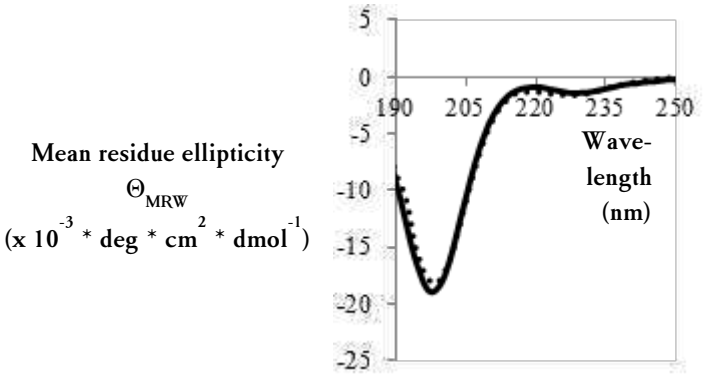

Figure 5. Effect of divalent calcium cations on the CD spectra of $P_5S_3(H)$. The polypeptide was expressed as KSI-$P_5S_3(H)$. Lyophilized polypeptide was dissolved in water and buffered to 20 mmol/L Tris-HCl pH 7.5. $P_5S_3(H)$ was examined at 100 µmol/L in the absence (*solid line*) or presence of 50 mmol/L (*dotted line*) calcium chloride.

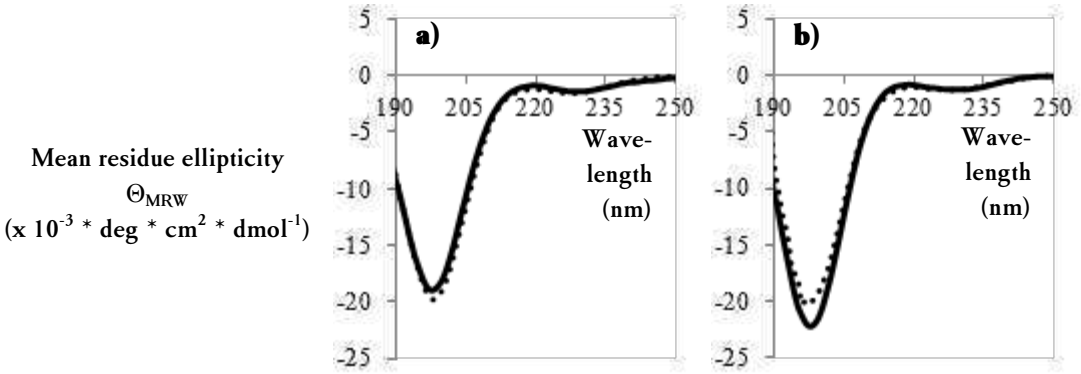

Figure 6. Effect of ammonium and guanidium ions on the CD spectra of $P_5S_3(H)$. The polypeptide was expressed as KSI-$P_5S_3(H)$. Lyophilized polypeptide was dissolved in water and buffered to 20 mmol/L Tris-HCl pH 7.5. $P_5S_3(H)$ was examined at 100 µmol/L in the absence (*solid line*) or presence of 50 mmol/L (*dotted line*) of (*a*) ammonium chloride or (*b*) guanidium chloride.

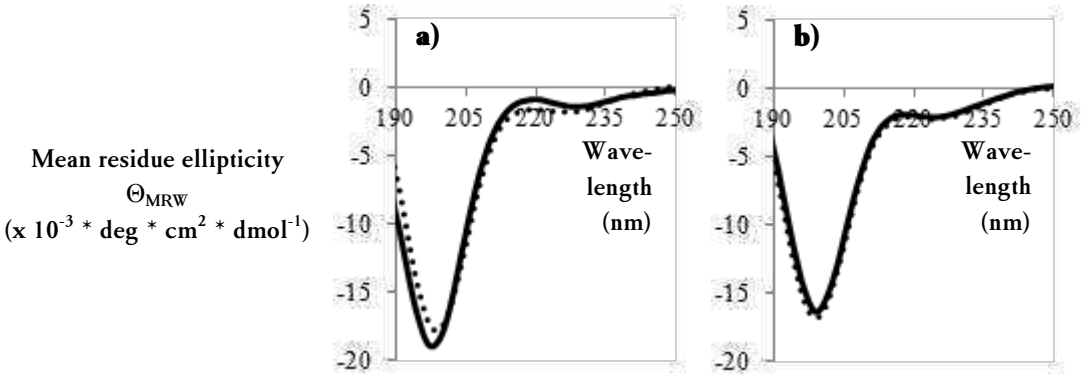

Figure 7. Effect of phosphate on the CD spectra of $P_5S_3(H)$. The polypeptide was expressed as KSI-$P_5S_3(H)$. Lyophilized $P_5S_3(H)$ was dissolved in water and buffered to 20 mmol/L Tris-HCl at (*a*) pH 7.5 or (*b*) pH 9.5. $P_5S_3(H)$ at 100 µmol/L was analysed in the absence (*solid line*) and presence (*dotted line*) of 50 mmol/L phosphate buffer (pH-adjusted with sodium hydroxide).

Self-assembly of P_xS_y polypeptides studied by UV/vis spectroscopy, dynamic light scattering and transmission electron microscopy

The self-assembly of polycationic silaffins in the presence of phosphate is a prerequisite for their silica-biomineralizing activity [12]. Hence, the self-assembling capability of P_xS_y polypeptides was evaluated under different conditions. The UV/vis absorption spectrum revealed that P_5S_3(H) self-assembled even in the absence of polyvalent anions like phosphate (Figure 8). Both at basic pH (sodium hydroxide), and after neutralization (by adding hydrochloric acid) P_5S_3(H) gave rise to a scattering curve. Scattering was found in the wavelength region from 200 to 400 nm, without a clear peak of tyrosine-absorption around 280 nm [16,22]. However, in the presence of sodium dodecylsulfate (0.4 % w/v), the scattering diminished and characteristic absorption signals of the peptide bond (below 230 nm [23,24]) and tyrosine (around 280 nm [16,22]) appeared.

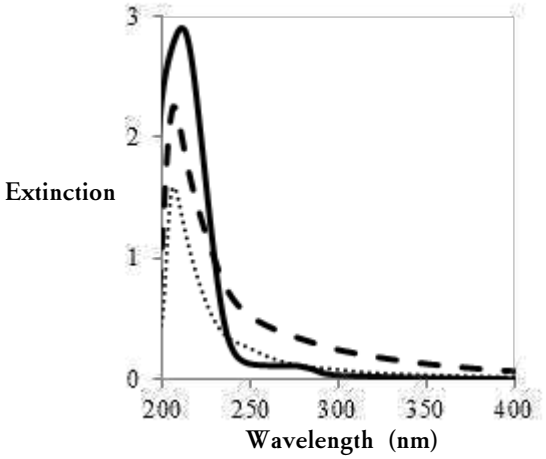

Figure 8. UV/vis absorption spectra of P_5S_3(H) under different solution conditions. The polypeptide was expressed as pelB-SilL fusion. After dissolving lyophilized polypeptide samples in 50 mmol/L sodium hydroxide solution, samples were diluted and neutralized by hydrochloric acid. If desired, SDS and Tris-HCl pH 7.5 were added. Measurement conditions were: 0.4 % w/v SDS, 2.6 mmol/L sodium chloride, neutral pH (*solid line*); 2.6 mmol/L sodium chloride, 20 mmol/L Tris-HCl pH 7.5 (*dashed line*) and 2.6 mmol/L sodium hydroxide, basic pH (*dotted line*). Since different polypeptide preparations were used, the peptide concentrations were only similar in the latter two samples.

The self-assembly of P_xS_y peptide versions at neutral pH 7.5 was further examined by dynamic light scattering (DLS). Preparations of tyrosine-free P_5S_3(-Y,H) and P_3S_1(-Y,H) as well as tyrosine-containing P_5S_3(H) and P_3S_1(H) were subjected to DLS (Figure 9) and their concentration was checked to be in a similar range by applying equal volumes of the solutions to SDS-PAGE (Figure 10). All DLS-profiles showed the presence of structures with both small (1 - 10 nm) and large (> 100 nm) hydrodynamic radii, at sodium chloride concentrations of 5, 15 or 255 mmol/L. Thus, a self-assembly was indicated at all sodium chloride concentrations investigated. The DLS-profiles were only interpreted in terms of intensity weight distribution. That is, in comparison of different aggregate sizes, small aggregates are underestimated in number since larger aggregates scatter light more efficiently (the intensity I of the scattered light is considered to be proportional to the sixth power of the hydrodynamic radius, R_H^6, for spherical aggregates), but to convert the profiles, a form factor would be required [25]. In case of P_xS_y, consecutive recordings resulted in different spectra, which is an indication for polydispersity or unstable assemblies, so a form factor cannot be assigned. Therefore, the occurrence of large assemblies was indicated, but no in depth analysis of size distributions was done.

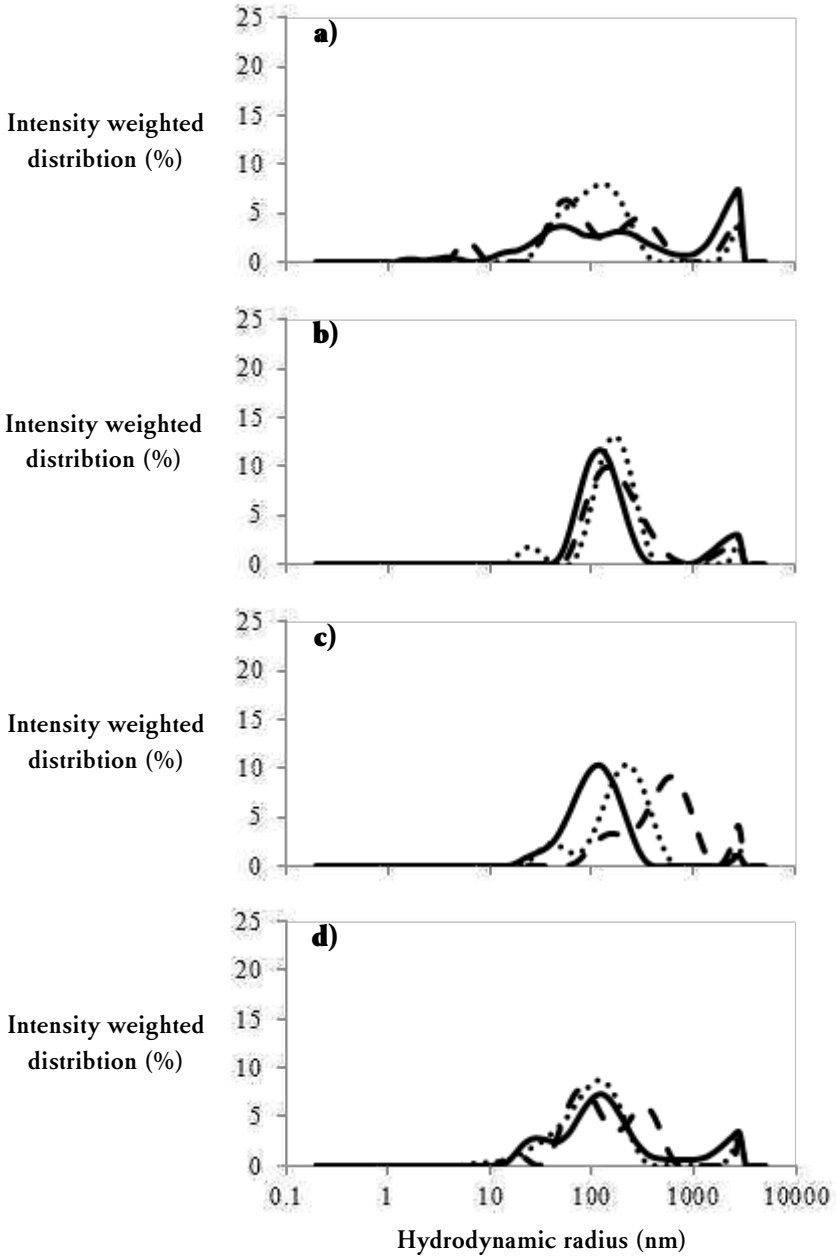

Figure 9. Exemplary dynamic light scattering profiles of P_xS_y(H) with and without tyrosine. All peptide versions were expressed as charge-compensated pelB-SilL fusion. Profiles were recorded for (*a*) P_3S_1(-Y,H) and (*b*) P_3S_1(H), as well as (*c*) P_5S_3(-Y,H) and (*d*) P_5S_3(H). Lyophilized polypeptide samples were dissolved in 50 mmol/L sodium hydroxide and subsequently diluted ten-fold, thereby neutralized with hydrochloric acid, buffered to 50 mmol/L Tris-HCl pH 7.5, and sodium chloride was added as desired (additional to 5 mmol/L sodium chloride from dissolution in sodium hydroxide and subsequent neutralization). Spectra were recorded at sodium chloride concentrations of 5 mmol/L (*solid line*), 15 mmol/L (*dotted line*) and 255 mmol/L (*dashed line*). Polypeptide concentrations were not exactly determined since in part tyrosine-free versions were examined, however, all peptide solutions were similarly concentrated as judged from SDS-PAGE (Figure 10).

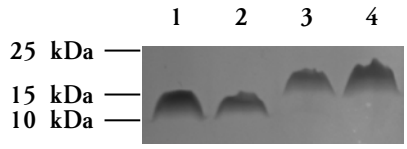

Figure 10. SDS-PAGE analysis of P_xS_y for estimation of the polypeptide concentration. Lyophilized polypeptides were dissolved in 50 mmol/L sodium hydroxide solution and each 1 µl per sample was subjected to SDS-PAGE. The lanes contain: P_3S_1(-Y,H) (*1*), P_3S_1(H) (*2*), P_5S_3(-Y,H) (*3*) and P_5S_3(H) (*4*).

P_5S_3(H) and P_5S_3, both expressed as KSI-fusion, were analyzed at a fixed concentration of 50 µmol/L to study the effects of the hexahistidine-tag and different pH environments on the self-assembly of the polypeptides. The intensity-weighted distribution profiles showed aggregates above 100 and above 1,000 nm in all conditions (pH 4.5 to 9.5) adjacent to signals in the 1 - 10 nm range which may derive from individual peptide chains. However, again a large variability in consecutive DLS-measurements was noted, indicating polydispersity.

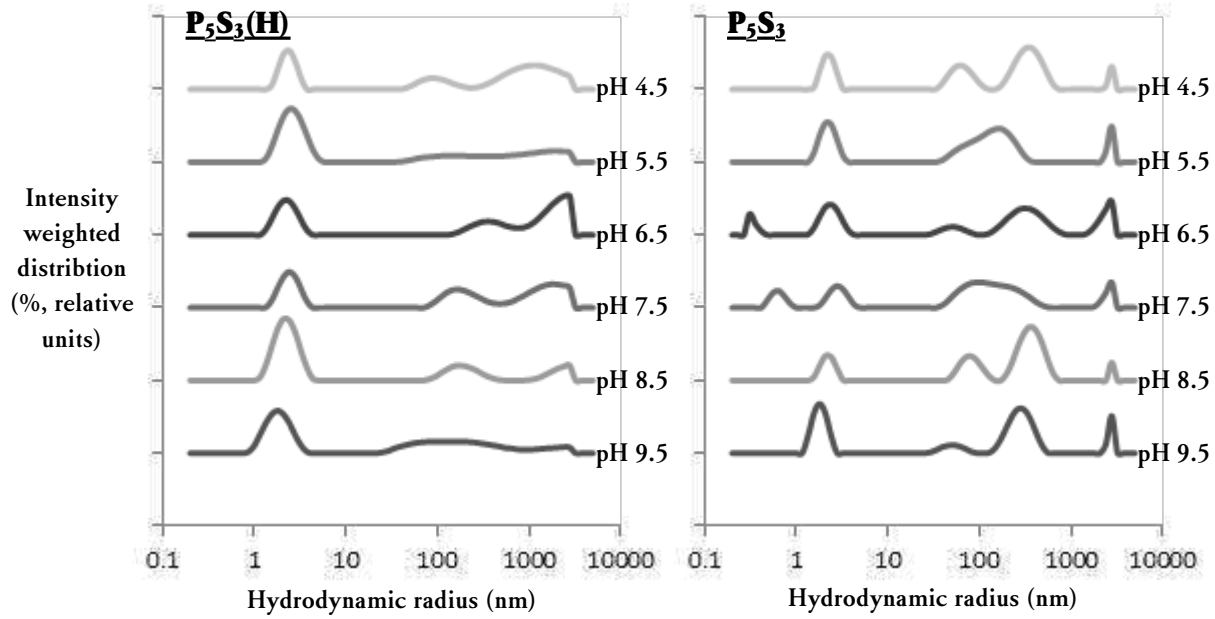

Figure 11. Exemplary dynamic light scattering profiles of hexahistidine-tagged and untagged P_5S_3(H) and P_5S_3. The polypeptides were expressed as KSI fusion. Both P_5S_3(H) (*left spectra*), and P_5S_3 (*right spectra*) were examined at different pH values, which were achieved by adding each 50 mmol/L of pre-adjusted buffers sodium acetate (pH 4.5), Bis-Tris-HCl (pH 5.5 and 6.5) or Tris-HCl (pH 7.5, 8.5, 9.5). Polypeptide concentrations were 50 µmol/L.

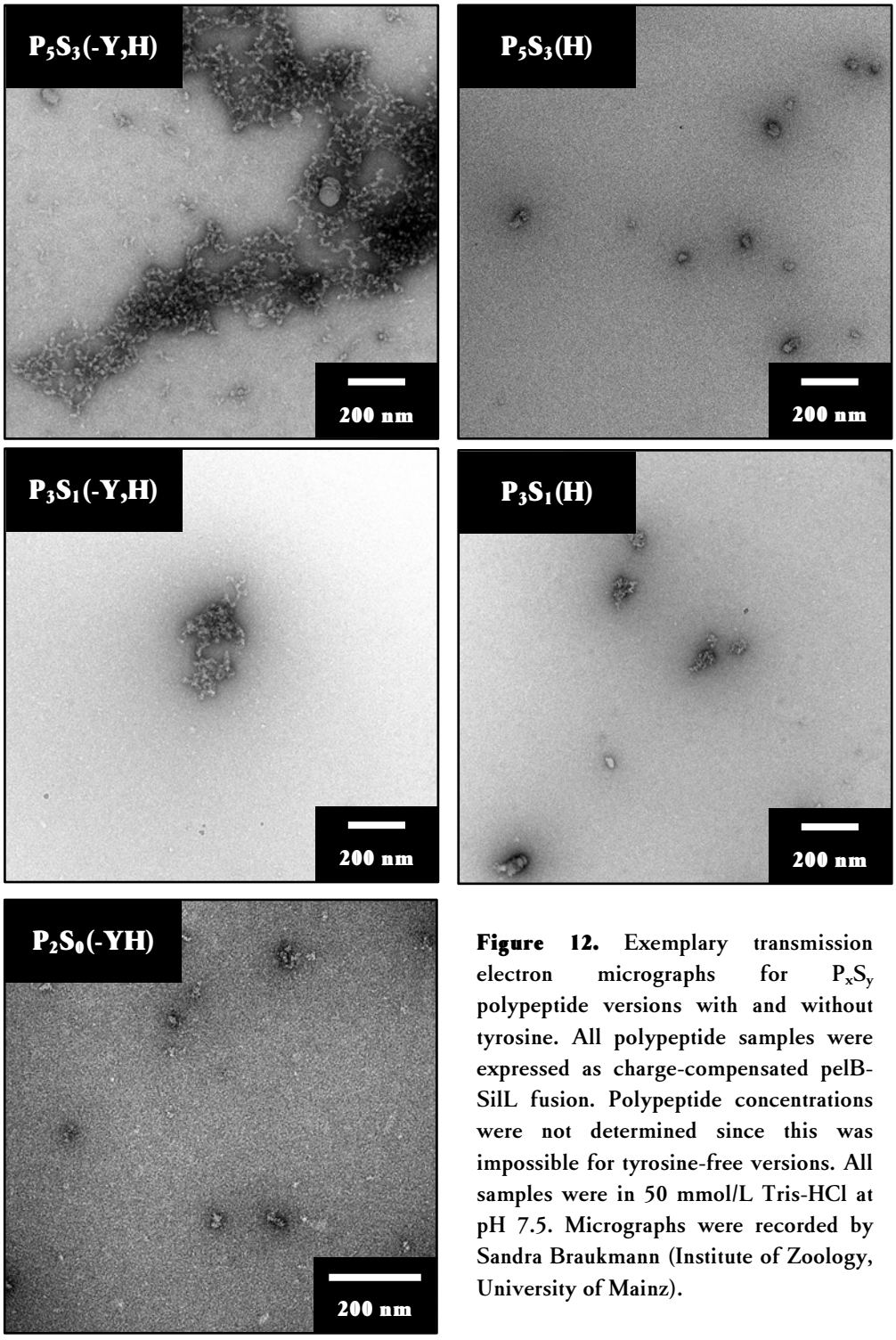

Figure 12. Exemplary transmission electron micrographs for P_xS_y polypeptide versions with and without tyrosine. All polypeptide samples were expressed as charge-compensated pelB-SilL fusion. Polypeptide concentrations were not determined since this was impossible for tyrosine-free versions. All samples were in 50 mmol/L Tris-HCl at pH 7.5. Micrographs were recorded by Sandra Braukmann (Institute of Zoology, University of Mainz).

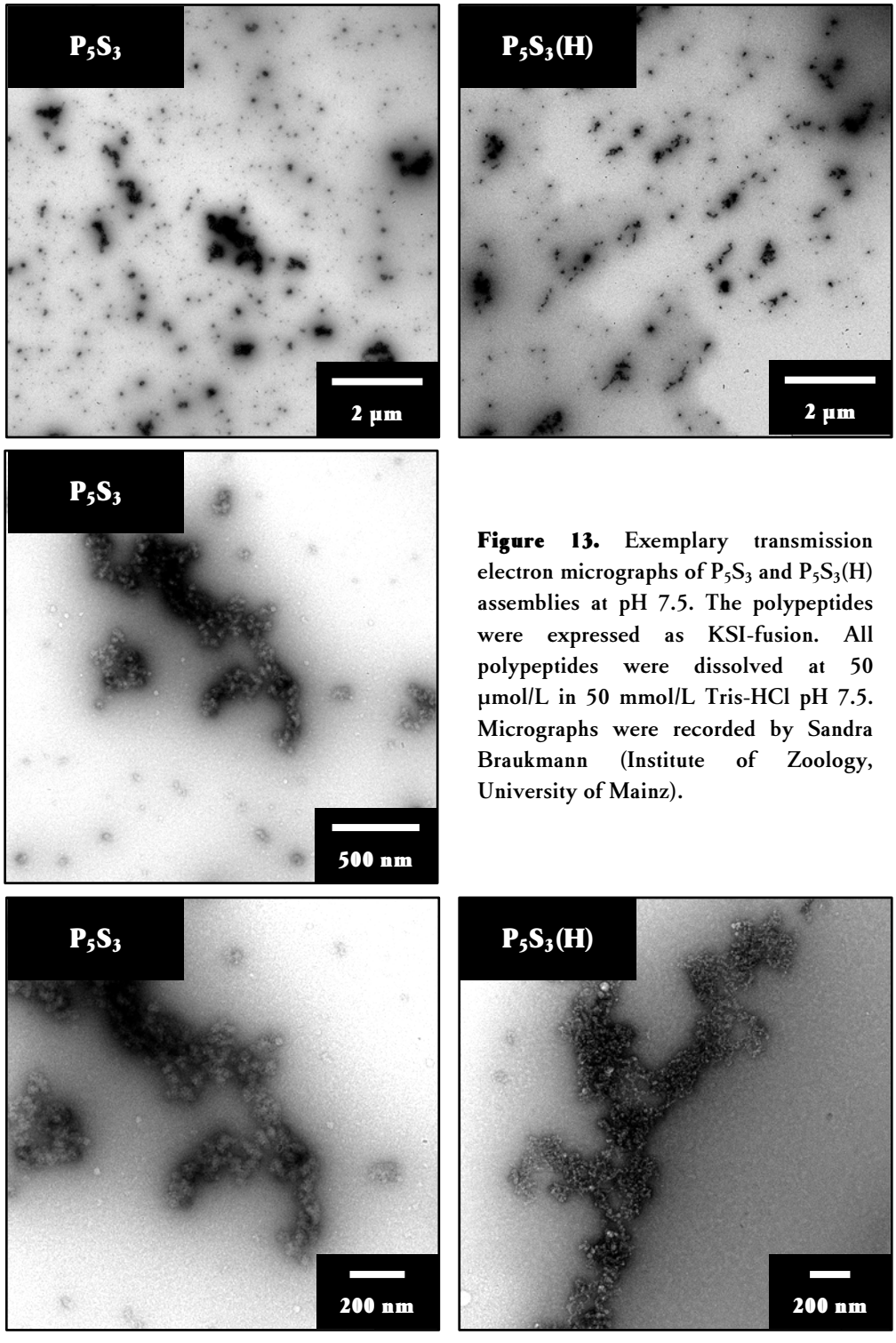

Figure 13. Exemplary transmission electron micrographs of P_5S_3 and $P_5S_3(H)$ assemblies at pH 7.5. The polypeptides were expressed as KSI-fusion. All polypeptides were dissolved at 50 µmol/L in 50 mmol/L Tris-HCl pH 7.5. Micrographs were recorded by Sandra Braukmann (Institute of Zoology, University of Mainz).

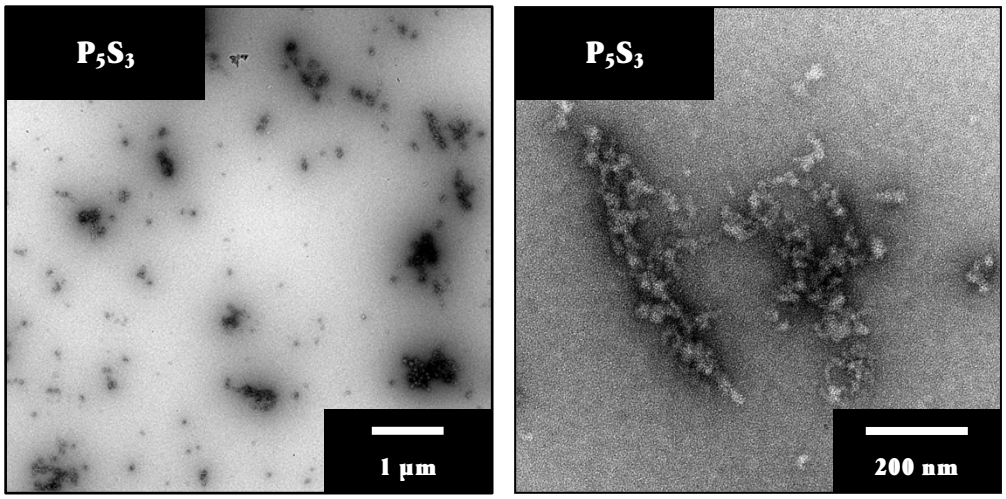

Figure 14. Exemplary transmission electron micrographs for P_5S_3 assembly at pH 9.5. The polypeptide was expressed as KSI-fusion. Concentration of P_5S_3 was 50 µmol/L, in 50 mmol/L Tris-HCl. Micrographs were recorded by Sandra Braukmann (Institute of Zoology, University of Mainz).

Transmission electron microscopy (TEM) analyzes gave further insight into the self-assembly of P_xS_y polypeptides (Figures 12, 13 and 14). Assemblies were examined after adsorption to hydrophilized carbon coated copper grids and negative staining by uranyl acetate or formiate. At pH 7.5, especially P_5S_3(-Y,H) was found to be highly self-assembled (Figures 12 and 13). Structures were found in the range below 50 nm (both spherical and wormlike extended) which were further assembled into structures of several hundreds of nanometers. The shorter polypeptide P_3S_1(-Y,H) (Figure 12) assembled into similar structures which however did not reach the extensions achieved by P_5S_3(-Y,H). Introduction of a tyrosine in both cases led to smaller structures, which were mostly below 50 nm in extension and seemed denser, especially for P_5S_3(H). The shortest version P_2S_0 (Figure 12) as well showed some assemblies below 50 nm, though large extended structures were not seen. A tyrosine-containing version of P_2S_0 was not examined. The self-assembly of P_5S_3 versions was similar irrespective whether they were derived from a pelB-SilL charge-compensated or KSI inclusion body-inducing fusion (Figures 12 and 13), whether they bore a hexahistidine tag (Figure 13) or whether they were analyzed at pH 7.5 or 9.5 (Figures 13 and 14), in accordance with the DLS results.

Discussion

Structural characterization of different P_xS_y versions revealed that these polypeptides, bearing numerous cationic side chains, preferentially adopt an extended polyproline II conformation [7], a typical structural motif of polypeptides with high net charge [6,9] which are often classified as "intrinsically disordered" [1]. The polyproline II (PPII) conformation can be identified from CD-spectra by a strong negative ellipticity in the 195 to 200 region. It is usually associated with a positive ellipticity around 220 nm, which however in some polypeptides only occurs at low temperature [7,26,27] and may be superimposed when α-helices and β-sheets are present as

folding motifs, since these have a negative ellipticity in this region [20]. The CD-spectra of P_5S_3 at room temperature and at different pH were negative below around 240 nm and exhibited a negative peak in the 195 to 200 nm region, indicating PPII. A similar CD-spectrum was reported for the silaffin-derived peptide R5, for which nuclear magnetic resonance spectroscopy of the dissolved peptide further indicated an ensemble of different unordered structures [28].

The conformation of P_xS_y as judged from the CD-spectrum was not affected by the presence of 50 mmol/L of the salts studied (chloride-salts of sodium, calcium, ammonium and guanidium; phosphate-buffer adjusted with sodium hydroxide). In the case of phosphate, this concentration sufficed to induce the assembly of silaffins, so that a response in the CD-spectra of P_xS_y could be expected in case of an interaction. At an increased sodium chloride concentration of 1,000 mmol/L, the PPII-typical spectral features were reduced in P_xS_y. A similar behavior was described for poly-L-glutamate and poly-L-lysine [29,30] as well as heptamers of lysine and proline [31]. Comparison of different salts revealed that the CD-spectral changes are due to chaotropic effects [30] which e.g. chloride anions excert at high concentrations [32,33]. An additional effect of salts as sodium chloride on high-charged polypeptides as P_5S_3 could derive from screening of side-chain charges by anions, thereby reducing the repulsion exerted by like-charged side chains [34,35]. Since this repulsion is a driving force for the adoption of the extended PPII structure [5], this might hence result in a conformational change.

All P_xS_y versions had an intrinsic tendency to self-assemble, without the necessity of polyvalent anions as mediator for electrostatic interactions as described for polycationic silaffins [12]. The dynamic light scattering (DLS) profiles of different P_5S_3 and P_3S_1 versions at different pH contained signals with hydrodynamic radii of several hundreds of nanometers, and in part also structures in the low nanometer range. Transmission elelctron microscopy (TEM, surface-adsorption of P_xS_y, negative stain mode) revealed structures around 50 nm which in part were assembled into larger strucutures. The variety of different assemblies seen in TEM prevents an in-depth analysis of the DLS-profiles, since the evaluation of the relative number of structures with different sizes from DLS requires a form factor [25], which is difficult to define for an ensemble of inhomogenous structures. However, it remains unclear whether the structures seen in TEM will also form in solution. Since the DLS profiles were unstable with time (i.e. differed in consecutive measurements), the assemblies may be too diverse, but could also only form transiently.

The mechanism behind the P_xS_y self-assembly remains to be elucidated. As a hypothesis, the high number of charged side chains in combination with the extended conformation could allow a side-chain (especially arginine) to main-chain (peptide bonds) interaction. This type of interaction is common in the folded state of several proteins [36] and for example contributes to the binding of peptide ligands to the human major histocompatibility complex which presents peptide-antigens to T-cells in a PPII-like conformation, in order to cause an immune response [37]. In tyrosine-bearing versions, additional cation-π interactions between lysine or arginine and the aromatic tyrosine [38], wich occur in proteins to a considerable extent [39], are possible, and

this interaction was sufficiently strong to mediate the fibrillization e.g. of the octameric peptide *Ac*-A_4KA_2Y [40] or the collagen-related peptide $RG(POG)_mF$, which assembled by head-to-tail interactions [41] (*Ac*: acetylation, A: alanine, F: phenylalanine, G: glycine, K: lysine, O: hydroxyproline, P: proline, R: arginine, Y: tyrosine, "m": repeat number (8 or 10)).

It is interesting to note that DLS indicated the peptide to remain assembled at sodium chloride concentrations of 250 mmol/L. This behavior may either hint to a high number of hydrogen bonds which in sum are stable at increased salt concentration, or to additional hydrophobic interactions (e.g. by leucine residues in P_xS_y) which are rather reinforced by increasing the salt concentration [42].

To understand the actual mechanism in P_xS_y-assembly, an analysis of the polypeptides by nuclear magnetic resonance (NMR) spectroscopy is intended in order to identify the amino acid residues which mediate the interaction.

References

[1] V.N. Uversky, J.R. Gillespie, A.L. Fink, Proteins 41 (2000) 415.

[2] R. van der Lee, M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D.T. Jones, P.M. Kim, R.W. Kriwacki, C.J. Oldfield, R. V Pappu, P. Tompa, V.N. Uversky, P.E. Wright, M.M. Babu, Chem. Rev. 114 (2014) 6589.

[3] L. Kalmar, D. Homola, G. Varga, P. Tompa, Bone 51 (2012) 528.

[4] E.E. Oren, R. Notman, I.W. Kim, J.S. Evans, T.R. Walsh, R. Samudrala, C. Tamerler, M. Sarikaya, Langmuir 26 (2010) 11003.

[5] A.H. Mao, S.L. Crick, A. Vitalis, C.L. Chicoine, R. V Pappu, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 8183.

[6] M.L. Tiffany, S. Krimm, Biopolymers 6 (1968) 1379.

[7] A.A. Adzhubei, M.J.E. Sternberg, A.A. Makarov, J. Mol. Biol. 425 (2013) 2100.

[8] Y.P. Myer, Macromolecules 2 (1969) 624.

[9] A. V Mikhonin, N.S. Myshakina, S. V Bykov, S.A. Asher, J. Am. Chem. Soc. 127 (2005) 7712.

[10] L. Ma, Z. Ahmed, S.A. Asher, J. Phys. Chem. B 115 (2011) 4251.

[11] B.J. McKenna, H. Birkedal, M.H. Bartl, T.J. Deming, G.D. Stucky, Angew. Chemie 116 (2004) 5770.

[12] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, Science 298 (2002) 584.

[13] E. Brunner, K. Lutz, M. Sumper, Phys. Chem. Chem. Phys. 6 (2004) 854.

[14] H. Cui, V. Krikorian, J. Thompson, A.P. Nowak, T.J. Deming, D.J. Pochan, Macromolecules 38 (2005) 7371.

[15] M.M. Tomczak, D.D. Glawe, L.F. Drummy, C.G. Lawrence, M.O. Stone, C.C. Perry, D.J. Pochan, T.J. Deming, R.R. Naik, J. Am. Chem. Soc. 127 (2005) 12577.

[16] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, Protein Sci. 4 (1995) 2411.

[17] H. Schaegger, G. von Jagow, Anal. Biochem. 166 (1987) 368.

[18] G. von Jagow, H. Schaegger, A Practical Guide to Membrane Protein Purification, Academic Press, San Diego, 1994.

[19] H. Schägger, Nat. Protoc. 1 (2006) 16.

[20] N.J. Greenfield, Nat. Protoc. 1 (2006) 2876.

[21] C. Zerfaß, S. Braukmann, S. Nietzsche, S. Hobe, H. Paulsen, Protein Expr. Purif. 108 (2015) 1.

[22] F.C. Smith, Proc. R. Soc. B Biol. Sci. 104 (1929) 198.

[23] A.R. Goldfarb, L.J. Saidel, Science 114 (1951) 156.

[24] A.R. Goldfarb, L.J. Saidel, E. Mosovich, J. Biol. Chem. 193 (1951) 397.

[25] W. Schärtl, Light Scattering from Polymer Solutions and Nanoparticle Dispersions, Springer, Berlin and Heidelberg, 2007.

[26] M.L. Tiffany, S. Krimm, Biopolymers 11 (1972) 2309.

[27] A.F. Drake, G. Siligardi, W.A. Gibbons, Biophys. Chem. 31 (1988) 143.

[28] L. Senior, M.P. Crump, C. Williams, P.J. Booth, S. Mann, A. Periman, P. Curnow, J. Mater. Chem. B 3 (2015) 2607.

[29] K. Xiong, L. Ma, S.A. Asher, Biophys. Chem. 162 (2012) 1.

[30] M.L. Tiffany, S. Krimm, Biopolymers 8 (1969) 347.

[31] A.L. Rucker, T.P. Creamer, Protein Sci. 11 (2002) 980.

[32] M.G. Cacace, E.M. Landau, J.J. Ramsden, Q. Rev. Biophys. 30 (1997) 241.

[33] Y. Zhang, P.S. Cremer, Annu. Rev. Phys. Chem. 61 (2010) 63.

[34] A. Norouzy, K.I. Assaf, S. Zhang, M.H. Jacob, W.M. Nau, J. Phys. Chem. B 119 (2015) 33.

[35] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, B. Schuler, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14609.

[36] C.L. Worth, T.L. Blundell, BMC Evol. Biol. 10 (2010) 161.

[37] T.S. Jardetzky, J.H. Brown, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, D.C. Wiley, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 734.

[38] D.A. Dougherty, Science 271 (1996) 163.

[39] J.P. Gallivan, D.A. Dougherty, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9459.

[40] T.J. Measey, K.B. Smith, S.M. Decatur, L. Zhao, G. Yang, R. Schweitzer-Stenner, J. Am. Chem. Soc. 131 (2009) 18218.

[41] C.C. Chen, W. Hsu, T.C. Kao, J.C. Horng, Biochemistry 50 (2011) 2381.

[42] H. Dong, S.E. Paramonov, L. Aulisa, E.L. Bakota, J.D. Hartgerink, J. Am. Chem. Soc. 129 (2007) 12468.

Section III - High yield recombinant production of a self-assembling polycationic peptide for silica biomineralization

Christian Zerfaß[a,b], Sandra Braukmann[c], Sandor Nietzsche[d], Stephan Hobe[a], Harald Paulsen[a]

[a] Institute of General Botany, Johannes Gutenberg University, Müllerweg 6, 55128 Mainz, Germany
[b] Graduate School Materials Science in Mainz, Staudinger Weg 9, 55128 Mainz, Germany
[c] Institute of Zoology, Johannes Gutenberg University, Müllerweg 6, 55128 Mainz, Germany
[d] Center for Electron Microscopy, University Hospital, Ziegelmühlenweg 1, 07743 Jena, Germany

Corresponding author: Harald Paulsen
Phone: +49-6131-3924633
Fax: +49-6131-3924670
E-mail: paulsen@uni-mainz.de

This manuscript was published in "Protein Expression and Purification" with the reference:

Zerfaß, C.; Braukmann, S.; Nietzsche, S.; Hobe, S.; Paulsen, H.
Protein Expression and Purification **2015**, 108, 1 - 8

Highlights

- Strongly basic biomineralization peptides are difficult to synthesize in vitro.
- Recombinant biomineralizing peptide is expressed at high yield (15 umol per liter of culture).
- The polycationic peptide self-assembles and precipitates silica in the absence of anions.
- Enzymatic phosphorylation affects the biomineralizing activity.

Abbreviations

KSI: Ketosteroid isomerase; PKA: Protein kinase A; Tris: Tris(hydroxymethyl)aminomethane; Bis-Tris: Bis(2-hydroxyethyl)aminotris(hydroxymethyl)methane.

Abstract

We report the recombinant bacterial expression and purification at high yields of a polycationic oligopeptide, P_5S_3. The sequence of P_5S_3 was inspired by a diatom silaffin, a silica precipitating peptide. Like its native model, P_5S_3 exhibits silica biomineralizing activity, but furthermore has unusual self-assembling properties. P_5S_3 is efficiently expressed in *Escherichia coli* as fusion with ketosteroid isomerase (KSI), which causes deposition in inclusion bodies. After breaking the fusion by cyanogen bromide reaction, P_5S_3 was purified by cation exchange chromatography, taking advantage of the exceptionally high content of basic amino acids. The numerous cationic charges do not prevent, but may even promote counterion-independent self-assembly which in turn leads to silica precipitation. Enzymatic phosphorylation, a common modification in native silica biomineralizing peptides, can be used to modify the precipitation activity.

Keywords

Polycationic peptide; Ketosteroid isomerase fusion; Silica biomineralization

Introduction

Silaffins are highly post-translationally modified polypeptides used by diatoms to deposit silica when building their mineralized cell wall [1]. Silaffins are also capable of silica precipitation *in vitro*, either in their highly modified native state [2,3] or in non-modified recombinant and synthetic versions [4-6]. Since their discovery, much attention has been drawn to biomimetic, peptide-controlled mineral formation [7,8]. Hydrophilic and especially cationic amino acids are characteristic key factors for silicification by silaffins [2,9-12] and also serve as targets for post-translational modification. Phosphorylations, sulfations, as well as glycosylations with acidic sugar-compounds [3,9,12] introduce anionic charges which mediate electrostatic self-assembly of silaffins, a prerequisite for their precipitation activity. However, presence of polyanions such as phosphate and sulfate free in solution can compensate the lack of the covalent modification [3]. Lysines often are post-translationally methylated or connected with polyamines [2,13], increasing the silaffin's cationic charge and hence its interaction with weakly acidic silica. Polyamines themselves are able to precipitate silicic acid and are not necessarily bound to a peptide compound in diatoms [14]. Vice versa, the polycationic silaffin peptide part without polyamines attached can as well be active in biomineralization, but with a different pH optimum [2].

Short silaffin-polypeptides with and without modifications have been produced by solid phase peptide synthesis [5,6,15,16]. However, due to incomplete coupling reactions, yields of chemical syntheses will decrease with increasing peptide size [17]. Recombinant protein expression in *Escherichia coli* is hence an attractive alternative to achieve larger polypeptides and proteins, of uniform length and sequence [18], but post-translational modifications generally have to be performed after protein expression and purification. Among the silaffin-modifications,

phosphorylations can easily be introduced by utilizing kinases. This modification changes the peptide net charge and affects its assembly-state by mediating electrostatic interactions [3]. Silaffins have been partially phosphorylated *in vitro* by a silaffin-specific diatom kinase [19]. Other peptides with biomineralizing activities have been phosphorylated by a variety of commercially available kinases [20-22] or extracts from pig golgi apparatus which contains a set of mammalian kinases [23]. To study the general impact of phosphorylations on biomineralizing peptides, recombinant silaffins can be engineered to contain kinase recognition sites. Especially kinases that phosphorylate sequences which contain basic amino acid residues, such as protein kinase A (PKA) [24,25], appear suitable for the highly cationic silaffins.

Recombinant production of polypeptides with numerous cationic charges is usually challenging in bacterial expression systems. Presumably, this is due to interactions with anionic cellular compounds such as nucleic acids or anionic membrane lipids, which render such polypeptides toxic to bacteria [26,27]. Additionally, these polypeptides may become degraded by cellular proteases [28]. The polycationic domain of the silaffin precursor-protein Sil1p from *Cylindrotheca fusiformis* has been expressed in *E.coli*, albeit with yields in the low milligram range per liter of culture [4]. In the case of cationic, amphiphilic antimicrobial peptides [29,30], fusion strategies have been established for detoxification and overexpression in bacteria [31,32]. One efficient fusion is ketosteroid isomerase (KSI) [33], a hydrophobic protein that induces the formation of inclusion bodies. This fusion was established with a small peptide mating factor from yeast [34], and efficiently adapted for a hydrogel-forming peptide [35,36], antimicrobial partial sequences from bovine milk lactophoricin [37], antimicrobial human dermcidin [38], an aggregating amyloid protein [39], a highly-charged antimicrobial peptide with sequence PFWRIRIRR [40] as well as an intrinsically disordered domain of human cardiac troponin [41].

In this contribution, we report on the expression of a designed polycationic, biomineralizing polypeptide, P_5S_3, by fusing it to ketosteroid isomerase (KSI). Introduction of a kinase target side allows specific *in vitro* phosphorylation with protein kinase A (PKA). P_5S_3 is modeled after *C. fusiformis* silaffin, bearing numerous basic amino acid side chains, which are either lysine-clusters as in silaffin, or arginine-doublets being part of the PKA phosphorylation site.

Materials and methods

Gene design and plasmid construction

We designed a polypeptide based on silaffin motifs from *C. fusiformis* [2] (Genbank AF191634.1). Numerous construction details are of no special importance for the present work but served other purposes. To keep this section compact, only the major steps will be described here. Readers interested in the detailed cloning procedure are referred to ESI. The cloning procedures were done following standard protocols for molecular biology procedures [42]. All

restriction enzymes were from New England Biolabs GmbH (Frankfurt (Germany)) and used with appropriate buffers according to the supplier.

Gene construction was achieved by ligating three primer hybrids (PIN 1, 2, and 3, all primer sequences are given in Table 1). Corresponding primer pairs (sense and antisense, each at 40 pmol/l) were hybridized by denaturation (10 min at 100 °C) and subsequent cooling to room temperature, in annealing buffer composed of 40 mmol/l Tris/HCl pH 7.5, 20 mmol/l magnesium chloride and 50 mmol/l sodium chloride. To 5′ phosphorylate the DNA double strands, hybridization samples were diluted 10-fold with buffer to achieve a final buffer composition of 9 mmol/l Tris/HCl pH 7.5, 3 mmol/l magnesium chloride, 5 mmol/l sodium chloride, 1 mmol/l dithiothreitol and 1.1 mmol/l adenosine triphosphate, and reaction with T4 polynucleotide kinase (2 units, New England Biolabs GmbH, Frankfurt (Germany)) was run for 1 h at 37 °C. Afterwards, the enzyme was heat-inactivated (20 min at 65 °C). Each of the primer hybrids contained single strand ends complementary to its neighbor to allow for oriented sticky end ligation (ESI, Fig. S1A). Hybrid PIN 2 was intended to be inserted multiple times to create repetitive sequences with variable lengths. Therefore, it was self-ligated for 30 min at 16 °C (T4 DNA Ligase, New England Biolabs GmbH, Frankfurt (Germany)), before hybrids PIN 1, 3 and (pre-ligated) PIN 2, together with the target plasmid were combined. The primary target plasmid was pET22b(+) (Merck Chemicals GmbH, Schwalbach (Germany)). Because of an earlier insertion into the multiple cloning site, the plasmid was double-digestable with *Bam*HI and *Xho*I for sticky end ligation with the corresponding 5′ and 3′ end of the assembled PIN construct. For ligation, the components were combined at molar ratios plasmid : PIN 1: PIN 2: PIN 3 of 1:2:6:2 with T4 DNA ligase. The new construct had *Ava*I restriction sites immediately 3′ and some bases 5′ of the P_5S_3-coding region for excision from the plasmid and insertion into the *Ava*I-digested expression plasmid pET 31b(+) (Merck Chemicals GmbH, Schwalbach (Germany)) which contains a ketosteroid isomerase (KSI) coding region upstream of the multiple cloning site, later serving as N-terminal fusion for recombinant fusion protein expression. The plasmid map and sequence of the expression vector containing KSI- P_5S_3 are shown in Fig. 1 and ESI Fig. S1, respectively. Please note that some single site mutageneses (codon exchange, codon insertion) were made after gene construction and are described in detail in Supplementary Information. Plasmid constructs were used to transform *E. coli* strain DH5-α (Life Technologies GmbH, Darmstadt (Germany)). All constructs were verified by sequencing (StarSeq GmbH, Mainz (Germany)) with primers T7PrompET22b(+) or pET22b(+)rev (Table 1)).

Protein expression

Protein expression was performed in *E. coli* strain Rosetta (Merck4Biosciences GmbH, Schwalbach (Germany)). All LB media [43] for protein expression were supplemented with 0.1 % (w/v) glucose, 100 μg/ml ampicillin and 34 μg/ml chloramphenicol.

Bacteria of an LB overnight culture (37 °C and 185 rpm shaking) were harvested (5 min at 5000*g*) and suspended in LB-medium at a 25-fold volume of the original culture. Cultures were grown (37 °C, 185 rpm) again to an optical density OD_{600} of 2.5 (early stationary phase) before expression was induced using 1 mmol/l isopropyl β-D-1-thiogalactopyranoside (IPTG) and continued overnight.

Table 1

Primer sequences used in this study. All primers were received from Biomers GmbH (Ulm (Germany)). Abbreviations: s - sense, as - antisense, f - forward, r - reverse.

Primer Name	Sequence (all: 5' to 3')	Function
PIN 1_s	GATCCTGTCGACGTGCTTCCTTA	Target gene, 5' region
PIN 1_as	CCCTAAGGAAGCACGTCGACAG	
PIN 2_s	GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA	Target gene, repetitive unit
PIN 2_as	CCCTAGGCTAGCGCGACGCAGCTTTTTAGATTT	
PIN 3_s	GGGAAGCTGCGTCGCGCTAGCC	Target gene, 3' region
PIN 3_as	TCGAGGCTAGCGCGACGCAGCTT	
T7PrompET22b(+)fw	TAATACGACTCACTATAGGG	Sequencing pET22 b(+)
pET22b(+)rev	GCTCAGCGGTGGCAGC	Sequencing pET31 b(+)

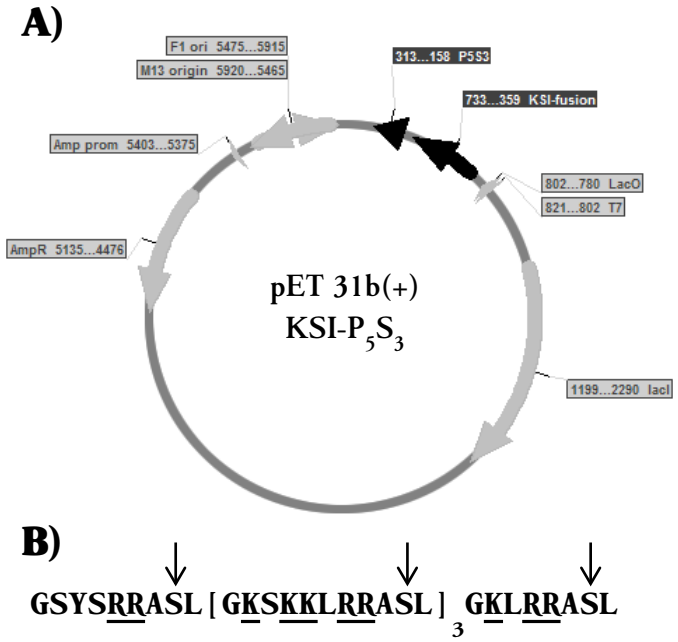

Figure 1

P_5S_3 expression vector and target sequence. **A)** Plasmid scheme of KSI-fused P_5S_3 in pET31b(+). **B)** Amino acid sequence of P_5S_3. Basic amino acids are underlined, serines within a protein kinase A target site are marked with arrows.

Protein purification

Cells from 1 l expression culture were harvested (5 min, 5000*g*), re-suspended in 50 ml lysis buffer (50 mmol/l Tris/HCl pH 8, 1 mmol/l ethylenediaminetetraacetic acid) and treated with lysozyme (0.5 mg/ml, 30 min at 37 °C). After addition of 2 mmol/l magnesium chloride and 20 mmol/l dithiothreitol, DNA-digestion was performed with Benzonase nuclease (Merck Chemicals GmbH, Schwalbach (Germany), 50 units, 30 min at 37 °C). The mixture was centrifuged 10 min at 17,500*g* and the pellet was thoroughly washed in Triton washing buffer adapted from Kyle et al. [35] (0.1 % (w/v) Triton X-100, 2.5 % (w/v) saccharose, 0.5 mmol/l magnesium chloride, 5 mmol/l Tris/HCl pH 7.5) supported by alternate ultrasonic bath incubation and vortexing (until pellet was dispersed) and subsequently centrifuged again (10 min at 10,000*g*). The precipitate was dissolved in 50 ml 70 % (v/v) formic acid for cyanogen bromide cleavage [44-46]. Nitrogen was flushed in before adding 1 g cyanogen bromide, dissolved in 1 ml acetonitrile. Cleavage was performed overnight, under constant stirring and light-protection. After evaporating the solvent, the polycationic target protein was separated from the insoluble KSI-fusion by suspending the precipitate in 50 ml of 50 mmol/l sodium hydroxide and centrifugation for 10 min at 10,000*g*. The supernatant was buffered by adding an equal volume of 40 mmol/l Tris/HCl pH 7.5, and the target polypeptide P_5S_3 was isolated by cation exchange chromatography on SP-Sepharose (GE-Healthcare GmbH, München (Germany)). The resin was washed and the target polypeptide eluted with 350 and 500 mmol/l ammonium carbonate (Merck Chemicals GmbH, Schwalbach (Germany)) solution, respectively, and the eluate was lyophilized. Peptide yields were usually determined by UV/vis spectrophotometry in 1 % (w/v) SDS applying the molar extinction coefficient for one tyrosine of 1490 l mol^{-1} cm^{-1} [47]. Peptide identity and purity was confirmed by total amino acid analysis (Genaxxon Bioscience GmbH, Ulm (Germany)).

SDS-PAGE analyses

Electrophoresis was done as described [48,49] using the Tris/Tricine system with 16.5 % acrylamide total concentration in the separating gel.

Phosphorylation with protein kinase A

Assays contained 100 mmol/l Tris/HCl pH 7.5, 2 mmol/l adenosine triphosphate (sodium salt), 1 mmol/l magnesium chloride and 100 µmol/l polypeptide. Protein kinase A (mouse, catalytic subunit, Biaffin GmbH, Kassel (Germany)) was added at an activity to complete the reaction within one hour. The phospho-polypeptide was isolated by cation exchange chromatography. The assay was diluted 1:1 with water and applied to SP-Sepharose (GE-Healthcare GmbH, München (Germany)). Washing and elution was done with ammonium carbonate solution of 50 mmol/l and 500 mmol/l, respectively, if preparative scale purification was desired. For analytical purposes of

binding strength to the resin, smaller increments in concentration were done which in this case is noted.

Silica precipitation assay

All experiments were done with 30 mmol/l silicic acid at a total volume of 50 µl, resulting in 1500 nmol total silicic acid available. As silicic acid source, either tetramethoxysilane (TMOS, Sigma-Aldrich GmbH, Fluka, Taufkirchen (Germany)) hydrolyzed with hydrochloric acid (1 mmol/l, 15 min at room temperature with shaking at 1400 rpm), or water glass (sodium metasilicate, C. Roth GmbH, Karlsruhe (Germany)) neutralized with an equal volume of hydrochloric acid in appropriate concentration, immediately before addition to the precipitation assay, was used. To maintain the pH during precipitation, samples contained 50 mmol/l of either Bis-Tris (pH 6.5) or Tris (pH 7.5, 8.5 or 9.5) adjusted with HCl to the desired pH. P_5S_3 was present at various concentrations. Precipitation was allowed to proceed for 30 min at 23 °C with shaking at 1400 rpm, after which precipitates were collected by centrifugation (5 min at 18400*g*), resuspended in water for washing and again sedimented by centrifugation. The water used was ultra-pure (resistance $R > 18$ MΩ). For scanning electron microscopy, precipitates were again suspended in water and a fraction placed on a glass slide, air-dried, and analyzed with a field emission electron microscope LEO-1530 (Zeiss, Oberkochen, Germany). To prevent surface charging, the samples were sputter coated with a 2 nm platinum layer using a MED 020 coating system (BAL-TEC, Liechtenstein). For quantitative analysis of precipitated silicic acid, samples were hydrolyzed in 1 mol/l sodium hydroxide (1 h at room temperature) and subjected to molybdenum blue analysis [50]. Fractions of hydrolyzed silica solutions were diluted to 200 µl with water. To build the yellow silicomolybdic acid complex, 4 µl of 10 % w/v ammonium heptamolybdate tetrahydrate in 1.6 mol/l sulfuric acid were added and samples were left undisturbed for 10 min. After addition of 4 µl 10 % w/v oxalic acid and 1 min incubation, 4 µl of 100 mmol/l ascorbic acid were added and samples were left stand for 15 min to reduce the silicomolybdic acid to the molybdenum blue complex. Reactions were done in 96 well-plates (Greiner Bio-ONE GmbH, Frickenhausen (Germany)), and the complex quantified spectroscopically (Plate Reader, Tecan GmbH, Mainz (Germany)) by measuring the absorbance at 810 nm and comparing this with a reference containing known amounts of silicic acid.

Transmission electron microscopy of polypeptide aggregates

Sample preparation for TEM was done as described earlier [51]. Briefly, peptides were bound to a glow-discharged carbon coated copper grid from a 50 µmol/l solution, in Tris/HCl buffer (50 mmol/l, pH 7.5). Negative staining was achieved with 1 % w/v uranyl acetate solution, after which the grid was washed three times with water. Between all steps, excess solution was blotted with tissue paper, and the sample was allowed to air-dry after the last step. TEM pictures were recorded on a Tecnai 12 electron microscope (FEI, Gräfeling, Germany).

Circular dichroism spectroscopy

To analyze the secondary structure of the polypeptide, CD-spectra were recorded on a JascoJ 810 spectropolarimeter (Jasco GmbH, Gross-Umstadt, Germany) equipped with a Jasco CDF-426S Peltier temperature controlling element. A 75 µmol/l polypeptide solution in 5 mmol/l Tris/HCl (pH 7.5) was measured in a quartz cuvette with 0.1 mm path length. Spectra were recorded in 1 nm steps, with 50 nm per minute and fivefold accumulation.

Results and discussion

The present paper reports the successful overexpression of the designed polypeptide P_5S_3 as a fusion construct with an inclusion body-inducing KSI-presequence [33,34] (Fig. 1, complete sequence in ESI Fig. S1). P_5S_3 is composed of five protein kinase A target sites (sequence LRRASLG [25]) interspersed with lysine-rich motifs (KSKK) that are typical for silaffins. After removal of the KSI-fusion (see below), the target polypeptide (50 amino acids, 5.6 kDa) contains 40 % lysine or arginine residues (10 residues each), leading to a pI of 12.6 and presumably exhibiting approximately 20 positive charges at pH 7.5. Serine (10 residues) and leucine (8 residues) make up for about another third of the primary structure.

High-yield expression of the cationic peptide P_5S_3

The KSI- P_5S_3 fusion was efficiently overexpressed in *E. coli* cells, as is readily seen by comparing the SDS-PAGE separated total cellular protein before and after induction (Fig. 2A, lane 1 and 2). Upon cell disruption with lysozyme, the insoluble fraction containing the fusion protein was washed with Triton buffer [35] which notably removed bacterial protein impurities (lane 3). The precursor was efficiently cleaved with cyanogen bromide. After evaporation of the solvent from the cleavage-reaction, water-soluble cleavage products, including P_5S_3, were extracted by alkaline solution (lanes 5), whereas the KSI-fragment remained in the insoluble debris (lane 4). Taking advantage of the strong cationic charge of P_5S_3, cation exchange chromatography on SP-Sepharose was used to isolate the protein (lane 8). Elution with volatile ammonium carbonate allowed both drying and desalting through lyophilization. The purity of P_5S_3 after chromatography was supported by amino acid analysis of the elution fraction, which confirmed the expected polypeptide sequence (see Table S1 in the supplementary information). Only the expected amino acids were detected, and deviations of more than 10 % of the expected values were found only for those amino acids (serine, tyrosine) that are known to decompose (serine may hydrolyze, tyrosine may become halogenated) during the analysis.

Yields of P_5S_3, as determined by UV/vis-spectroscopy, were repeatedly in a range of 15 µmol, corresponding to about 80 mg, per liter bacterial culture when expression was induced in the early stationary phase. A gravimetric quantitation of one batch of the expressed protein was performed in comparison to the spectroscopic determination. Purified recombinant polypeptide

was lyophilized for 38 h, weighed (10 mg in total) and dissolved in water at a concentration of 5 mg/ml. The calculated absorption at 280 nm [47] after 10-fold dilution was 0.134, whereas the measured value was 0.167. The deviation of 20 % is within experimental error. An additional deviation may have been caused by potential protein impurities, indicated in SDS-PAGE by a coomassie-signal in the 60 kDa range (Fig. 2A, lanes 4-8). However, the staining intensity was estimated with ImageJ [52] to be only 5 % of the P_5S_3 band, and is assumed to be a gel artefact since it also appeared in other lanes, irrespective of the presence of proteins.

Yield of the purified protein was lower by a factor of 3 when expression was induced during the exponential growth phase (Fig. 2B), which has been shown for some other recombinant proteins as well [53].

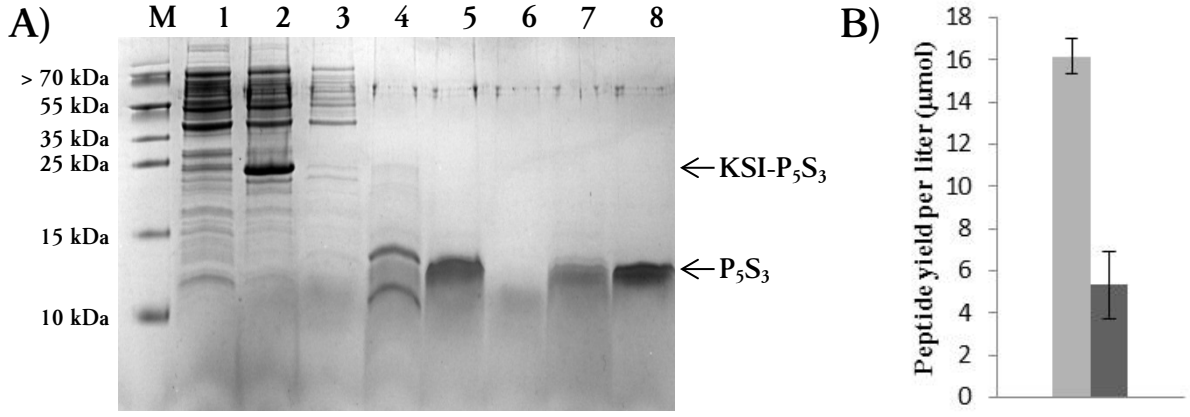

Figure 2

Expression and purification of P_5S_3 from KSI-precursor. **A)** SDS-PAGE analysis of the progress of expression and purification. M: PageRuler Prestained Protein ladder, 1: Total protein before induction, 2: Total protein at least 16 h after induction, 3: Triton washing fraction, 4: Water-insoluble residue after BrCN-cleavage, 5: Water-soluble protein fraction after BrCN-cleavage, 6: Flow through cation exchange chromatography (CEC), 7: Washing fraction CEC, 8: Elution fraction CEC. **B)** Summary of yields of P_5S_3 measured via UV/vis spectroscopy. Values are per 1 liter bacterial culture. Shown are yields when IPTG-induction was at OD_{600} around 2.5 (light grey, 3 preparations) or between 1 and 2 (dark grey, 5 preparations).

An alternative strategy has been reported for the recombinant expression of antimicrobial peptide buforin II [54], based on the charge compensation by a polyanionic fusion. This was also tried for P_5S_3 (sequence described in Supplementary Information as cloning intermediate); however, although purification and processing worked well, yields were below 0.5 µmol (corresponding to 2.5 mg) per liter of culture (not shown).

P_5S_3-induced silica precipitation and self-assembly of the peptide

Since P_5S_3 contained motifs of biomineralizing diatom polypeptides, its capacity to condense silicic acid was analyzed in Tris buffered assays at pH 7.5. P_5S_3 induced the rapid condensation

of silicic acid, producing silica spheres (diameters between 50 nm and 200 nm) which were mostly fused into a network (Fig. 3). In the absence of peptide no precipitates were found. This activity of P_5S_3 is not surprising considering its high cationic charge. However, a remarkable difference in these assays as compared to standard silaffin based silica biomineralization is the absence of any multivalent anionic components such as phosphate. Silica precipitation activity of peptide R5, a cationic repeat of the silaffin-precursor, as well as silaffin Sil-1A, carrying cationic polyamine modifications, strongly depends on the presence of phosphate [3,55]. In these cases, phosphates were proposed to mediate self-assembly, which was described as prerequisite for silica precipitating activity.

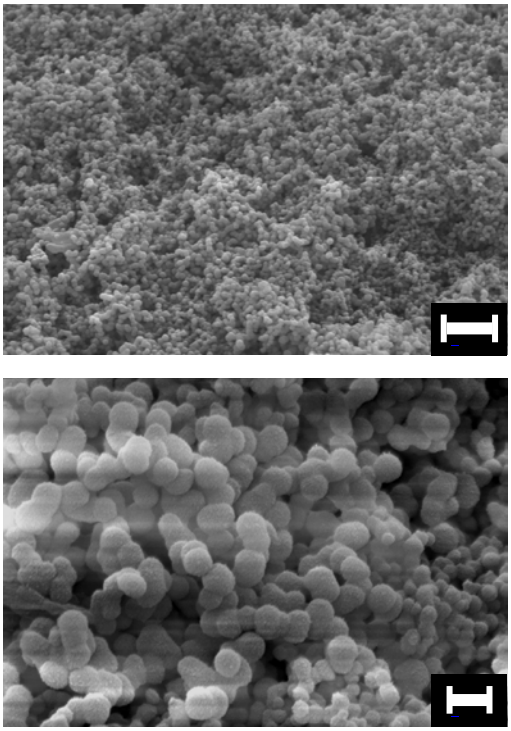

Figure 3

Silica precipitates analyzed with Scanning Electron Microscopy. Polypeptide concentration was 30 µmol/l in 30 mmol/l silicic acid (waterglass as source), buffered with Tris/HCl to pH 7.5. Shown is the same sample at different resolutions, with scale bars representing 1 µm (upper figures) or 200 nm (lower figure).

The observation that silica condensation by P_5S_3 does not require polyanions raises the question whether it self-assembles via an alternative mechanism. Transmission electron microscopy (TEM, Fig. 4) revealed that indeed it has an intrinsic tendency to assemble, and hence overcomes electrostatic repulsion of the cationic side chains. Polypeptide samples were studied in Tris buffer

at pH 7.5, similar to the conditions for silica precipitation analysis. Very large loosely connected assemblies (up to and even larger than 500 nm) of smaller building blocks (10-50 nm) appeared in TEM. This assembled structure presumably mediates a rapid deposition when silica precipitates on its surface, leading to the fused structure observed in SEM (Fig. 3).

The unexpected self-assembly in the absence of polyanionic additives clearly distinguishes P_5S_3 from *C. fusiformis* silaffin, which self-assembles only when phosphorylated or in the presence of added phosphate [3]. Obviously, this modification is not a general requirement for polycationic peptides. Using the polyanion-mediated self-assembly could hence be a regulatory strategy to deliberately control the onset of peptide-mediated silica biomineralization, but also an impact of phosphate on the silicic acid precipitation mechanism itself is possible. Therefore, the effect of phosphorylation on the silica biomineralizing activity of P_5S_3 was evaluated to compare this with the silaffin-system.

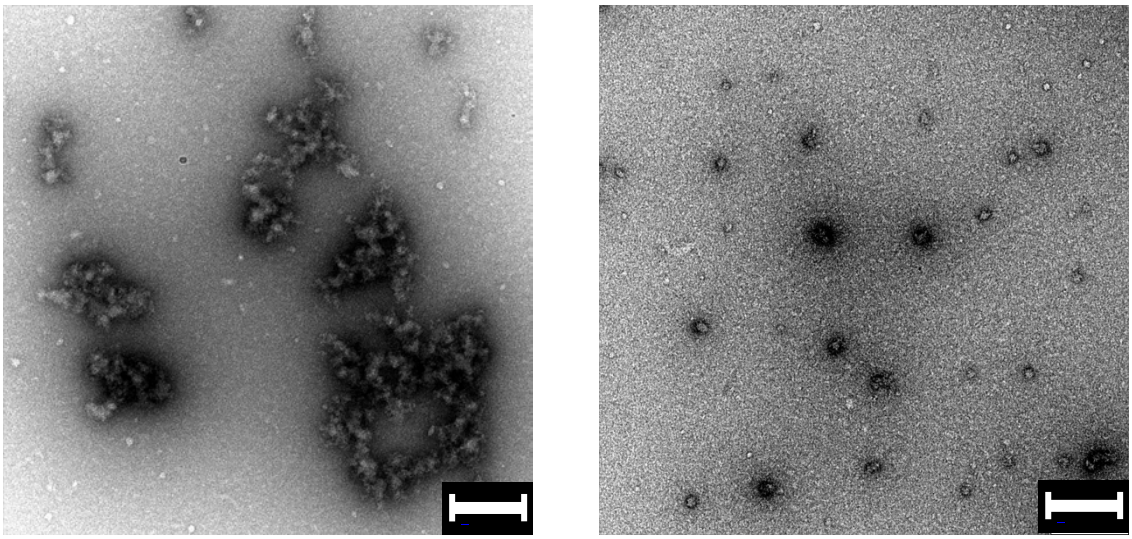

Figure 4

Transmission electron microscopy analysis of P_5S_3 assembly. Polypeptide samples were prepared in Tris buffer pH 7.5. The two micrographs were recorded on different areas of the TEM grid. Scale bars represent 200 nm.

Effect of peptide phosphorylation on silica precipitation

Phosphorylation of P_5S_3 was achieved with protein kinase A. The success of the reaction was obvious from a significantly increased electrophoretic mobility in SDS-PAGE (Fig. 5A and B). In cases where the phosphorylation was incomplete, up to five additional peptide bands appeared (Fig. 5A and larger representation in Fig. S1). We therefore assume quantitative phosphorylation of all five sites when only one distinct band with the highest mobility is seen (Fig. 5B). The phosphorylated polypeptide was purified by cation exchange chromatography. In a step gradient of ammonium carbonate, the non-modified polypeptide eluted at 400 mmol/l, whereas the

phosphorylated polypeptide already eluted at 200 mmol/l. This is to be expected from a reduced net charge and hence an additional proof of successful phosphorylation.

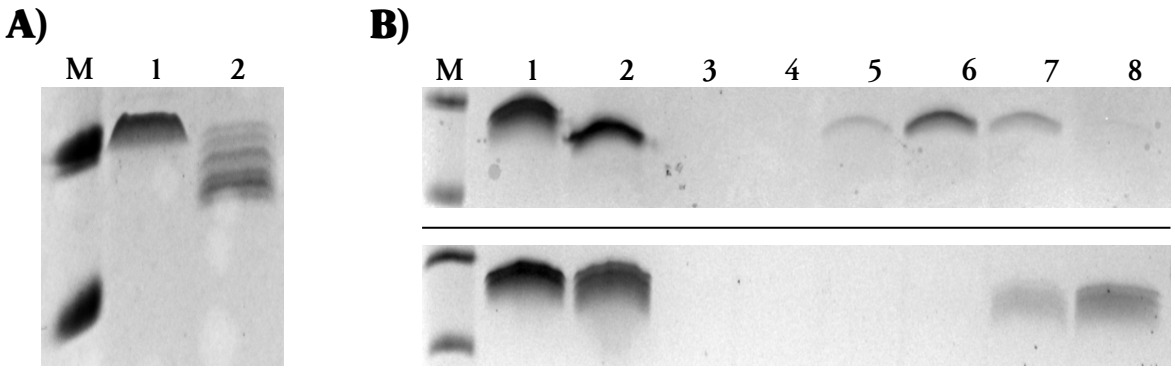

Figure 5

Phosphorylation and cation exchange chromatography. (A) SDS-PAGE analysis of an incomplete phosphorylation. M: PageRuler Prestained, 1: before PKA-addition, 2: after the reaction. (B) SDS-PAGE analysis of enzymatic phosphorylation and subsequent cation exchange chromatography. The upper gel represents the phosphorylation assay, the lower one the negative control without kinase. M: PageRuler Prestained, 1: before kinase-addition (not added in negative control in lower gel), 2: after phosphorylation incubation, 3: flow through of cation exchange chromatography, 4-8: elution with stepwise increased ammonium carbonate concentration of 50, 200, 300, 400 and 500 mmol/l.

Phosphorylation and cation exchange chromatography. (A) SDS-PAGE analysis of an incomplete phosphorylation. M: PageRuler Prestained, 1: before PKA-addition, 2: after the reaction. (B) SDS-PAGE analysis of enzymatic phosphorylation and subsequent cation exchange chromatography. The upper gel represents the phosphorylation assay, the lower one the negative control without kinase. M: PageRuler Prestained, 1: before kinase-addition (not added in negative control in lower gel), 2: after phosphorylation incubation, 3: flow through of cation exchange chromatography, 4-8: elution with stepwise increased ammonium carbonate concentration of 50, 200, 300, 400 and 500 mmol/l.

For comparison of the precipitation activities of non-modified and phosphorylated P_5S_3, precipitated silica was quantitatively analyzed by the molybdenum blue method [50] (Fig. 6). With both polypeptides, the amount of precipitated silica increased with increasing peptide concentration up to 150 µmol/l. Highest silica yields were obtained at pH 6.5 and yields were gradually lowered with increasing pH. Phosphorylation generally led to an overall decrease of precipitation activity, but to rather similar yields above pH 7.5. The decrease of precipitating activity upon phosphorylation is in accordance with observations for biosilicifying, silaffin-derived peptide R5 [6], although silica condensation by this peptide principally depends on the presence of phosphate in the assay [55].

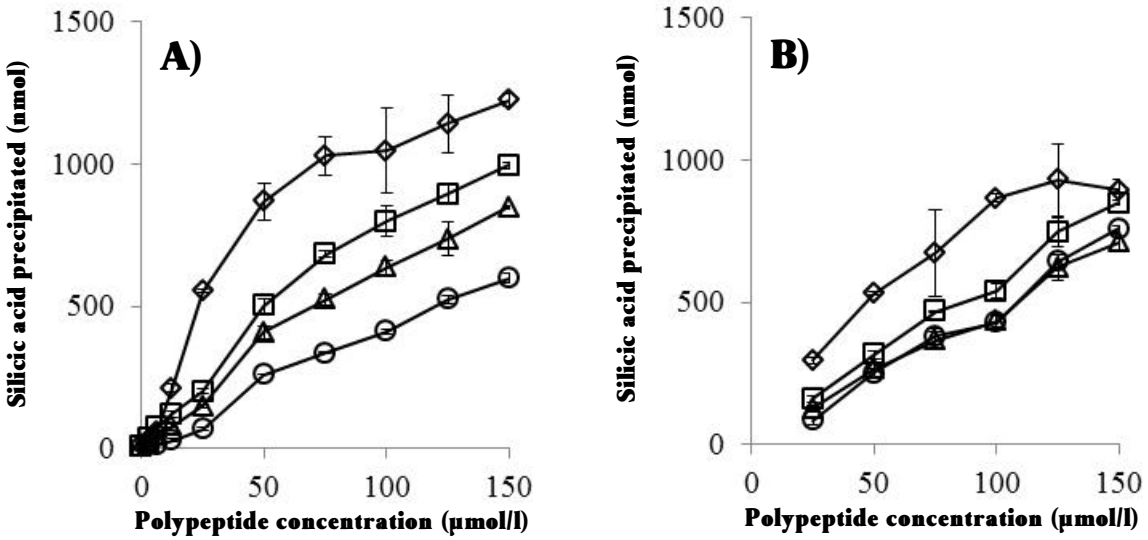

Figure 6

Precipitation activity of **(A)** unmodified and **(B)** phosphorylated P_5S_3. Precipitation was done in buffers (50 mmol/l) BisTris pH 6.5 (diamonds) or Tris pH 7.5 (squares), 8.5 (triangles) or 9.5 (circles). All assays were performed with 30 mmol/l silicic acid (acidically hydrolyzed TMOS as a source) with a total amount of 1500 nmol.

Taken together, P_5S_3 is active in silica biomineralization both in its non-modified, and in its phosphorylated state, the latter being slightly less active. For self-assembly, which is seen in transmission electron microscopy, the peptide does not require any anionic additives. This is in contrast to its native silaffin model which strictly depends on the presence of phosphates to mediate self-assembly of both the polyamine-modified [3] and the unmodified peptide-part [55], which makes them in turn active in biomineralization. Hence, the question arises which interaction is the driving force behind P_5S_3 self-assembly.

Secondary structure determination and possible assembly mechanism of P_5S_3

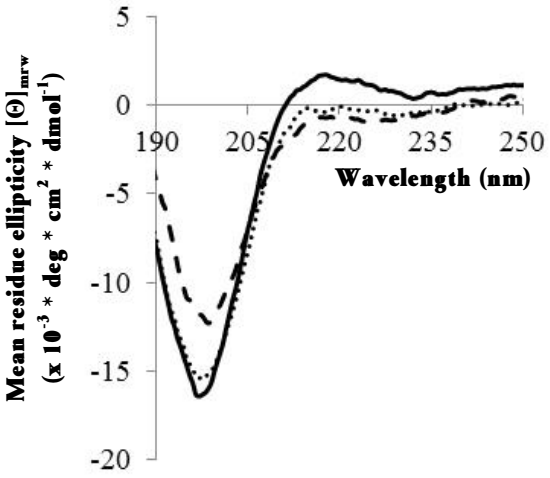

Figure 7

Circular dichroism spectroscopy for secondary structure analysis. Spectra were recorded in the ultraviolet range at temperatures of 4 °C (solid), 23 °C (dotted) and 60 °C (dashed).

The presence of an aromatic tyrosine residue together with many cationic lysines and arginines may imply cation-π interactions [56], which have been used elsewhere to design self-assembling peptides [57,58]. However, this cannot be the major driving force for P_5S_3 assembly since a version without tyrosine exhibited a similar assembly pattern in transmission electron microscopy (not shown). A possible alternative explanation is an interaction between cationic side chains and the peptide bond in the main chain. It is known that particularly arginine is involved in amide-bonding in folded proteins [59]. This interaction depends on exposure of the peptide bond, which in α-helical protein structures is shielded by hydrogen bonds but is accessible in more extended conformations such as β-sheets. For the P_5S_3-polypeptide, circular dichroism spectroscopy revealed a high fraction of unordered conformation (Fig. 7, calculated fractions of secondary structure motifs in Table S2 in the supplementary information). At 4 °C, the CD-spectrum had a dominant negative peak around 197 nm and a much less intense positive peak around 220 nm. Upon heating to 23 and 60 °C, the first peak diminished gradually, and the latter disappeared. These spectral characteristics were originally interpreted as unordered or random coil conformation [60], but have been attributed later-on to the PPII secondary structure motif, an extended helical conformation with high exposure of the main chain towards the solvent [61-63]. The PPII conformation is typical for intrinsically disordered kinase target sites [64]. The protein kinase A site, which was inserted in P_5S_3, has been described to exhibit an extended structure [65], and a sequence-related inhibitor sequence was more explicitly assigned to adopt PPII [66]. This suggests that the PPII structure is at least a local conformation in P_5S_3, exposing amide groups that may participate in main chain - side chain interactions. The question of whether cationic charge and extended conformation suffice to mediate self-assembly, independent of

counterions will further be addressed by NMR spectroscopy with ^{13}C and ^{15}N labeled polypeptides.

Conclusion

The ketosteroid isomerase fusion was efficiently used in the overproduction of the tailored polypeptide P_5S_3 carrying an exceptionally high number of cationic amino acids. Both the unmodified and phosphorylated form of P_5S_3 precipitate silicic acid. The self-assembly of P_5S_3, even in the absence of phosphate and without protein phosphorylation, is intriguing since assembly would be expected to be inhibited by electrostatic repulsion.

Author contributions

C.Z., S.H. and H.P. designed the work and wrote the manuscript. C.Z., S.B. and S.N. were responsible for the experiments. All authors contributed to data analysis.

Acknowledgments

C.Z. is a recipient of a fellowship through the Excellence Initiative (DFG/GSC 266) in the context of the graduate school of excellence "MAINZ" (Materials Science in Mainz). We thank Professor Jürgen Markl (Institute of Zoology, Johannes Gutenberg University Mainz, Germany) for scientific support with transmission electron microscopy and Susanne Linde (University hospital, Jena, Germany) for scanning electron microscopy. We would like to acknowledge the contribution of several Diploma and Bachelor candidates who did cloning steps towards the final expression system in their theses: Anna Hippmann, Christian Roos, Konstantin Müller, Janica Wiederstein, Anna Katharina Resch.

References

[1] M. Sumper, N. Kröger, J. Mater. Chem. 14 (2004) 2059.

[2] N. Kröger, R. Deutzmann, M. Sumper, Science 286 (1999) 1129.

[3] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, Science 298 (2002) 584.

[4] N. Kröger, M.B. Dickerson, G. Ahmad, Y. Cai, M.S. Haluska, K.H. Sandhage, N. Poulsen, V.C. Sheppard, Angew. Chem. Int. Ed. Engl. 45 (2006) 7239.

[5] M.R. Knecht, D.W. Wright, Chem. Commun. (2003) 3038.

[6] C.C. Lechner, C.F.W. Becker, Chem. Sci. 3 (2012) 3500.

[7] J.M. Galloway, S.S. Staniland, J. Mater. Chem. 22 (2012) 12423.

[8] C.L. Chen, N.L. Rosi, Angew. Chemie - Int. Ed. 49 (2010) 1924.

[9] N. Poulsen, N. Kröger, J. Biol. Chem. 279 (2004) 42993.

[10] A. Scheffel, N. Poulsen, S. Shian, N. Kröger, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3175.

[11] M. Nemoto, Y. Maeda, M. Muto, M. Tanaka, T. Yoshino, S. Mayama, T. Tanaka, Mar. Genomics 16 (2014) 39.

[12] N. Poulsen, M. Sumper, N. Kröger, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 12075.

[13] N. Kröger, R. Deutzmann, M. Sumper, J. Biol. Chem. 276 (2001) 26066.

[14] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14133.

[15] C.C. Lechner, C.F.W. Becker, J. Pept. Sci. 20 (2014) 152.

[16] R. Wieneke, A. Bernecker, R. Riedel, M. Sumper, C. Steinem, A. Geyer, Org. Biomol. Chem. 9 (2011) 5482.

[17] L. Andersson, L. Blomberg, M. Flegel, L. Lepsa, B. Nilsson, M. Verlander, Biopolymers 55 (2000) 227.

[18] A.L. Demain, P. Vaishnav, Biotechnol. Adv. 27 (2009) 297.

[19] V. Sheppard, N. Poulsen, N. Kroger, J. Biol. Chem. 285 (2010) 1166.

[20] S. Jono, C. Peinado, C.M. Giachelli, J. Biol. Chem. 275 (2000) 20197.

[21] M. Wojtas, M. Wolcyrz, A. Ozyhar, P. Dobryszycki, Cryst. Growth Des. 12 (2012) 158.

[22] E. Salih, Connect. Tissue Res. 44 Suppl 1 (2003) 223.

[23] B. Manconi, T. Cabras, A. Vitali, C. Fanali, A. Fiorita, R. Inzitari, M. Castagnola, I. Messana, M.T. Sanna, Protein Expr. Purif. 69 (2010) 219.

[24] L.A. Pinna, M. Ruzzene, Biochim. Biophys. Acta 1314 (1996) 191.

[25] B.E. Kemp, D.J. Graves, E. Benjamini, E.G. Krebs, J. Biol. Chem. 252 (1977) 4888.

[26] D.E. Olins, A.L. Olins, Von Hippel, P H, J. Mol. Biol. 24 (1967) 157.

[27] W. Hartmann, H.J. Galla, Biochim. Biophys. Acta 509 (1978) 474.

[28] K.L. Piers, M.H. Brown, R.E. Hancock, Gene 134 (1993) 7.

[29] N.Y. Yount, M.R. Yeaman, Ann. N. Y. Acad. Sci. 1277 (2013) 127.

[30] J. Jarczak, E.M. Kościuczuk, P. Lisowski, N. Strzalkowska, A. Jóźwik, J. Horbanczuk, J. Krzyzewski, L. Zwierzchowski, E. Bagnicka, Hum. Immunol. 74 (2013) 1069.

[31] Y. Li, Protein Expr. Purif. 80 (2011) 260.

[32] N.S. Parachin, K.C. Mulder, A.A.B. Viana, S.C. Dias, O.L. Franco, Peptides 38 (2012) 446.

[33] A. Kuliopulos, Proc. Natl. Acad. Sci. 84 (1987) 8893.

[34] A. Kuliopulos, C.T. Walsh, J. Am. Chem. Soc. 116 (1994) 4599.

[35] S. Kyle, A. Aggeli, E. Ingham, M.J. McPherson, Biomaterials 31 (2010) 9395.

[36] J.M. Riley, A. Aggeli, R.J. Koopmans, M.J. McPherson, Biotechnol. Bioeng. 103 (2009) 241.

[37] T.-J. Park, J.-S. Kim, S.-S. Choi, Y. Kim, Protein Expr. Purif. 65 (2009) 23.

[38] I. Cipáková, J. Gasperik, E. Hostinová, Protein Expr. Purif. 45 (2006) 269.

[39] S. Sharpe, W.-M. Yau, R. Tycko, Protein Expr. Purif. 42 (2005) 200.

[40] M. Zorko, B. Japelj, I. Hafner-Bratkovic, R. Jerala, Biochim. Biophys. Acta 1788 (2009) 314.

[41] P.M. Hwang, J.S. Pan, B.D. Sykes, Protein Expr. Purif. 85 (2012) 148.

[42] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 3. ed., 3., Wiley, New York, 1995.

[43] G. Bertani, J. Bacteriol. 62 (1951) 293.

[44] E. Gross, B. Witkop, J. Am. Chem. Soc. 83 (1961) 1510.

[45] E. Gross, B. Witkop, J. Biol. Chem. 237 (1962) 1856.

[46] W.A. Schroeder, J.B. Shelton, J.R. Shelton, Arch. Biochem. Biophys. 130 (1969) 551.

[47] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, Protein Sci. 4 (1995) 2411.

[48] H. Schaegger, G. von Jagow, Anal. Biochem. 166 (1987) 368.

[49] G. von Jagow, H. Schaegger, A Practical Guide to Membrane Protein Purification, Academic Press, San Diego, 1994.

[50] T. Coradin, D. Eglin, J. Livage, Spectroscopy 18 (2004) 567.

[51] J.R. Harris, J. Markl, Eur. J. Biochem. 225 (1994) 521.

[52] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, Nat. Methods 9 (2012) 671.

[53] J. Ou, L. Wang, X. Ding, J. Du, Y. Zhang, H. Chen, A. Xu, Biochem. Biophys. Res. Commun. 314 (2004) 174.

[54] J.H. Lee, I. Minn, C.B. Park, S.C. Kim, Protein Expr. Purif. 12 (1998) 53.

[55] F. Rodriguez, D.D. Glawe, R.R. Naik, K.P. Hallinan, M.O. Stone, Biomacromolecules 5 (2004) 261.

[56] D.A. Dougherty, Science 271 (1996) 163.

[57] C.C. Chen, W. Hsu, K.C. Hwang, J.R. Hwu, C.C. Lin, J.C. Horng, Arch. Biochem. Biophys. 508 (2011) 46.

[58] C.C. Chen, W. Hsu, T.C. Kao, J.C. Horng, Biochemistry 50 (2011) 2381.

[59] C.L. Worth, T.L. Blundell, BMC Evol. Biol. 10 (2010) 161.

[60] N. Greenfield, G.D. Fasman, Biochemistry 8 (1969) 4108.

[61] A.A. Adzhubei, M.J.E. Sternberg, A.A. Makarov, J. Mol. Biol. 425 (2013) 2100.

[62] N.J. Greenfield, Nat. Protoc. 1 (2006) 2876.

[63] M.L. Tiffany, S. Krimm, Biopolymers 11 (1972) 2309.

[64] W. Austin Elam, T.P. Schrank, A.J. Campagnolo, V.J. Hilser, Protein Sci. 22 (2013) 405.

[65] P.R. Rosevear, D.C. Fry, A.S. Mildvan, M. Doughty, C. O'Brian, E.T. Kaiser, Biochemistry 23 (1984) 3161.

[66] D. Knighton, J. Zheng, L. ten Eyck, N. Xuong, S. Taylor, J. Sowadski, Science 253 (1991) 414.

Electronic Supplementary Information

Detailed description of cloning steps in P_5S_3 creation before KSI-fusion

The earliest version created through insertion of the assembled PIN Primer constructs (see main publication for construction scheme) resulted in the target polypeptide P_5S_3, fused to a *pelB* export leader donated by the pET 22b(+) plasmid. An interspersing protease cleavage site for Factor Xa protease had earlier been introduced by primer hybrid insertion. The P_5S_3 gene assembled from PIN hybrids contained a cysteine following the first codon (methionine) since it had to fit into a BamHI restriction site in the target plasmid.

Factor Xa (FXa) showed digestion not only at its cleavage site (sequence IEGR | S), but also within the target gene P_5S_3, and therefore was exchanged by a HRV3C protease cleavage site (sequence LEFLFQ | PAM). The FXa site was excised with NcoI and EcoRI (New England Biolabs GmbH, Frankfurt (Germany)), and phosphorylated primer hybrid HRV3C_s (5' AATTCCCTGGAAGTTCTGTTCCAGGGGCCGGC 3') and HRV3C_as (5' CATGGCCGGCCCCTGGAACAGAACTTCCAGGG 3') were combined with gel-purified digested plasmid (primer insertion as described for PIN hybrids). Then, the cysteine at position 2 of the target sequence was exchanged for serine with QuikChange II Site Directed Mutagenesis kit (Agilent Technologies GmbH, Waldbronn (Germany)) and primer hybrid CtoSinPxSy_f (5`-CCGGCCATGGGATCCTCTCGACGTGCTTCC- 3`) and CtoSinPxSy_r (5`-GGAAGCACGTCGAGAGGATCCCATGGCCGG- 3`), to avoid intermolecular disulfide formation, which potentially complicates purification. It should be noted that no expression was observed without *pelB*-leader but that *pelB*-containing protein versions were hardly soluble. Purification with metal ion affinity chromatography was not satisfactory, and protease digest was incomplete.

To enhance solubility of the expression product, a charge-compensating anionic fusion was introduced. We examined the anionic pro-sequence of *Cylindrotheca fusiformis* silaffin gene (Genbank **AF191634.1**), kindly provided by Nils Kröger on a pUC 18 plasmid. The sequence was PCR-amplified with primers SilL-f (5'-AAACATATGGTGGCCAGCGACTCCTCGGATGACGC-3') and SilL-r (5'-AAAAAGAATTCTCGGG GATACGAAGTTCTTCCTCTTCGG-3') with *Pwo*-Polymerase (Peqlab Biotechnology GmbH, Erlangen (Germany)) and the following cycle conditions: initial denaturation 2 min 95 °C, cycle (35x) [denaturation 1 min 95 °C, annealing 30 sec 52 °C, elongation 30 sec 72 °C], final elongation 5 min 72 °C. The amplificate was purified with peqGOLD Cycle Pure kit (Peqlab Biotechnology GmbH, Erlangen (Germany)). Both the fragment and the plasmid described above were digested with *Nde*I and *Eco*RI, excising the *pelB*-leader. The opened target plasmid was dephosphorylated (Calf Intestinal Phosphatase, New England Biolabs GmbH, Frankfurt (Germany)), purified in a 1.5 % (w/v) agarose gel (TAE-buffer, see main publication) and gel-extracted with peqGOLD Gel Extraction kit (Peqlab

Biotechnology GmbH, Erlangen (Germany)). Ligation was achieved with Quick Ligation Kit (New England Biolabs GmbH, Frankfurt (Germany)) with a molar ratio of insert to plasmid 3:1.

Expression of this charge-compensated fusion polypeptide was not observed, so the pelB-leader was re-integrated. Source plasmid with pelB, HRV3C-site and target polypeptide coding sequence as well as the plasmid having the anionic leader sequence without *pelB*-leader were digested with *Eco*RV and *Msc*I. *Eco*RV cuts within the *lacI* gene, far enough away from the gene of interest to allow transfer of a larger DNA-fragment for handling purposes. *Msc*I cuts in the 3' region of *pelB* and in the 5' region of the anionic silaffin pro-sequence, respectively. Ligation gave a three-domain fusion protein of *pelB*, anionic leader, (HRV3C-site) and target protein.

This construct was expressed in the bacterial host, but with low yields. Therefore, the expression plasmid pET 31b(+), used in the main publication, was chosen. The three-domain fusion had *Ava*I restriction sites for excision and reintegration into pET 31b(+). It was, however, not directly excised, but PCR-amplified with primers T7PrompET22b(+)fw and pET22b(+)rev (Table 1 in main publication) and *Pwo* DNA Polymerase (see above) to achieve larger amounts of the insert fragment. PCR-conditions were: initial denaturation 1 min 95 °C, cycle (35x) [denaturation 30 s 95 °C, anneal 30 sec 50 °C, elongate 1 min 72 °C], final elongation 5 min 72 °C. The PCR-product was purified (Cycle Pure Kit, see above), *Ava*I digested, gel purified (agarose gel, see main publication) and the desired fragment was ligated into the digested pET 31b(+) plasmid.

Both in pET 22b(+) in the three-domain protein, and in the pET 31b(+) KSI-fusion, a tyrosine (for UV/vis) quantification was introduced by site-directed mutagenesis with primer hybrid InsY-s (5' GGCCATGGGATCCTATTCTCGACGTGCTTCC 3') and InsY-as (5' GGAAGCACGTCGAGAATAGGATCCCATGGCC 3'), which adds an additional amino acid. Expression of the hexahistidyl-tag was suppressed in both versions through site-directed mutagenesis with primer hybrids EtoStopPxSy_s (5' CGCGCTAGCCTCTAGCACCACCACC 3') and EtoStopPxSy_as (5' GGTGGTGGTGCTAGAGGCTAGCGCG 3').

Supplementary Tables and Figures

Figure S1

DNA-Sequence of P_5S_3. A) Assembly of P_5S_3 gene from primer hybrids PIN 1, 2, 3 with important restriction sites (in grey). B) Sequence of translated region for KSI-fused P_5S_3. Three codons were changed compared to the gene initially assembled from the PIN-primer hybrids (site-directed mutagenesis) and are marked in red (noted in Materials and Methods and in the Supplementary Information).

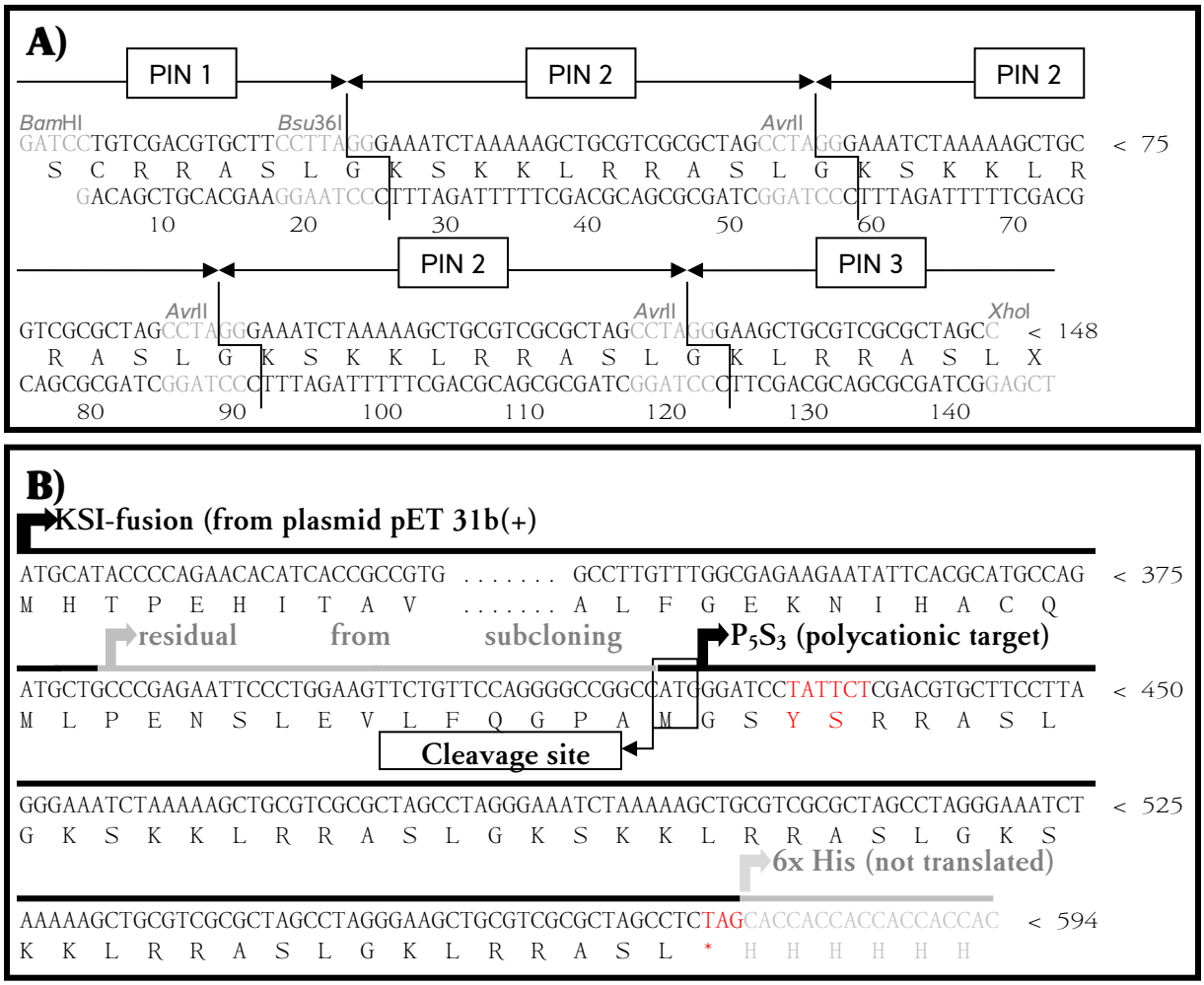

Table S1

Amino acid analysis of purified P_5S_3. A sample was sent to Genaxxon bioscience GmbH (Ulm, Germany) for analysis. The polypeptide was acidically hydrolyzed in 6 mol/l hydrochloric acid and the amino acids separated by cation exchange chromatography. For quantification, amino acids were subjected to post-column derivatization with ninhydrin. Genaxxon bioscience GmbH states a 1 to 10 % deviation from the expected number of amino acids generally acceptable in the experimental set-up, whereas serine may be biased further because of partial destruction.

Amino acid	number in sequence	number revealed	relative deviation
Asp (D/N)		0.0	
Thr (T)		0.0	
Ser (S)	10	7.8	-21.5
Glu (E/Q)		0.0	
Pro (P)		0.0	
Gly (G)	5	4.9	-1.3
Ala (A)	5	5.2	4.1
Cys (C), (Cys(O3H); Cys2)		0.0	0.0
Val (V)		0.0	
Met (M)		0.0	
Ile (I) (+ allo-Ile)		0.0	
Leu (L)	9	9.2	2.3
Tyr (Y)	1	0.9	-13.0
Phe (F)		0.0	
His (H)		0.0	
Lys (K)	10	9.1	-8.7
Trp (W)		0.0	
Arg (R)	10	10.4	3.7
		Mean absolute deviation	2.9%

Table S2

Secondary structure assignment. Underlying CD-spectra are shown in Figure 8 in the main publication. Spectra were analysed with the CDpro software package **(Sreerama, N., Woody, R.W., *JMolBiol* 1994, 242, 497 and Sreerama, N., Woody, R.W., *AnalBiochem* 2000, 287, 252)** to calculate the contribution of different secondary structure motifs to the spectra. The partially negative values from CDSSTR are indicative for an inaccuracy in the assignment which is hence only a rough estimation. The main fraction is found to be unordered at all temperatures examined, with a considerable contribution of turn structures.

	P5S3 4 °C			**P5S3 23 °C**			**P5S3 60 °C**		
	Selcon	Contin	CDSSTR	Selcon	Contin	CDSSTR	Selcon	Contin	CDSSTR
Helix (regular)	0.034	0	-0.019	0.035	0.002	-0.014	0.028	0.001	-0.016
Helix (distorted)	0.079	0.055	-0.001	0.074	0.06	0	0.055	0.055	0.016
Sheet (regular)	0.071	0.202	0.232	0.119	0.188	0.243	0.158	0.207	0.237
Sheet (distorted)	0.097	0.135	0.11	0.11	0.124	0.128	0.116	0.126	0.137
Turn	0.205	0.221	0.227	0.241	0.234	0.227	0.242	0.23	0.245
Unordered	0.451	0.386	0.398	0.465	0.392	0.386	0.404	0.38	0.352
Sum	**0.937**	**0.999**	**0.947**	**1.044**	**1**	**0.97**	**1.003**	**0.999**	**0.971**

Section IV - Triple-Role of Recombinant Silaffin-Like Cationic Polypeptide P_5S_3: Peptide-Induced Silicic Acid Stabilization, Silica Formation and Inhibition of Silica Dissolution

Argyro Spinthaki,[#] Christian Zerfaß,[§,+] Harald Paulsen,[§] Stephan Hobe[§,*] and Konstantinos D. Demadis [#,*]

[#] Crystal Engineering, Growth and Design Laboratory, Department of Chemistry, University of Crete, Voutes Campus, Heraklion, Crete, GR-71003, Greece.

[§] Institut für Allgemeine Botanik, Johannes-Gutenberg-Universitaet Mainz, Muellerweg 6, D-55099 Mainz, Germany.

[+] Graduate School Materials Science in Mainz, Staudinger Weg 9, 55128 Mainz, Germany

This section is a publication manuscript in preparation.

Abstract

Diatoms and other silica condensing organisms have to deal with the basic aspects of silica chemistry such as condensation and dissolution when transporting and concentrating the precursor orthosilicic acid and then condensing it into polymeric, amorphous silica in a spatially and time-controlled manner. Isolation of biomineralizing proteins as well as systematic analyses of synthetic biomineralization substances has generated a large number of useful functional insights, however, a comprehensive picture about the molecular interactions guarding all aspects of morphogenesis between orthosilicic acid and complex mesoscopic silica structures has not emerged yet, although this knowledge can be valuable for technical applications. Here we describe a recombinant approach, using the polycationic polypeptide P_5S_3 with triple impact on silica solubility: (i) stabilization of silicic acid at supersaturated but low concentration of 500 ppm (8.3 mmol/L), (ii) deceleration of amorphous silica hydrolysis at a globally non saturated silicic acid concentration of 60 ppm (1 mmol/L) as well as (iii) enhancement of silica precipitation at 1,800 ppm (30 mmol/L) silicic acid with the polypeptide quantitatively trapped within silica. In a neutral environment the polypeptide sequence is positively charged due to its high isoelectric point at 12.64, therefore the three effects probably occur due to the interaction of the oppositely charged silanol groups with lysine (pKa = 10.79) plus arginine (pKa = 12.48) side chains (most abundant amino acids in the polypeptide).

Introduction

Silicon is the second most abundant element in the earth's crust [1,2] and is abundant in its oxidized form (silicates) in argillaceous minerals, which contribute to soil fertility because they absorb water and trace elements essential for the plants' nutrition [3,4]. Several organisms use

silicon in the form of silica (SiO_2), and thus take part in the world's Si cycle [5]. Diatoms, an example of such inspirational biomineralizing organisms, form their cell wall starting out from monosilicic acid ($Si(OH)_4$), in neutral pH environment [6-8]. They import silicic acid via silicon transport proteins (SITs) [9-12], which play an important role in silicon transport because the concentration of silicic acid in natural waters is low [13-15], and condensation of $Si(OH)_4$ in undersaturated solutions, induced by externally-added organic additives, has not yet been reported. Silicic acid is transported into the silica deposition vesicles (SDVs) and accumulated there to up to 100 mmol/L (6,000 ppm), much higher than the saturation concentration at pH <9 of 150 ppm/2.5 mmol/L [16,17]. The condensation of silicic acid in diatoms is enhanced by e.g. post-translationally modified peptides with high cationic net charge, known as silaffins [18-20]. The high supersaturation of silicic acid *in vivo* appears necessary since no organic biomineralizing agent is known to induce the formation of silica from undersaturated solutions. By contrast, oligoamines, polyamines and poly-L-lysine, which are known to be precipitants of silica in supersaturated silicic acid solutions, were reported to accelerate the decondensation of silica in globally undersaturated solutions [21], and it is an open question how diatom cell walls are protected from dissolution [22].

The aim of this study is the systematic investigation of the tailor-made, silica biomineralizing polypeptide P_5S_3 (ref [23] and section III of this thesis) with regard to its influence on three aspects of silicon chemistry: (i) stabilization of the chemically unstable silicic acid, (ii) enhancement of silica formation, and (iii) retardation of the dissolution of amorphous silica. P_5S_3 (see Figure 1) was designed based on the amino acid sequence of *Cylindrotheca fusiformis* silaffin-1A [18]. The polypetide bears numerous basic amino acid side chains, which are either lysine-clusters or arginine-doublets, the latter being part of a protein kinase A phosphorylation site [24]. P_5S_3 can be expressed in *Escherichia coli* by fusing it to ketosteroid isomerase (KSI) [25], with subsequent chemical cleavage of the fusion by using cyanogen bromide ([23] and section III). By analyzing the impact of this polypeptide on silicic acid stabilization, silica formation and silica dissolution, we provide a comprehensive study of the silicic acid-affecting action of the biomineralizing polypeptide P_5S_3.

NH_2-GSYS**RR**AS*LG[**K**S**KK**L**RR**AS*LG]$_3$**K**L**RR**AS*L-*COOH*

Figure 1. The sequence of P_5S_3 in amino acid single letter code [26]. The cationic amino acids lysine (K) and arginine (R) are shown underlined and bold. The amino- and carboxy-terminus are indicated by "NH_2" and "COOH", respectively. The central sequence (squared brackets) is repetitive and occurs three times (marked by the index) in P_5S_3. Serines in the proteins kinase A target site [24] are marked by an asterisk.

Experimental Section

Definitions and Si species nomenclature.

In this paper we adopt the terminology "molybdate-reactive silica" described by Icopini *et al.* [27] and Iler [28] as those "Si" species that are responsive to the silicomolybdate test, described in detail below. In accordance to this definition,"molybdate-reactive" includes two or three species, namely, monomeric (monosilicic acid, H_4SiO_4), dimeric (disilicic acid, $H_6Si_2O_7$), and possibly trimeric ($H_8Si_3O_{10}$) silica.

Reagents, chemicals and materials.

Sodium silicate pentahydrate $Na_2SiO_3 \cdot 5H_2O$ was purchased from EM Science (Merck). Tetramethoxysilane (TMOS) was from Fluka (Sigma Aldrich). Ammonium molybdate ($(NH_4)_6Mo_7O_{24} \cdot 4H_2O$) and oxalic acid ($H_2C_2O_4 \cdot 2H_2O$) were obtained from EM Science (Merck) and Fluka (Sigma Aldrich), ascorbic acid from Roth. Sodium hydroxide (NaOH) was purchased from Merck and Sigma Aldrich, hydrochloric acid 37 % from Riedel de Haen and Merck, and sulfuric acid (95 - 98 %) from Merck. As buffer ions, tris(hydroxymethyl)-aminomethane (Tris, Biomol) and bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane (Bis-Tris, Sigma Aldrich) were pH-adjusted with hydrochloric acid, and sodium acetate trihydrate (Merck) was adjusted with glacial acetic acid (Merck). Trichloroacetic acid (TCA) for protein precipitation was purchased from Merck. Poly-L-lysine (MW = 15 - 30 kg/mol) was from Sigma Aldrich, polyacrylic acid (MW = 2000 Da) from Polysciences, Inc.. In-house deionized water from an ion-exchange resin was used for all experiments. This water was tested for molybdate-reactive silica and was found to contain negligible amounts. Acrodisc filters (0.45 μm and occasionally 0.20 μm) from Pall-Gelman Corporation were used.

Preparation of supersaturated sodium silicate solutions and silicomolybdate test reagents.

A solution containing silicate (500 ppm) was prepared by dissolving 4.08 g of $Na_2SiO_3 \cdot 5H_2O$ in 2 L of deionized water (a non-glass container must be used), followed by overnight rigorous stirring. Alternatively, TMOS was acidically hydrolyzed in 1 mmol/L hydrochloric acid at a concentration of 150 mmol/L, for 15 - 30 minutes with shaking at 1,400 rpm, immediately before using it in the experiment. Stock solutions of the additives (polypeptide P_5S_3, poly-L-lysine (PLL) and polyacrylic acid (PAA)) were prepared in water.

Two methods were used for the determination of molybdate reactive silica. For the molybdenum yellow method, the following solutions were prepared for the silicate spectrophotometric detection test: (a) Ammonium molybdate solution: 10 g of ammonium molybdate was dissolved

in 100 mL of water, and its pH was adjusted between 7 and 8 with NaOH to avoid precipitation of ammonium molybdate. (b) HCl solution: one volume 37 % HCl was mixed with an equal volume of water. (c) Oxalic acid solution: 8.75 g of oxalic acid was dissolved in 100 mL of water. All solutions were kept in polyethylene containers (glass containers must be avoided to minimize SiO_2 dissolution and silicate leaching into the test solutions). HCl and ammonium molybdate were kept in the refrigerator, while silica and oxalic acid solution were kept at room temperature.

For the molybdenum blue analysis, a 10 % w/v ammonium heptamolybdate tetrahydrate solution was prepared in 1.6 mol/L sulfuric acid. Oxalic acid dihydrate and ascorbic acid were both dissolved in water, at 10 % w/v and 100 mmol/L, respectively.

Determination of molybdate reactive silica.

Frequently, there is a misconception about the use of the terms "silica", silicate" and "silicic acid". In this paper the term "silica" indicates the product of silicic acid polycondensation. The term "silicate" indicates all the forms of $Si(OH)_4$ at various deprotonated states ($H_{4-x}SiO4^{x-}$, x = 0, 1, 2, 3, 4; 4 indicates that silicic acid is fully deprotonated) which exist at pH regions > 9. [29] Note that mono- and disilicic acid rapidly interconvert and they both contribute to molybdate-reactive silica.

Molybdenum yellow analysis. For our studies, we used a modification of (several) published procedures of the molybdenum yellow method [30-35] (for which a precision of ± 5 % is reported) as follows: A quantity of 2 mL aliquot from the working solution is filtered through a 0.45 μm syringe filter, and then diluted to 25 mL in a special cylindrical cell (the Hach Co.) of 1 cm pathlength. Then, 1 mL of ammonium molybdate stock solution and 0.5 mL HCl (1:1 dilution of the concentrated solution) are added to the sample cell, the solution is mixed well by shaking and finally left undisturbed for 10 min. After that, 1 mL of oxalic acid stock solution is added and thoroughly mixed again. The solution is set aside for 2 min. After the second time period, the photometer is set to "zero absorbance" using a sample of water plus all chemicals used for the silicomolybdate test except for silicic acid ("blank"). Finally, the sample absorbance is measured (at 420 nm) and is expressed as "ppm SiO_2". The detectable concentration range for this specific protocol is 6 - 75 ppm. In order to calculate the concentration in the original solution an appropriate dilution factor is applied. The silicomolybdate method is based on the principle that ammonium molybdate reacts only with "reactive" silica (see definitions above) and any phosphate present, to yield heteropoly acids, yellow in color, at low pH (about 1.2). Oxalic acid is added to destroy the molybdophosphoric acid (in case phosphate is present in the water) leaving the silicomolybdate complex intact, and thus eliminating any color interference from phosphates. It should be emphasized that molybdate reacts only with the monomeric, dimeric, and possibly trimeric forms as stated above, but is totally unreactive to colloidal silica particles, if the measurements are performed with strict time control (see above). This was verified

experimentally in our laboratory. Within the strict time control of the above-described measurements, only mono- and disilicic acids are reactive.

Molybdenum blue analysis. We used this method [30] to quantify silicic acid from the precipitation assay (pellets hydrolysed in 1 mol/L sodium hydroxide for 1 hour at room temperature) or in the decondensation assay (an aliquot of the assay was added directly and supplemented with an appropriate amount of 2 mol/L hydrochloric acid sufficient to completely protonate the Tris base from the buffer, see below). Samples were diluted to 200 µl with water, and 4 µl 10 % w/v ammonium heptamolybdate tetrahydrate in 1.6 mol/L sulfuric acid were added. After 10 minutes, 4 µl of 10 % w/v oxalic acid dihydrate were added, and after 1 min of incubation, the silicomolybdic acid complex was reduced by adding 4 µl of 100 mmol/L ascorbic acid. After a further 15 min period, absorbance was measured at 810 nm in a 96 well plate (UV star, Greiner Bio-ONE) in a plate reader Tecan M1000 (Tecan). The concentration of silicic acid was determined by comparing the absorbance with a standard curve obtained from known silicic acid amounts, generated from acidically hydrolysed TMOS.

Stabilization of molybdate-reactive silica in the absence ("control" protocol) and presence of P_5S_3 polypeptide.

A. Absence of P_5S_3 polypeptide ("control"). 100 mL of the 500 ppm sodium silicate stock solution (see above) was placed in a polyethylene beaker charged with a teflon-covered magnetic stirring bar. The pH of this solution was initially ~ 11.8 and subsequently adjusted under stirring to the desired pH value by addition of hydrochloric acid or sodium hydroxide (the change in the resulting volume was 3 % or less and was taken into account for subsequent calculations). The beaker was then covered with plastic membrane and set aside without stirring. The solutions were checked for molybdate-reactive silica by the silicomolybdate method every hour for the first eight hours and after 24, 48, 72 h time intervals after the onset of the pH adjustment. There must be strict time control in measuring molybdate-reactive silica, in order to avoid conversion of higher oligomers/colloidal silica to silicic and disilicic acids. Specifically, after the ammonium molybdate and hydrochloric acid solutions were added to a working sample, a period of 10 min has to pass until the solution of oxalic acid is added to the same sample. Then, another 2 min period has to follow until the final measurement. All samples (control and in the presence of polypeptide, or any other additive) were treated in precisely the same way. Separate experiments were performed in which the working solutions were stirred instead of simply being set aside, but no difference in molybdate-reactive silica levels was found, compared to the quiescent solutions.

B. Presence of P_5S_3 polypeptide. 100 mL portions of the 500 ppm sodium silicate stock solution (see above) were placed in polyethylene containers and charged with teflon-covered magnetic stir bars. In each container, different volumes of polypeptide (from a 10,000 ppm (1 % w/v) stock solution) were added to achieve the desired polypeptide concentration. These ranged from 20 - 60 - 100 - 165 ppm and the added volumes were 200 - 600 - 1000 - 1650 µL. If desired,

poly(acrylic acid) with a molecular weight of 2,000 Da was added to 10 - 30 - 50 - 100 ppm from a 10,000 ppm (1 % w/v) stock solution (100 - 300 - 500 - 1000 µL). After that, the same procedure as the "control" protocol was followed.

Silicic acid precipitation assay.

The silicic acid source used in this experiment was acidically hydrolyzed TMOS. A volume of the 150 mmol/L silicic acid stock solution was directly added to the samples of interest. These contained 50 mmol/L of one of three buffers (sodium acetate/acetic acid at pH 4.5, Bis-Tris-HCl at pH 5.5 and 6.5, Tris-HCl at pH 7.5 and 8.5) as well as P_5S_3 at different concentrations, as mentioned in the particular results. Precipitated silica was harvested by centrifugation (5 min at 18,400 g), and supernatant and pellet fractions were analyzed by quantification of silicic acid (pellet fraction only, hydrolysis in 1 mol/L sodium hydroxide for 1 hour followed by molybdenum blue quantification, see above) or examined for the partitioning of P_5S_3 (both fractions) by SDS-PAGE. In the latter case, silica was dissolved in 1 mol/L sodium hydroxide. This was achieved by adding 2 mol/L sodium hydroxide in equal volume to the supernatant or by dissolving the pellet fraction in a 1 mol/L solution. To facilitate SDS-PAGE, samples were desalted by trichloroacetic acid (TCA) precipitation of the polypeptide [36], both reducing the high salt load from dissolution in sodium hydroxide and removing silicic acid to avoid re-precipitation after pH neutralization. Aliquots of 50 µl of the basic hydrolysates were diluted (10-fold) by adding 250 µL water, 100 µL of 1 mol/L hydrochloric acid for acidification and 50 µL TCA (100 % w/v, final concentration 10 % w/v, induces polypeptide precipitation). Samples were vigorously vortexed and then placed on ice for 15 minutes, before the precipitated protein fraction was harvested by centrifugation for 10 minutes, at 14,000 g and 4 °C. The pellet was washed once by suspending it in 500 µL acetone (vortexing) followed by a second centrifugation at the same conditions. Eventually, the polypeptide was re-dissolved in water and applied to gel electrophoresis at an expected amount of 1 - 1.5 µg per gel lane, assuming all P_5S_3 polypeptide had either gone to the silica-pellet fraction, or remained in the supernatant fraction. The SDS-PAGE system used was a Tris/Tricine gel with 16.5 % total acrylamide concentration [37-39].

Silica dissolution in undersaturated solutions.

To monitor the dissolution of silica, the hydrolyzed TMOS was first polymerized by dilution to 30 mmol/L (1,800 ppm) silicic acid, in Tris-HCl pH 7.0 or 7.5 and shaken for 15 min at 1,400 rpm. After this pre-autocondensation time, the samples were further diluted to a total silicic acid concentration of 1 mmol/L (60 ppm), which is below the autopolycondensation threshold [28]. For assays in the presence of P_5S_3, the polypeptide was dissolved in water and added to 1 µmol/L (5.6 ppm) upon the step of diluting the silicic acid solution to 1 mmol/L. As an alternative biomineralizing agent, poly-L-lysine hydrochloride (Sigma-Aldrich) with a mean molecular weight of 22.5 kg/mol (15-30 kg/mol according to the supplier) was examined at 0.25 µmol/L,

which are 5.6 ppm as for P_5S_3. Assays were buffered to either pH 7.0 or 7.5 with 20 mmol/L Tris-HCl. The temperature was kept constant at 23 °C during the experiments, and samples were shaken at 1,400 rpm. To prove that the silicic acid concentration (1 mmol/L, 60 ppm) was below the supersaturation threshold in our conditions, an assay without pre-autocondensation (i.e. a freshly hydrolyzed TMOS solution) was monitored in parallel.

Results and Discussion

Stabilization of molybdate-reactive silica.

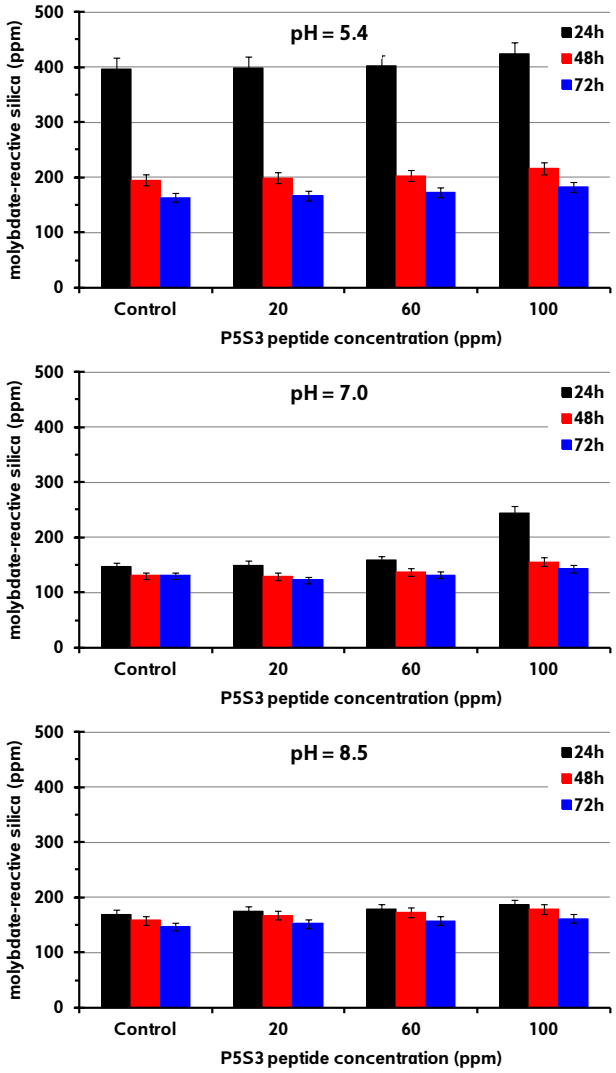

Figure 2. Stabilization of molybdate-reactive silica by polypeptide P_5S_3 during a 3-day period of the condensation reaction at pH 5.4, 7.0, and 8.5, as shown.

Initial attempts to use the polypeptide P_5S_3 as a stabilizing agent for molybdate-reactive silica included "long-term" experiments (up to 3 days). Silicic acid condensation is highly pH-dependent and for the purposes of our work we evaluated three pH values, 5.4, 7.0, and 8.5. As shown in Figure 2, there is no deviation from the "control" in all three pH values (with the exception of pH = 7.0 at 100 ppm (18.0 µmol/L) polypeptide concentration, after 24 h). Hence, based on these results we conclude that polypeptide P_5S_3 is ineffective in stabilizing molybdate-reactive silica for time periods beyond 24 hours.

In an effort to study the behavior of the polypeptide P_5S_3 at the initial stages of the condensation reaction, "short-term" experiments were performed. The pH value of 7.0 was selected for these studies. The results are shown in Figure 3, for an 8-hour duration. A clear stabilization effect is noted. At low P_5S_3 concentrations (20 and 60 ppm, or 3.6 and 10.8 µmol/L) silicic acid stabilization is insignificant (only 17 and 44 ppm above the control after 8 hours, respectively).

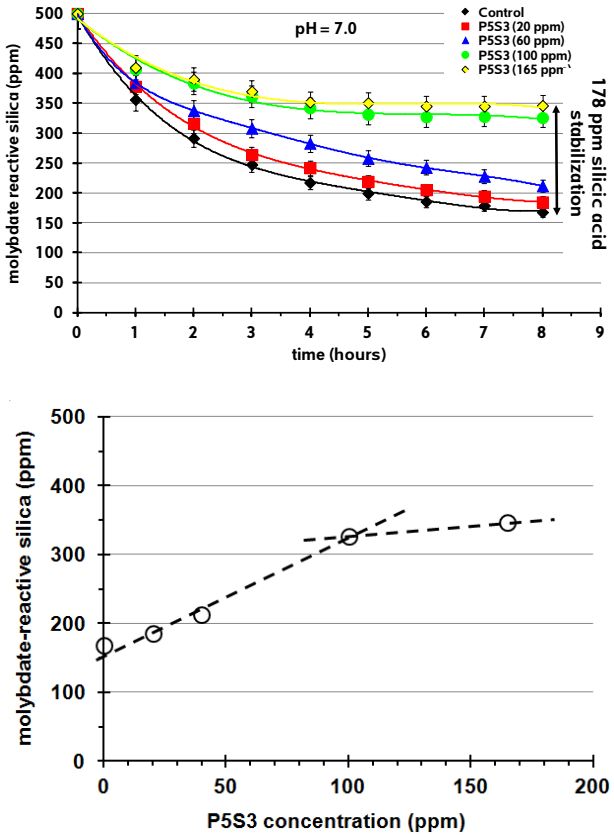

Figure 3. Evolution of loss of molybdate-reactive silica to colloidal silica and concentration dependence vs. time at pH 7.0 (upper). Plot of the measured molybdate-reactive silica levels after 8 hours vs. P_5S_3 polypeptide concentration (lower).

However, upon increase of polypeptide concentration to 100 and 165 ppm (18.0 and 29.7 µmol/L), silicic acid stabilization increases to 158 and 178 ppm above the control, after 8 hours, respectively. The concentration-dependent enhancement in silicic acid stabilization follows an

almost linear trend up to 100 ppm polypeptide concentration. Above that, the effect seems to level off, see Figure 3, lower.

In a circumneutral environment the polypeptide P_5S_3 is positively charged due to its high isoelectric point at 12.64, therefore silicic acid stabilization most likely occurs due to the interaction of the oppositely charged, deprotonated silanol groups and lysine (pK_a = 10.79) plus arginine (pK_a = 12.48) side chains (the most abundant amino acids in the polypeptide). Therefore, we decided to test this hypothesis by "blocking" these cationic residues by a polyanionic electrolyte. We selected a simple homopolymer of acrylic acid, PAA of MW = 2000 Da, which possesses negative charges at pH = 7.0. Figure 4 shows the silicic acid stabilization results at constant P_5S_3 concentration (100 ppm, 18.0 µmol/L) and variable PAA concentration (0, 10, 30, 50, and 100 ppm).

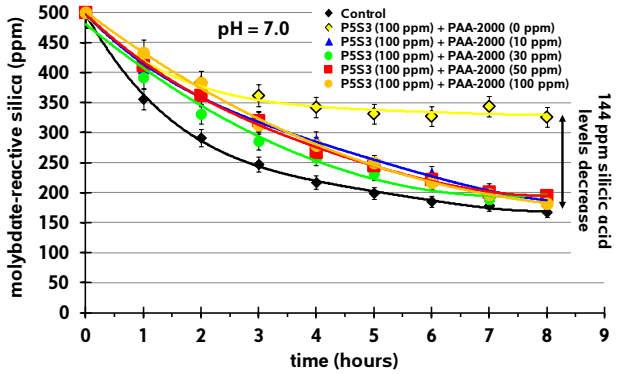

Figure 4. The effect of polyacrylic acid on silicic acid stabilization by the polypeptide P_5S_3.

PAA causes a dramatic reduction (~ 144 ppm after 8 hours) in molybdate-reactive silica, compared to the silicic acid levels in the presence of 100 ppm P_5S_3 alone. This strongly supports the hypothesis that cationic moieties on the polypeptide backbone are responsible for the stabilization effect. PAA "blocks" these moieties through their (partial) neutralization via the $-COO^-$ groups.

Assays at similarly low concentration of silicic acid (5 mmol/L, 300 ppm) were checked for the presence of aggregates by centrifugation after 30 minutes incubation. This time-interval is common for the analysis of silica-formation supported by polycationic agents, which is usually accomplished within a few minutes [18,40-44]. Furthermore, the decrease of molybdate reactive silica in the stabilization assays is fastest both in the control and P_5S_3 samples (Figures 3 and 4). No silica or polypeptide precipitates were found after centrifugation (see Supporting Information Figures S4 and S5), in contrast to assays with higher silicic acid concentration (see below). Instead, the polypeptide remains soluble irrespective of the pH tested or the polypeptide concentration. So the stabilizing interaction between P_5S_3 and silicic acid leaves both partners in a low molecular, soluble form, irrespective of pH (as tested within the range of 4.5 - 8.5) and the applied polypeptide concentration (15 - 75 µmol/L).

Characterization of silica-polypeptide precipitates.

During the 8 hour silicic acid stabilization experiments small amounts of white precipitates appear. These precipitates were studied by FT-IR and SEM, in order to identify the presence (or absence) of polypeptide P_5S_3 in the silica precipitate matrix (with FT-IR) and to evaluate morphological effects of the polypeptide on silica particle texture (SEM). In Figure 5 a portion of the FT-IR spectrum is shown of pure precipitated silica, pure P_5S_3 polypeptide, and precipitated silica in the presence of P_5S_3 polypeptide.

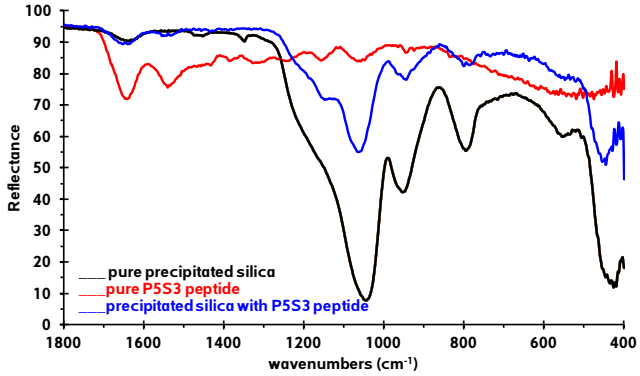

Figure 5. Stacked FT-IR spectra of pure precipitated silica (black), pure P_5S_3 polypeptide (red), and precipitated silica in the presence of P_5S_3 polypeptide (blue).

The unequivocal presence of entrapped P_5S_3 polypeptide in the precipitated silica matrix is evidenced by the amide stretches at 1645 and 1540 cm^{-1} [45].

Furthermore, the silica precipitates in the presence of P_5S_3 polypeptide were studied by SEM and the images are shown in Figure 6.

SEM images were recorded on precipitates formed at pH 7.0, after 8 hours of condensation time, in the absence (control) or presence of various concentrations of polypeptide P_5S_3, as indicated. The control samples display a non-descript morphology, aggregation, and they are lacking any significant geometrical features. Precipitates formed in the presence of 20 ppm P_5S_3 resemble fibers that are loosely attached to one another. Upon concentration increase of P_5S_3 to 60 ppm, a distinct spherical particle morphology becomes evident. These spherical primary particles display a low degree of aggregation. Further polypeptide concentration increase to 100 ppm increases the aggregation tendency, with no apparent change in primary particle size (sub-micron dimensions). Lastly, at 165 ppm polypeptide concentration an increased aggregation tendency is noted. This is in accord with the FT-IR results that strongly suggest that the polypeptide is entrapped into the silica matrix, thus acting as an "organic glue" between the particles. Interestingly, the spherical, aggregated morphology had also been seen in SEM analyses of P_5S_3-silica composites precipitated at higher silicic acid concentrations (i.e. 1,800 ppm, 30 mmol/L) (ref [23] and section

III in this thesis), which reveals that the presence of P_5S_3 has a structure-guiding effect during silica-formation.

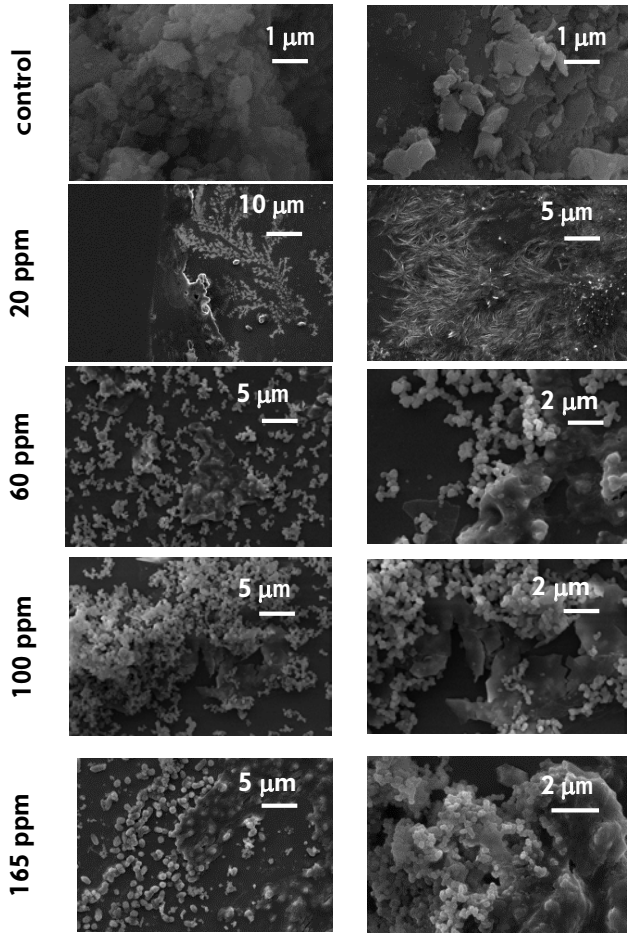

Figure 6. Selected SEM images of silica precipitates (formed at pH 7.0, after 8 hours) in the presence of various concentrations of polypeptide P_5S_3, as indicated. The "control" image was taken on a dried gel sample.

Induced Silica Formation.

P_5S_3 was shown before to strongly enhance silica formation at silicic acid concentrations of 1,800 ppm (30 mmol/L) (ref [23] and section III in this thesis) which is well above the autopolycondensation threshold of 60 ppm (1 - 2 mmol/L) [28] and 3- to 6-fold with respect to the 8.3 mmol/L (500 ppm) and 5 mmol/L (300 ppm) assays above, where P_5S_3 acts as a stabilizer for silicic acid solutions. At 30 mmol/L silicic acid, P_5S_3 was active in silica formation even at concentrations at and below 2 μmol/L (11.1 ppm) at pH 7.5 (Figure 7A), with a constant stoichiometry of silicic acid precipitated per polypeptide, and P_5S_3 being quantitatively trapped in the precipitate (Figure 7B). This co-precipitation of P_5S_3 at high silicic acid concentration is expected, considering its numerous cationic charges which are characteristic for many natural and

biomimetic polypeptides co-precipitating with silica, like silaffin-1A [18], peptide R5 [40,44,46], poly-L-lysine [47] as well as non-peptide compounds, such as diatom polyamines [48].

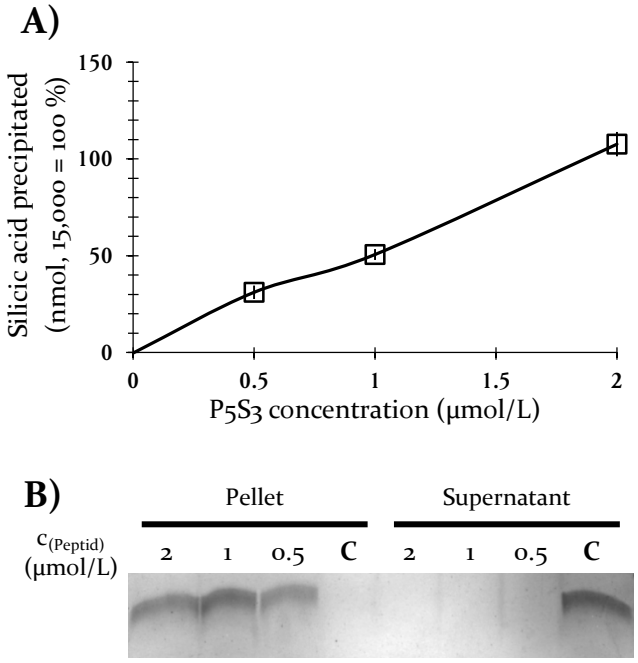

Figure 7. Silica formation at low concentrations of P_5S_3. Silicic acid concentration was 30 mmol/L in an assay volume of 500 µL (total available amount of 15 µmol silicic acid), buffered to pH 7.5 with Tris-HCl. Assays were centrifuged after 30 minutes and pellets were analyzed in terms of silica-quantification (A, a line was drawn to guide the eye) and presence of polypeptide compared with the supernatant (B, lane "C" denotes a control of 2 µmol/L P_5S_3 in absence of silicic acid).

The question arises how much the co-precipitation of P_5S_3 with silica is affected by the pH and P_5S_3-concentration, since P_5S_3 could in part also remain soluble, potentially even as soluble adduct with silica as indicated by the stabilization at 500 ppm (8.3 mmol/L) silicic acid where eventually co-precipitates occurred. This was analyzed at different pH-values, since the activity of biomineralizing peptides is strongly pH-dependent. For example, native silaffin-1A, in which lysine-residues are modified with polyamine chains of different lengths, accelerates silica-formation in slightly acidic environments, whereas the unmodified peptide (termed R5) is only active in near-neutral to basic solutions [18]. By comparison with the silica-biomineralizing peptide PRP1, which induces silica-precipitation already at around 1 unit of pH lower than R5, indications for the dependence on the density of cationically charged amino acids were found [49]. Earlier analysis of P_5S_3-induced silica formation, in a 30 minute interval revealed that the polypeptide is active at pH 6.5 and above (ref [23] and section III in this thesis), similar to the activity of R5 [18].

At the lowest pH values tested (pH 4.5 and 5.5), no precipitation of silica occurred in presence of P_5S_3 (see SI Figure S6) at silicic acid concentration of 30 mmol/L (1,800 ppm) and P_5S_3 concentrations of 30, 75 and 150 µmol/L (166, 417 and 833 ppm). The polypeptide also remained

in the supernatant after centrifugation (Figure 8). At pH 6.5, all P_5S_3 co-precipitated at 30 µmol/L polypeptide, whereas increasing amounts of P_5S_3 remained in the supernatant at the higher concentrations (75 and 150 µmol/L). At pH 7.5 and 8.5, all P_5S_3 was found within the co-precipitate at all concentrations tested. The trend of silica precipitation in presence of P_5S_3 follows the intrinsic pH-dependency of silicic acid to autocondense [28], indicating that P_5S_3 primarily acts by harvesting autocondensed silica from solution, eventually leading to precipitation. The observation that P_5S_3 at high concentrations remains partly in the supernatant at pH 6.5 could either be a result from insufficient silicic acid polymerization or an indication for a temporarily soluble P_5S_3-silica adduct, which will be examined in future investigations.

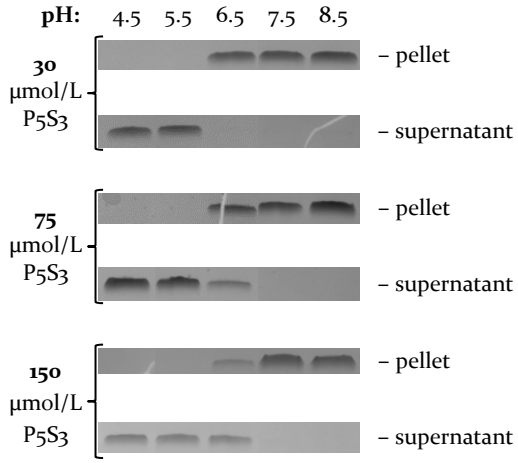

Figure 8. pH and concentration dependent silica formation and partitioning of P_5S_3 between soluble and precipitated fraction. Silicic acid concentration in the assays was 30 mmol/L, P_5S_3 concentrations are noted aside the gel pictures. Buffers used were sodium acetate/acidic acid (pH 4.5), Bis-Tris-HCl (pH 5.5 and 6.5) and Tris-HCl (pH 7.5 and 8.5).

Inhibition of Silica Dissolution.

The fact that P_5S_3 can both stabilize silicic acid and promote its condensation, depending on the silicic acid concentration, which was in both cases well above the autopolycondensation threshold of 1 - 2 mmol/L [28], raises the question whether it may also influence silica dissolution below this threshold. Before having a closer look at P_5S_3-dependent dissolution of precondensed silica we checked that at both pH 7.0 and 7.5 (SI Figures S7 and S8), a solution of 1 mmol/L (60 ppm) monosilicic acid remained stable irrespective of the presence of P_5S_3. The molybdenum blue analysis proved that no condensation occurred within 8 hours, as expected at this low silicic acid concentration.

By contrast, the decondensation of silicic acid oligomers was significantly affected by P_5S_33 (Figures 9 and 10). In both assays (pH 7.0 and 7.5), the initial concentration of molybdate-reactive silica was around 0.2 mmol/L and the decondensation process immediately commenced. In the absence of P_5S_3 it was completed after about 300 minutes at pH 7.0 and, much faster, after 120 minutes at pH 7.5. The presence of 1 µmol/L P_5S_3 decreased the dissolution rate both at pH

7.0 and 7.5. At pH 7.5, where the inhibiting effect was more pronounced, the time required for complete retrieval of 1 mmol/L molybdate reactive silicic acid nearly doubled.

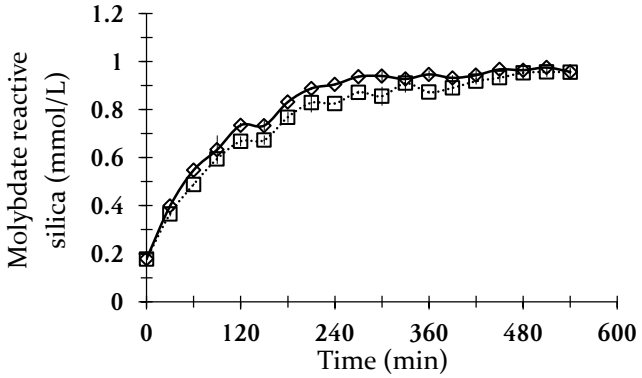

Figure 9. Dissolution of autocondensed silica at pH 7.0. Silicic acid was precondensed at 30 mmol/L (1,800 ppm) for 15 minutes. Dissolution was started by dilution to a globally undersaturated concentration of 1 mmol/L (60 ppm) in the absence (diamonds, solid line) and presence (squares, dotted line) of 1 µmol/L (5.6 ppm) P_5S_3.

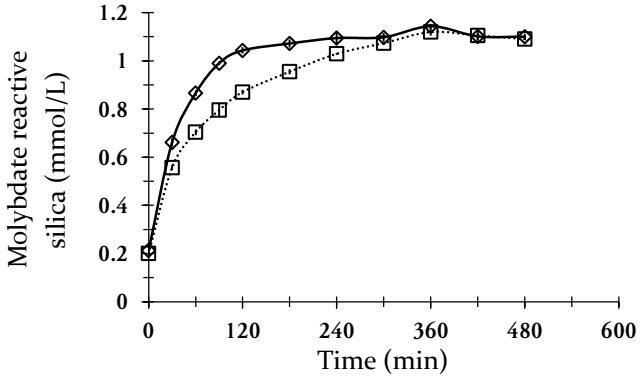

Figure 10. Dissolution of autocondensed silica at pH 7.5. Silicic acid was precondensed at 30 mmol/L (1,800 ppm) for 15 minutes. Dissolution was started by dilution to a globally undersaturated concentration of 1 mmol/L (60 ppm) in the absence (diamonds, solid line) and presence (squares, dotted line) of 1 µmol/L (5.6 ppm) P_5S_3.

Poly-L-lysine, which like P_5S_3 induces the formation of silica at high silicic acid concentrations (e.g. ref's[50,51]), was analyzed in a dissolution assay for comparison. At 0.25 µmol/L (5.6 ppm), which is a similar mass per volume concentration as used for P_5S_3, poly-L-lysine impeded the dissolution of silica at pH 7.5 (SI Figure S9), though this was less pronounced than with P_5S_3. At higher concentration of around 30 µmol/L (672 ppm) however, poly-L-lysine of similar degree of polymerization was reported to accelerate the dissolution of silica in globally undersaturated solution of silicic acid [21]. The retardation of silica dissolution is hence presumably dependent on the ratio of silicic acid per peptide in the assays. The retarding effect exerted by poly-L-lysine presented in this manuscript indicates that electrostatic interactions with silica play a role in this stabilization. However, while all lysine side-chains in poly-L-lysine can be charged by protonation, P_5S_3 has a lower charge density (40 % of the side chains) though is more effective at similar mass per volume concentration. This might hint to further interactions, e.g. *via* hydrogen bonds between silica and serine, which bear hydroxyl-groups that may interact

with silica [52]. This question will be addressed by the tailored design of recombinant peptides, systematically screening the impact of polypeptide size, charge density, and sequence.

Conclusion

The effect of the biomineralizing, polycationic polypeptide P_5S_3 on the condensation and dissolution of silicic acid and silica was examined. P_5S_3 was found to retard the condensation of silicic acid in a supersaturated 500 ppm (8.3 mmol/L) silicic acid solution for an 8-hour term, eventually co-precipitating with silica. The effectiveness in slowing down the condensation increased from 20 to 100 ppm (0 - 18 µmol/L) P_5S_3, after which it leveled off. At increased silicic acid concentration of 1,800 ppm (30 mmol/L), P_5S_3 vice versa induced the formation of silica within 30 minutes. Precipitation of peptide-silica composite occurred above pH 6.5 at a wide range of P_5S_3 concentrations from 2.8 to 833 ppm (0.5 to 150 µmol/L). P_5S_3 was quantitatively trapped in the silica matrix at pH 7.5 and above, while it was partly retrieved from the soluble fraction after the silica-formation reaction at pH 6.5. When pre-condensed silica was diluted to an undersaturated silicic acid concentration of 60 ppm (1 mmol/L), the presence of P_5S_3 at 5.6 ppm (1 µmol/L) transiently stabilized the silica from dissolution, though eventually the silica was fully dissolved. In summary, this shows that P_5S_3 exerts different effects in on the silicic acid condensation and silica dissolution which depend on the respective concentrations of both the polypeptide, and silicic acid.

Corresponding Authors

*demadis@chemistry.uoc.gr (K.D.D.), hobe@uni-mainz.de (S.H.)

Funding Sources

KDD thanks the EU for funding the Research Program SILICAMPS-153, under the ERA.NET-RUS Pilot Joint Call for Collaborative S&T projects. C.Z. is a recipient of a fellowship through the Excellence Initiative (DFG/GSC 266) in the context of the graduate school of excellence "MAINZ" (Materials Science in Mainz).

Abbreviations

SDV = silica deposition vesicle, STV = silicon transport vesicle, SIT = silicon transporter, MW = molecular weight, SEM = scanning electron microscopy, PAA = polyacrylic acid, FTIR = fourier transform infrared spectroscopy, Tris = tris(hydroxymethyl)aminomethane, Bis-Tris: Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan.

References

[1] K.H. Wedepohl, Geochim. Cosmochim. Acta 59 (1995) 1217.

[2] D. Otzen, Scientifica 2012 (2012) Article ID 867562.

[3] N.C. Brady, R. Weil, The Nature and Properties of Soils, 14th ed., Prentice Hall, 2007.

[4] H. Wiese, M. Nikolic, V. Römheld, in:, B. Sattelmacher, W.J. Horst (Eds.), Apoplast High. Plants Compart. Storage, Transp. React., Springer, 2007.

[5] K.D. Bidle, M. Manganelli, F. Azam, Science 298 (2002) 1980.

[6] C.C. Perry, Rev. Mineral. Geochemistry 54 (2003) 291.

[7] H. Ehrlich, Biological Materials of Marine Origin. Invertebrates, Springer, 2010.

[8] J.P. Smol, E.F. Stoermer, eds., The Diatoms: Applications for the Environmental and Earth Sciences, 2nd ed., Cambridge University Press, 2010.

[9] M. Hildebrand, B.E. Volcani, W. Gassmann, J.I. Schroeder, Nature 385 (1997) 688.

[10] M. Hildebrand, K. Dahlin, B.E. Volcani, Mol. Gen. Genet. 260 (1998) 480.

[11] K. Thamatrakoln, A.J. Alverson, M. Hildebrand, J. Phycol. 42 (2006) 822.

[12] P. Curnow, L. Senior, M.J. Knight, K. Thamatrakoln, M. Hildebrand, P.J. Booth, Biochemistry 51 (2012) 3776.

[13] R. Siever, Am. Mineral. 42 (1957) 821.

[14] P. Tréguer, D.M. Nelson, A.J. Van Bennekom, D.J. Demaster, A. Leynaert, B. Quéguiner, Science 268 (1995) 375.

[15] S.-X. Li, Z.-H. Wang, B.A. Stewart, in:, Adv. Agron., 2013, pp. 205–397.

[16] W.E.G. Müller, M.A. Grachev, eds., Biosilica in Evolution, Morphogenesis, and Nanobiotechnology, Springer Berlin Heidelberg, 2009.

[17] V. Martin-Jézéquel, M. Hildebrand, M.A. Brzezinski, J. Phycol. 36 (2000) 821.

[18] N. Kröger, R. Deutzmann, M. Sumper, Science 286 (1999) 1129.

[19] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, Science 298 (2002) 584.

[20] N. Poulsen, N. Kröger, J. Biol. Chem. 279 (2004) 42993.

[21] S. V Patwardhan, G.E. Tilburey, C.C. Perry, Langmuir 27 (2011) 15135.

[22] H. Ehrlich, K.D. Demadis, O.S. Pokrovsky, P.G. Koutsoukos, Chem. Rev. 110 (2010) 4656.

[23] C. Zerfaß, S. Braukmann, S. Nietzsche, S. Hobe, H. Paulsen, Protein Expr. Purif. 108 (2015) 1.

[24] B.E. Kemp, D.J. Graves, E. Benjamini, E.G. Krebs, J. Biol. Chem. 252 (1977) 4888.

[25] A. Kuliopulos, Proc. Natl. Acad. Sci. 84 (1987) 8893.

[26] Eur. J. Biochem. 138 (1984) 9.

[27] G.A. Icopini, S.L. Brantley, P.J. Heaney, Geochim. Cosmochim. Acta 69 (2005) 293.

[28] R.K. Iler, The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry, Wiley, New York, 1979.

[29] T. Coradin, J. Livage, Colloids Surf. B. Biointerfaces 21 (2001) 329.

[30] T. Coradin, D. Eglin, J. Livage, Spectroscopy 18 (2004) 567.

[31] G.B. Alexander, J. Am. Chem. Soc. 75 (1953) 5655.

[32] V.W. Truesdale, P.J. Smith, C.J. Smith, Analyst 104 (1979) 897.

[33] V.W. Truesdale, C.J. Smith, Analyst 100 (1975) 797.

[34] V.W. Truesdale, C.J. Smith, Analyst 100 (1975) 203.

[35] A.D. Eaton, L.S. Clesceri, E.W. Rice, A.E. Greenberg, M.. Franson, Stand. Methods Exam. Water Wastewater, Am. Public Heal. Assoc. (2005).

[36] L. Jiang, L. He, M. Fountoulakis, J. Chromatogr. A 1023 (2004) 317.

[37] H. Schägger, Nat. Protoc. 1 (2006) 16.

[38] G. von Jagow, H. Schaegger, A Practical Guide to Membrane Protein Purification, Academic Press, San Diego, 1994.

[39] H. Schaegger, G. von Jagow, Anal. Biochem. 166 (1987) 368.

[40] C.C. Lechner, C.F.W. Becker, Chem. Sci. 3 (2012) 3500.

[41] S. Patwardhan, N. Mukherjee, S. Clarson, Silicon Chem. 1 (2002) 47.

[42] E. Brunner, K. Lutz, M. Sumper, Phys. Chem. Chem. Phys. 6 (2004) 854.

[43] R.R. Naik, P.W. Whitlock, F. Rodriguez, L.L. Brott, D.D. Glawe, S.J. Clarson, M.O. Stone, Chem. Commun. (2003) 238.

[44] L. Senior, M.P. Crump, C. Williams, P.J. Booth, S. Mann, A. Periman, P. Curnow, J. Mater. Chem. B 3 (2015) 2607.

[45] K. Nakamoto, Infrared and Raman Spectra of Inorganic and Coordination Compounds, Theory and Applications in Inorganic Chemistry, Wiley-Interscience: New York, 1997.

[46] M.R. Knecht, D.W. Wright, Chem. Commun. (2003) 3038.

[47] N. Li, X. Zhang, Q. Wang, F. Wang, P. Shen, RSC Adv. 2 (2012) 3288.

[48] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14133.

[49] H. Kauss, K. Seehaus, R. Franke, S. Gilbert, R.A. Dietrich, N. Kröger, Plant J. 33 (2003) 87.

[50] S. V Patwardhan, S.J. Clarson, Silicon Chem. 1 (2002) 207.

[51] D. Belton, G. Paine, S. V Patwardhan, C.C. Perry, J. Mater. Chem. 14 (2004) 2231.

[52] K. Spinde, K. Pachis, I. Antonakaki, S. Paasch, E. Brunner, K.D. Demadis, Chem. Mater. 23 (2011) 4676.

Supporting Information

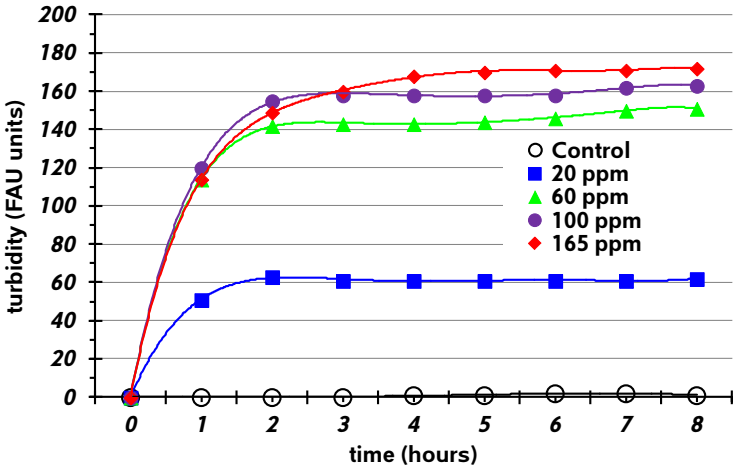

Figure S1. Evolution of turbidity over time in solutions containing various concentrations of P_5S_3.

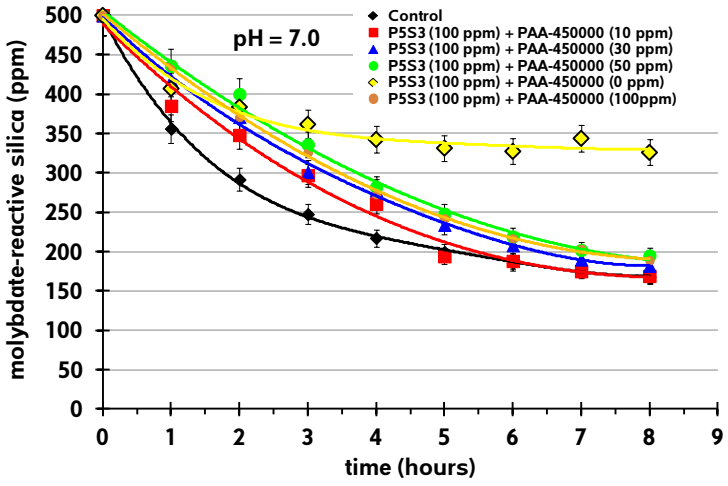

Figure S2. Stabilization of molybdate-reactive silica in the presence of P_5S_3 (100 ppm) and various concentrations of PAA (polyacrylic acid, MW 450 kDa).

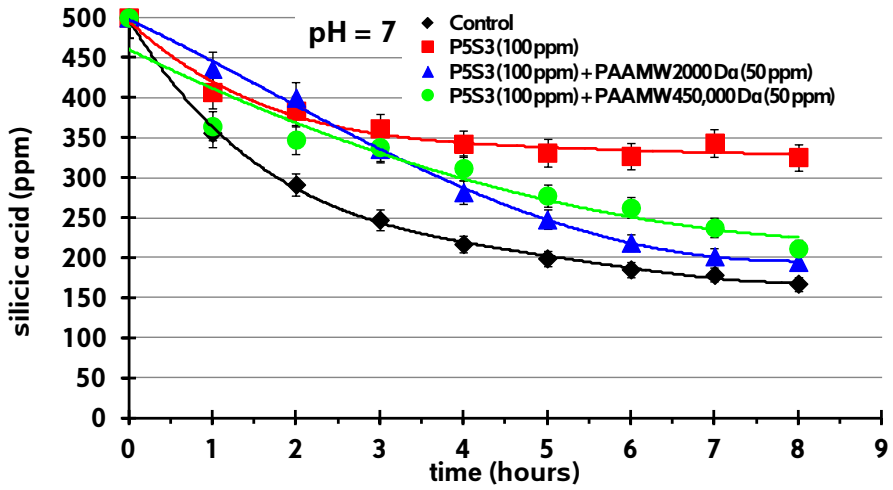

Figure S3. Comparison of stabilization of molybdate-reactive silica in the presence of P_5S_3 (100 ppm) and PAA (polyacrylic acid) of MW 2 kDa (short) and 500 kDa (long).

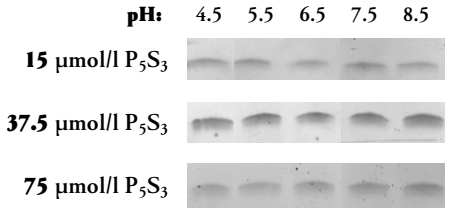

Figure S4. P_5S_3 at stabilizing condition stays soluble. P_5S_3 was incubated for 30 minutes at 5 mmol/L (300 ppm) silicic acid and different concentrations of P_5S_3 (noted in the left-hand margin) at the given pH values, and centrifuged 5 min at 18,400 g. Buffers used were sodium acetate/acetic acid (pH 4.5), Bis-Tris-HCl (pH 5.5) and Tris-HCl (pH 7.5 and 8.5).

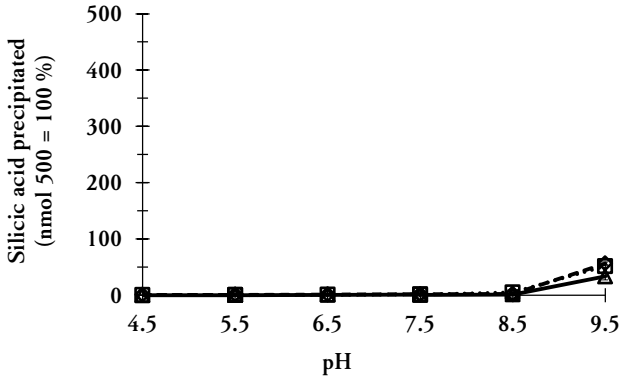

Figure S5. P_5S_3-induced silicic acid precipitation. Silicic acid concentration was 5 mmol/L in 100 µl assays (total available amount of 500 nmol). P_5S_3 was used at 15 (triangles, solid line), 37.5 (squares, dotted line) and 75 µmol/L (diamonds, dashed line). The assays were buffered with sodium acetate (pH 4.5 and 5,5), Bis-Tris-HCl (5.5, 6.5. 7.5) or Tris-HCl (pH 7.5, 8.5 and 9.5). Lines were drawn to guide the eye.

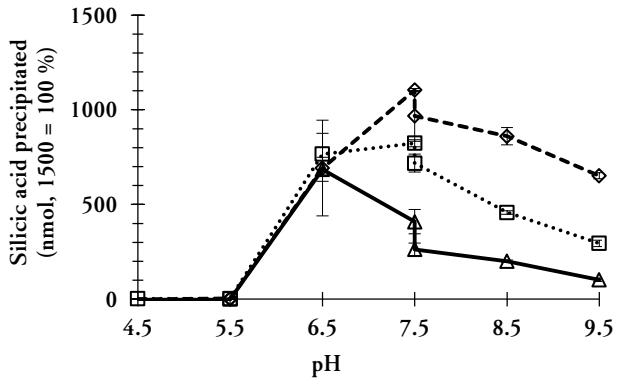

Figure S6. P_5S_3-induced silicic acid precipitation. Silicic acid concentration was 30 mmol/L in 50 µl assays (total available amount of 1500 nmol). P_5S_3 was used at 30 (triangles, solid line), 75 (squares, dotted line) and 150 µmol/L (diamonds, dashed line). The assays were buffered with sodium acetate (pH 4.5 and 5,5), Bis-Tris-HCl (5.5, 6.5. 7.5) or Tris-HCl (pH 7.5, 8.5 and 9.5). Lines were drawn to guide the eye.

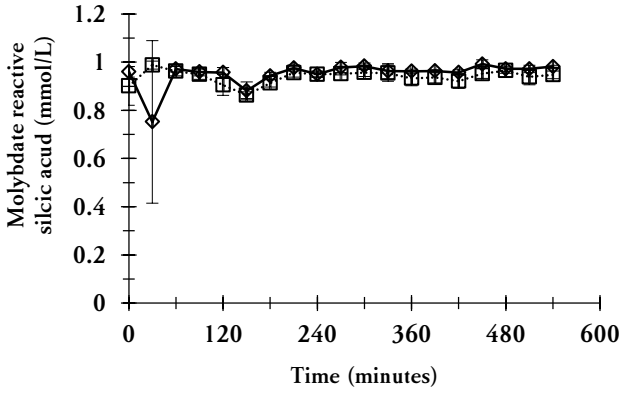

Figure S7. Stability of 1 mmol/L (60 ppm) monosilicic acid at pH 7.0 in the absence (diamonds, solid line) and presence (squares, dotted line) of 1 µmol/L (5.6 ppm) P_5S_3.

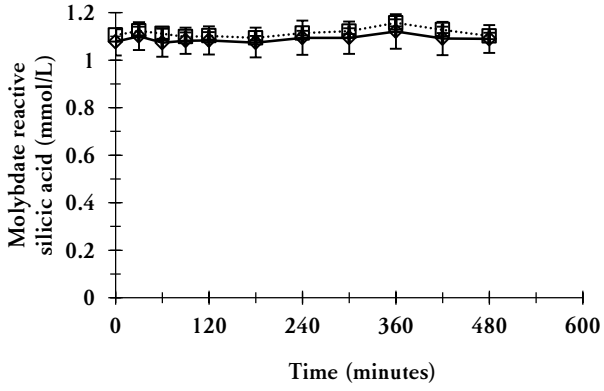

Figure S8. Stability of 1 mmol/L (60 ppm) monosilicic acid at pH 7.5 in the absence (diamonds, solid line) and presence (squares, dotted line) of 1 µmol/L (5.6 ppm) P_5S_3.

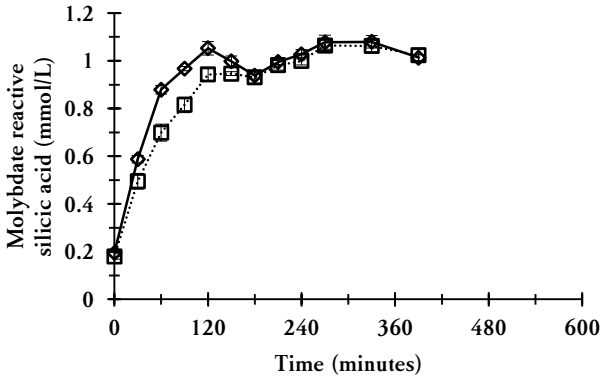

Figure S9. Dissolution of autocondensed silica at pH 7.5 in absence or presence of poly-L-lysine. Silicic acid was precondensed at 30 mmol/L (1,800 ppm) for 15 minutes. Dissolution was started by dilution to a globally undersaturated concentration of 1 mmol/L (60 ppm) in the absence (diamonds, solid line) and presence (squares, dotted line) of 0.25 µmol/L (5.6 ppm) poly-L-lysin.

Section V - Probing a designed biomineralizing polypeptide during the interaction with silica: A circular dichroism spectroscopy approach

Christian Zerfaß [a,b], Harald Paulsen [a], Stephan Hobe [a]

[a] Institute of General Botany, Johannes Gutenberg University, Muellerweg 6, 55128 Mainz, Germany

[b] Graduate School Materials Science in Mainz, Staudinger Weg 9, 55128 Mainz, Germany

This section is a publication manuscript in preparation.

Highlights

- Structural changes in a designed biomineralizing polypeptide are studied during the interaction with silica.
- Circular dichroism spectra reveal an extended conformation in solution, which is partly lost upon silica-formation and polypeptide-silica co-precipitation.
- Even at high salt concentrations, a polypeptide-silica interaction is observed and attributed to the contribution of non-electrostatic forces.
- Monitoring the silica dissolution in th epresence of the polypeptide by time-resolved CD reveals two distinct phases of polypeptide behavior.

Abstract

Polypeptides play an important role in biomineralization, but little is known about the impact of conformation and dynamics of biomineralizing polypeptides. In this work, circular dichroism spectroscopy was used to analyze a designed silica-biomineralizing polypeptide bearing numerous cationic charges. The polypeptide-silica interaction was investigated during co-precipitation starting out from silicic acid as well as during the dissolution of pre-formed silica. The polypeptide in part lost its disordered conformation when occluded in a silica matrix, and residual interaction remained at high-salt conditions. The changes in the polypeptide conformation were used to monitor the dynamics of the interaction with silica by time-resolved circular dichroism spectroscopy during silica-formation and -dissolution.

Keywords

Biomineralizing Peptide; Silica-formation and -dissolution; Circular Dichroism Spectroscopy

Abbreviations

Bis-Tris: Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane. CD: circular dichroism. DM: dodecyl maltoside. PAGE: polyacrylamide gel electrophoresis. SDS: sodium dodecyl sulfate. TMOS: tetramethoxysilane. Tris: tris(hydroxymethyl)aminomethane. UV: ultraviolet.

Introduction

Diatoms build their silicified cell walls by using silaffin peptides as well as poly(propylamine)s to accelerate and control the formation of silica from soluble pre-cursors [1-6]. In addition, further polypeptides [7] are associated with the silica-matrix, such as frustulins [8,9] and pleuralins [10,11], or the recently identified cingulins [12], which are thought to be structure-guiding agents. Silaffins from *Cylindrotheca fusiformis* were found to initially self-assemble mediated by the presence of phosphate [13], undergoing a phase separation [14] and thereby forming an environment in which silicic acid polycondensation is facilitated. Beyond, covalently attached polyamines are thought to catalyse the siloxane bond formation by acid-base catalysis, by an interplay between protonated and unprotonated amines [15].

Without the polyamine moiety, silaffins generally retain their ability to accelerate silica formation, albeit only at neutral pH and above whereas native silaffins are active in a slightly acidic environment [2]. This was shown with the peptide R5, a segment of the silaffin from *C. fusiformis*. It is thought that R5 first interacts with phosphate to form a micelle-like assembly [16], and that this self-assembly process is a prerequisite of silica biomineralizing activity [13]. In this micellar structure, only a limited number of amino acid side chains is exposed to the bulk solution, and NMR-analysis revealed that only amino-terminal amino acids of R5 interact with the silica-matrix upon precipitation [17]. Poly-L-lysin (PLL) can also phase-separate in the presence of phosphate to produce hollow silica spheres [18] or hexagonal platelets [19] via a surface deposition on the assembled matrix.

The conformation of silaffins in solution, in the absence of silicic acid, was predicted to be intrinsically disordered by computational approaches [20] and proven for the R5-peptide by circular dichroism and nuclear magnetic resonance spectroscopy [21]. This structural classification comprises ensembles of transient structures with local structural preferences [22]. Increasing the net charge per residue in intrinsically disordered proteins leads to a preference of extended structures, as shown with arginine-rich protamine, which has an increased content of polyproline II (PPII) conformation when the cationic charge is elevated [23].

In solution at neutral pH, PLL as well adopts a disordered conformation [24-27] which is comprised of a high fraction of extended conformations including PPII. During co-precipitation with silica, the disordered structure of PLL of low degree of polymerization (< 100) is retained [28-30] and spherical silica is deposited. However, at degrees of PLL-polymerization above 100 residues [31] and the presence of phosphate (at least 2.5 phosphate per lysine [32]) PLL

undergoes a transition into an α-helical conformation, in which it self-assembles into single crystalline sheets [31] which in turn template the deposition of hexagonal silica platelets [19,32,33].

Here, the structure of P_5S_3 [34], a designed polycationic polypeptide with silica-biomineralizing activity, was assessed during its interaction with silica to gain insight into the polypeptide dynamics during mineralization reactions. The P_5S_3 sequence is derived from the silaffin of *Cylindrotheca fusiformis* [2] and contains high numbers of lysine, arginine, and serine. Kinase target sites allow the *in vitro* enzymatic phosphorylation. In solutions of silicic acid far above the supersaturation threshold (i.e. at 30 mmol/L), P_5S_3 both in its phosphorylated and non-phosphorylated state accelerates the formation of silica at pH 6.5 to 9.5 [34]. However, at lower, but still supersaturated concentrations of silicic acid (8.3 mmol/L), it impedes the polymerization to silica, and additionally slows down the dissolution of silica in globally undersaturated (1 mmol/L) solutions (A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis). In this report, the steady-state conformation of P_5S_3 in the silica matrix and the structural changes during silica-interaction in silica-dissolution experiments was analyzed by circular dichroism (CD) spectroscopy.

Materials and Methods

Protein expression and purification.

Expression and purification of P_5S_3 was done as described earlier ([34] and section III of this thesis). The polypeptide consists of five protein kinase A target sites (sequence RRASL) interspersed with three silaffin-derived lysine-rich partial sequences (GKSKKL) and an amino-terminal tyrosine-containing extension (GSYS) for UV/vis-spectroscopic quantification via absorption at 280 nm (using the absorption coeficient of tyrosine [35]). Occasionally, a version with a carboxy-terminal hexahistidyl-tag (extension EHHHHHH) was used, but is not specified since it did not differ in its behavior compared with the tag-free polypeptide in the analyzes described here, even though it contained a single acidic amino acid (glutamate). For expression and purification, a fusion of P_5S_3 with ketosteroid isomerase [36] was used. The fusion mediates the polypeptide deposition in inclusion bodies upon expression and can be cleaved off by cyanogen bromide [37]. The P_5S_3 target polypeptide is eventually isolated by cation exchange chromatography.

Circular dichroism spectroscopy of P_5S_3 in solution and during interaction with silica.

Circular dichroism spectroscopy was done in a JascoJ 810 spectropolarimeter (Jasco GmbH, Gross-Umstadt, Germany) equipped with a Jasco CDF-426S Peltier temperature controlling element. All measurements were done at 23 °C. Quartz cuvettes with either 1 cm or 0.1 mm

(mountable cuvette) were used, depending on the P_5S_3 concentration analyzed, which was usually 1 or 100 µmol/L, respectively. Secondary structure fractions were estimated from the CD spectra by using the CDPro software package [38,39]. This package contains three different algorithms (ContinII, Selcon3, CDSTTR), but since Selcon3 and CDSTTR largely failed in the assignement (considerable deviation in the calculated from the measured spectra), only ContinII results are presented.

P_5S_3 spectra were recorded in solution at different pH, sodium chloride concentrations, and in the presence of the detergents sodium dodecylsulfate (SDS, C. Roth GmbH, Karlsruhe (Germany)) and dodecyl maltoside (DM, Merck Chemicals GmbH, Schwalbach (Germany)).

Orthosilicic acid was prepared by hydrolysis of 150 mmol/L TMOS (tetramethoxysilane) in 1 mmol/L hydrochloric acid for 15 minutes (23 °C, 1,400 rpm). Oligo- and polymeric silicic acid species were produced by pre-condensation at 30 mmol/L silicic acid, at different pH depending on the later measurement conditions (see below). For silica dissolution assays, a 30 mmol/L silicic acid solution was buffered with 15 mmol/L Tris-HCl to pH 7.0 or 7.5, according to the pH aspired in the assay. After 15 minutes at 23 °C and 1,400 rpm, the assay was diluted to 1 mmol/l, by which P_5S_3 was added to 1 µmol/L. No additional buffer was added, but the pH verified by using pH indicator strips. For time-resolved CD of the silica formation in the presence of P_5S_3 at pH 6.5, where the autopolycondensation of silicic acid is slowed down, a 60 mmol/L solution of orthosilicic acid was pre-incubated for 15 minutes (23 °C, 1,400 rpm) in 10 mmol/L Tris-HCl at pH 9.5 to accelerate the condensation [40,41]. The assay was diluted to 30 mmol/L silicic acid, thereby adding Tris-HCl pH 6.5 to 15 mmol/L, P_5S_3 to 100 µmol/L, and hydrochloric acid to protonate residual Tris from the pre-incubation (residual Tris was calculated according to the Henderson-Hasselbalch equation [42,43]).

P_5S_3-induced silica-formation in the presence of sodium chloride.

A solution containing P_5S_3 was supplemented with Tris-HCl buffer (50 mmol/L, pH 7.5), 30 mmol/L acidically hydrolysed TMOS (15 minutes in 1 mmol/L hydrochloric acid), and different concentrations of sodium chloride. After 30 minutes at 23 °C and 1,400 rpm, silica was harvested by centrifugation (5 min at 18,400 g), resuspended in water for washing, and harvested again. Both supernatant and pellet fraction were stored. For quantification of silicic acid, the pellet fraction was hydrolyzed (1 mol/L sodium hydroxide for 1 hour at room temperature) and subjected to molybdenum blue silicic acid quantification ([34,44] and section III of this thesis). For localization of P_5S_3, sodium hydroxide was additionally added to the supernatant fraction to achieve a concentration of 1 mol/L. After hydrolysis, P_5S_3 was enriched and desalted by trichloroacetic acid (TCA) protein precipitation as described elsewhere ([45] and A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis). Eventually, the isolated polypeptide-fraction was analyzed by SDS-PAGE (see below).

P_5S_3 desorption from pre-formed silica.

Silica was produced with 0.5 mmol/L spermine at 30 mmol/L acidically hydrolyzed TMOS (hydrolysis for 15 minutes in 1 mmol/L hydrochloric acid), buffered to pH 7.5 with 50 mmol/L Tris-HCl. After 30 minutes of silica formation reaction, silica was harvested by centrifugation (5 min at 6,570 g). The pellet was resuspended in 1,000 mmol/L sodium chloride buffered to pH 7.5 with 50 mmol/L Tris-HCl to remove any surface-bound spermine, and afterwards washed two times in buffer without salt. To collect silica after each washing procedure, the assays were centrifuged and the supernatants discarded.

The as prepared silica was subsequently suspended in 15 µmol/L P_5S_3, in 50 mmol/L Tris-HCl pH 7.5, and incubated for 2.5 minutes. After silica-collection by centrifugation, the pellet was again washed with 50 mmol/L Tris-HCl pH 7.5. Eventually P_5S_3 was desorbed with sodium chloride solution, silica removed by centrifugation, and the supernatant analyzed by SDS-PAGE. Nine parallel precipitation assays (with spermine) were prepared and P_5S_3 elution was examined in parallel assays with sodium chloride solutions at 300 to 1,000 mmol/L (in steps of 100 mmol/L) buffered by 50 mmol/L Tris-HCl pH 7.5.

Miscellaneous methods.

For SDS-PAGE analyzes, 16.5 % acrylamide gels (total concentration in the separating gel) in the Tris/Tricine system were used [46-48]. Several buffer systems were used for analysis of CD-spectra at different pH: sodium acetate/acetic acid at pH 4.5, Bis-Tris-HCl at pH 6.5, Tris-HCl at pH 7.5 and 9.5. All solutions for silica-formation and CD-spectroscopic analyzes as well as molybdenum blue silicic acid quantification were prepared with ultra pure water (electrical resistance 18.6 MΩ). TMOS and spermine were purchased from Sigma-Aldrich GmbH, Fluka, Taufkirchen (Germany).

Results and Discussion

In aqueous solution, P_5S_3 adopts to a large extent an extended polyproline II (PPII) structure ([34] and section III of this thesis). The CD spectrum of this motif is characterized by a strong negative ellipticity at 195 - 200 nm [49], which is often insufficiently interpreted as random coil, or unordered conformation [50,51]. A positive ellipticity around 220 nm is usually associated with PPII [49], but often only appears at low temperatures [52], where it was found for P_5S_3 [34].

Conformational changes in P_5S_3 upon interaction with silica.

The polyproline II-structure of P_5S_3 was prominent at pH 4.5, 7.5 and 9.5 as revealed by the strongly negative ellipticity at 195 - 200 nm, though most pronounced at the most acidic pH

(Figure 1). At basic pH 9.5, the 220 nm region was notably negative. This might be the result of a partial transition into either β-sheet or α-helix which both have a negative CD-spectrum in this region [51]. Considering the high content of lysine in P_5S_3, comparison with the pH-dependant conformation of poly-L-lysine renders it most likely that the negative ellipticity around 220 nm is a result from an increase in α-helical content. In poly-L-lysine, in which the pK_a of side-chain N_ε amines is around 9.8 [53], the reduction in charge during increase of the pH from neutral to around 9.5 is associated with a partial transition from extended to α-helical conformation [25,26,54]. It can hence be expected that the lysine-clusters in P_5S_3 undergo a similar transition with increasing pH, which may also have an impact on the conformation of adjacent residues [55,56].

The conformation of P_5S_3 in the presence of silicic acid and upon co-precipitation with silica was assessed by measuring the CD-spectra at different pH, hence allowing or prohibiting silicic acid autopolycondensation. At pH 4.5 no significant changes in the CD-spectra were seen in the presence of either 30 or 60 mmol/L silicic acid. Since silicic acid has a very slow condensation rate at this pH [41], it can be assumed that virtually only orthosilicic acid was present during the measurement, and that this silicic acid species did not affect the P_5S_3 conformation. However, at pH 7.5 and 9.5, where P_5S_3 rapidly precipitates with silica [34], strong spectral changes occurred. At pH 7.5 with ratios of silicic acid per peptide of 150 and 300 (15 and 30 mmol/L silicic acid) the polyproline II (or "disordered") fraction was reduced, as revealed by the reduction of the 195 - 200 nm negative ellipticity. The 220 nm region became more negative, indicating a partial adoption of α-helix or β-sheet structure [51]. At pH 9.5 and a ratio of 150 Si per peptide, the spectrum was very close to that of an α-helix (positive ellipticity below 200 nm, double negative ellipticity peak at 207 and 228 nm).

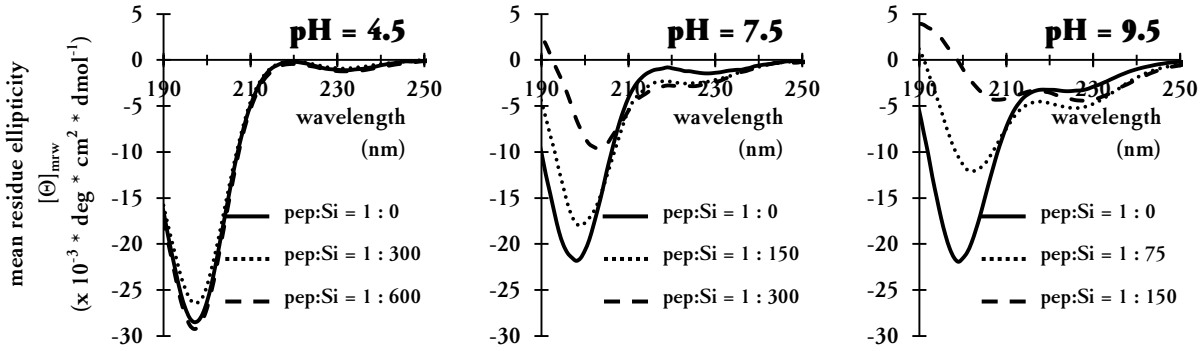

Figure 1. Changes in the CD-spectrum of P_5S_3 in response to interaction with silicic acid. Spectra were recorded at pH 4.5 (no precipitation of silica), 7.5 and 9.5 (both conditions with silica precipitation). Concentration of P_5S_3 was 100 µmol/L in all assays, buffer concentration was 20 mmol/L.

An estimation of the different secondary stucture motifs from the CD-spectra was done by using the CDPro software package [38,39] (Supplementary Information Figures S1-3). ContinII identified the decrease of disordered conformation (up to 15 % decrease in combined "turn" and

"unordered") with a concomitant increase of α-helical and β-sheet structure (alike around 15 %) when P_5S_3 was analyzed after silica formation.

The CD-spectrum of P_5S_3 in solution is similar to that reported for the silaffin-derived peptide R5 [21]. Nuclear magnetic resonance spectroscopy of R5 revealed that the peptide adopts an ensemble of disordered conformations. The effect of silicic acid on the structure of R5 was analyzed in 2 mmol/L silicic acid (above the autopolycondensation threshold), but did not have any effect on the CD-spectrum. However, a reduced spectral intensity of the disordered spectrum, as found here for P_5S_3, was reported for poly-L-lysine chains with degree of polymerization below 100 residues [18,19], which, like P_5S_3, were disordered in solution, with a high fraction of extended polyproline II conformation [24,25,27,49]. PLL above 100 residues adopted an α-helical conformation during silica-formation at 50 - 100 mmol/L silicic acid [19], which is similar to the P_5S_3 behavior at basic pH. However, the PLL conformational transition strictly relied on the presence of phosphate [32] and was driven by the assembly into hexagonal single-crystals [31]. Given that the α-helix conformation of poly-L-lysin was allowed after side-chain charge neutralization by countercharged phoshate (an effect which was also observed for other polypeptides bearing numerous cationic charges [23,57,58]), it seems likely that some residues in P_5S_3 adopt this conformation after silica-formation. Especially at pH 9.5, 50 % of the lysines are uncharged according to their pK_a [53], and additional neutralization is possible by ionic interaction with the acidic silica surface. Together, this would reduce the repulsion between charged side-chains, which in turn may lead to a reduction of the propensity of extended secondary structures [23,27,49]. The reduction of the polyproline II or disordered CD-spectrum may also result from an interaction between peptide-bond amides, which are exposed in extended conformations, and silica, an interaction proposed earlier for the silica-biomineralizing peptides KSL [59,60] and silaffin-1 [61].

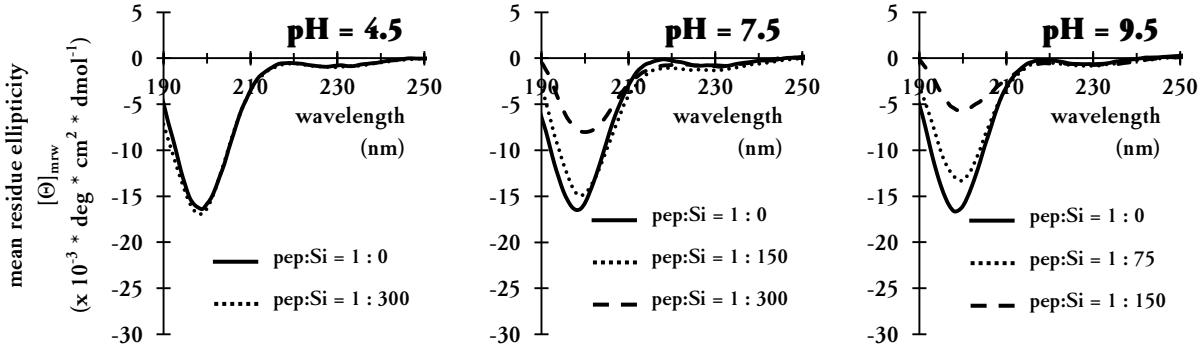

Figure 2. Changes in the CD-spectrum of phosphorylated P_5S_3 in response to interaction with silicic acid. Spectra were recorded at pH 4.5 (no precipitation of silica), 7.5 and 9.5 (both conditions with silica precipitation). Concentration of P_5S_3 was 100 µmol/L in all assays, buffer concentration was 20 mmol/L.

To examine the impact of the polypeptide net charge on the structure of P_5S_3, phosphorylations, a common modification in silaffins which mediate their self-assembly [13], were introduced by enzymatic phosphorylation using protein kinase A [34,62]. The phosphorylated P_5S_3 still

facilitates silica-formation, though with slightly reduced silica yield in comparison to the unmodified P_5S_3 ([34] and section III of this thesis).

The phosphorylated P_5S_3 was still in a dominantly disordered polyproline II conformation (Figure 2), but far less so than the unmodified polypeptide, as judged from the spectral intensity in the 195 - 200 nm region. The mean residue ellipticity from 190 to 210 nm did not significantly differ between pH 4.5, 7.5 and 9.5 indicating that the conformation remained stable. Likewise, the ellipticity at 220 nm was equally close to zero under all conditions. At pH 4.5, the spectrum remained stable after addition of silicic acid, not indicating an interaction with orthosilicic acid, and at pH 7.5 and 9.5, where silica-formation is expected, the intensity of the spectrum of PPII decreased with increasing the silicic acid per peptide ratio. While this was similar to the observations with the unmodified P_5S_3, interaction of phosphorylated P_5S_3 with silica did not produce any CD signature indicative for α-helix and β-sheet conformation. The estimation of secondary structures by ContinII (Figures S4-6) supported the interpretation of a reduction in disordered conformation (around 15 %) and found primarily an increase in β-sheet structure (around 10 %). However, since the ellipticity at 210 - 230 nm remained unchanged, an increase in α-helix and β-sheet conformation remains questionable since this would imply an increasingly negative ellipticity in this region [51].

The different behavior of phosphorylated and unmodified P_5S_3 could in principle be a result of two different effects. Firstly, the reduced net cationic charge of P_5S_3 might result in a less intimate association between the polypeptide and silica. The local repulsion between phosphate-groups and the acidic silica surface should thereby affect the overall peptide conformation and structural adaptation upon binding to silica. Secondly, phosphorylations on serine side chains can form hydrogen bonds with adjacent peptide bonds [63] by which they stabilize the polyproline II extended conformation [64,65]. This would imply that in the unmodified P_5S_3, the site of α-helix adoption upon silica-binding lies within the protein phosphorylation site (sequence RRASL), but this remains to be elucidated.

SDS micelles as a model anionic surface.

To further investigate the impact of side-chain charge neutralization in P_5S_3 upon interaction with anionic surfaces, CD-spectra of P_5S_3 were recorded in the presence of sodium dodecylsulfate (SDS). SDS micelles were used to mimic membrane-like environments [66] in studies of small polypeptides which are intrinsically disordered in solution, but adopt more rigid conformations upon interaction with SDS below and above its critical micellar concentration. Examples of polypeptides which showed this behavior include the biomineralizing enamel protein amelogenin [67], Parkinson's disease associated α-synuclein [68], Alzheimer's disease associated amyloid Aβ peptide [69], and porcine peptide hormone motilin [70]. Polypeptides bearing cationic side chains are thought to be bound to the surface of SDS micelles [71,72]. In this case, the interaction of

P_5S_3 with SDS and silica, as far as the electrostatic contribution is concerned, should be quite similar.

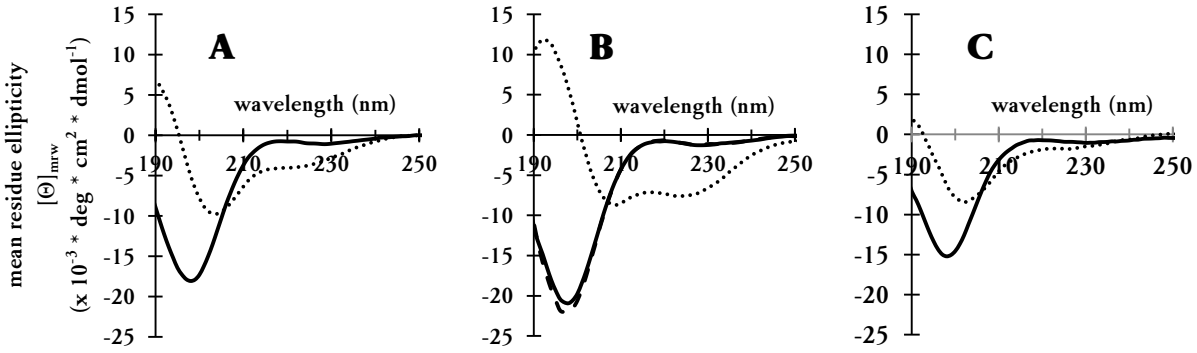

Figure 3. Detergent-induced changes in P_5S_3 secondary structure examined by CD-spectroscopy. ***A)*** 100 µmol/L P_5S_3 in the absence (*solid line*) or presence (*dotted line*) of 34.7 mmol/L (1 % w/v) sodium dodecylsulfate. ***B)*** 1 µmol/L P_5S_3 without detergent (*solid line*) and in the presence of 1 mmol/L sodium dodecylsulfate (SDS) (*dotted line*) or dodecyl maltoside (DM) (*dashed line*). ***C)*** 1 µmol/L phosphorylated P_5S_3 without detergent (*solid line*) as well as in the presence of 1 mmol/L SDS (*dotted line*). All experiments were done at pH 7.5.

Both at SDS concentrations above (34.7 mmol/L) and below (1 mmol/L) the critical micellar concentration [73-75], P_5S_3 (100 and 1 µmol/L, respectively) underwent strong structural transitions (Figure 3). In both environments, the CD-spectral features of polyproline II or disordered conformation were lost, and instead, a spectrum typical for α-helices appeared. At 34.7 mmol/L SDS and 100 µmol/L P_5S_3 (molar ratio 347 to 1), the P_5S_3 CD-spectrum was positive below 195 nm and exhibited two negative peaks around 205 and 220 nm, respectively (Figure 3A). This indicates an increase in α-helical conformation, which was supported by the secondary structure estimation by ContinII (Figure S7). At 1 mmol/L SDS and 1 µmol/L P_5S_3 (molar ratio 1,000 to 1), the spectrum was characterized by a positive ellipticity peak centered at 192 nm, and two negative peaks centered at 207 and 225 nm, respectively (Figures 3B, S8). The transition towards α-helix was far more pronounced than in Figure 3A, because of the higher molar ratio of SDS per P_5S_3, covering binding sites on the polypeptide more effectively. Additionally, SDS above the critical micellar concentration forms micelles with an aggregation number at around 60 at moderate salt concentrations as applied here [76,77], which may have further reduced the effective concentration. Above the critical micellar concentration of SDS, it is likely that P_5S_3 was bound to the micelle surface, whereas below the interaction should be primarily an interaction with individual SDS molecules [71,72]. The electrostatic interaction presumably reduces the repulsion of cationically charged side chains, allowing the adoption of the more compact α-helix conformation. This effect was demonstrated with synthetic peptide XAO (acetyl-$X_2A_7O_2$-amine, X: diaminobutyric acid, A: alanine, O: ornithine), which adopts the extended polyproline II conformation in water, but acquires α-helical structure upon interaction with SDS [78].

To exclude that this structural transition was mainly induced by the hydrophobic tail of SDS, the effect of dodecyl maltoside (DM) at similar conditions was examined (Figure 3B and S8). This detergent has the same hydrophobic tail as dodecylsulfate, but instead of a charged headgroup bears an uncharged, hydrophilic disaccharide moiety. No structural transition of P_5S_3 was observed in the presence of DM. From this it is concluded that the P_5S_3-SDS interaction, inducing an α-helical conformation, is mainly driven by electrostatic forces. As with P_5S_3, no structural changes were observed in XAO when dodecyl maltoside was added [78].

Since phosphorylated P_5S_3 behaved differently than the unmodified P_5S_3 in the presence of silicic acid (Figure 2), the conformation in the presence of SDS was compared as well (Figures 3C and S9). At 1 mmol/L SDS and 1 µmol/L phosphorylated P_5S_3, the polypeptide CD-spectrum showed a reduction in polyproline II or disordered conformation, with again a concomitant increase of α-helical conformation, seen from the increased negative ellipticity in the 207 to 235 nm range. However, the structural transition was far less pronounced than for the non-phosphorylated P_5S_3. As for the silica-interaction, this is most likely a result of reduced electrostatic interaction.

Overall, there are similarities in the structure of SDS- and silica-bound P_5S_3, probably beause of the contribution of electrostatic interactions in the binding.

The effect of sodium chloride on P_5S_3-structure in silica.

To further investigate the electrostatic contribution to the interaction between P_5S_3 and silica, the polypeptide behavior in the presence of sodium chloride and silica was examined. The strength of binding to a silica surface was assessed by using silica precipitated biomimetically by the aid of spermin [15], binding of P_5S_3 buffered with Tris-HCl pH 7.5 and eventually desorption with sodium chloride solutions at different concentrations, as well buffered to pH 7.5. The desorption assay (Figure 4) revealed that P_5S_3 was not retained on the silica-surface at sodium chloride concentrations of 600 mmol/L and above, and a little amount of P_5S_3 was already desorbed at 500 mmol/L.

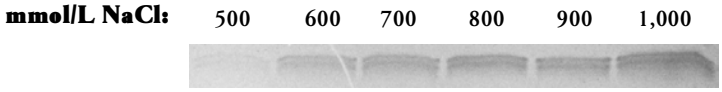

Figure 4. Desorption of P_5S_3 from silica by sodium chloride. Silica was produced by spermine-induced precipitation. P_5S_3 (15 µmol/L) was adsorbed onto the precipitate from Tris-HCl (50 mmol/L, pH 7.5) buffered solution. After washing with buffer, desorption was achieved with sodium chloride at different concentrations, noted above the gel picture. Eluted peptides were separated from the silica precipitate by centrifugation and the supernatants applied to PAGE. All solutions were buffered with 50 mmol/L Tris-HCl pH 7.5.

Silica formation in the absence and presence of P_5S_3 was examined at different sodium chloride concentrations. Silicic acid at 30 mmol/L and buffered with 50 mmol/L Tris-HCl pH 7.5 was incubated for 30 minutes with shaking at 1,400 rpm and subsequently silica was harvested by centrifugation (18,400 g). No precipitate was found in the absence of polypeptide and salt (Figure 5A). 300 and 600 mmol/L sodium chloride induced the precipitation of 50 and virtually 100 % of silicic acid available, respectively. In the presence of 100 µmol/L P_5S_3 50 % of the original silicic acid precipitated in the absence of salt. This yield was increased by 300 mmol/L sodium chloride but was not further enhanced by higher concentrations. At 300 mmol/L sodium chloride the presence of P_5S_3 led to an increased yield compared to the polypeptide-free reference, demonstrating that P_5S_3 remained active as biomineralizing agent.

The pellet and supernatant fractions after centrifugation were additionally analyzed for the presence of P_5S_3, and the polypeptide turned out to be fully co-precipitated irrespectively of the sodium chloride concentration (Figure 5B). Since P_5S_3 was already eluted from pre-formed silica at 600 mmol/L sodium chloride and above (Figure 4), retrieving P_5S_3 from the silica precipitate even at 1,000 mmol/L sodium chloride indicates that non-electrostatic interactions contribute to the P_5S_3-silica interaction. When the polypeptide becomes occluded in the mineral matrix during the silica-formation reaction, the interaction could be fostered by confinement. The non-electrostatic binding might be mediated by hydrophobic interactions, which were found for other silica-binding peptides [79,80] and polyelectrolytes [81,82].

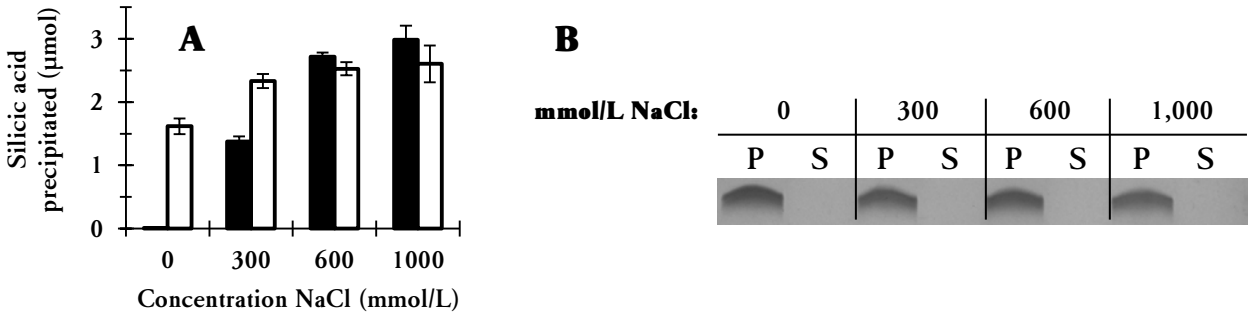

Figure 5. Impact of sodium chloride on P_5S_3-silica co-precipitation. Silica-formation in the absence or presence of 100 µmol/L P_5S_3 was examined with 30 mmol/L silicic acid solutions buffered with Tris-HCl (50 mmol/L, pH 7.5), at different concentrations of sodium chloride. After 30 minutes, precipitates were collected by centrifugation. ***A)*** Silicic acid in the pellet fraction was quantified by molybdenum blue analysis for assays in the absence (*black bars*) or presence (*white bars*) of P_5S_3. The total amount of silicic acid available in the assays was 3 µmol. ***B)*** Hydrolyzed supernatant (*S*) and pellet (*P*) fractions from the silica formation assay were analyzed for the presence of P_5S_3 by TCA-precipitation of peptides and subsequent SDS-PAGE.

CD-spectroscopy was used to analyze whether a P_5S_3-silica interaction could be proven at high sodium chloride concentrations. Silica-formation in the presence of P_5S_3 was examined at pH 9.5 where the spectral changes in the absence of salt were most pronounced (see Figure 1).

In solution, i.e. in the absence of silicic acid, the P_5S_3-spectrum was virtually unaffected in the presence of 150 mmol/L sodium chloride (Figure 6A). At 1,000 mmol/L sodium chloride,

ellipticity at 210 - 235 nm was notably decreased, which could indicate the adoption of a fraction of α-helix or β-sheet conformation [51]. The occurrence of an α-helical local fold would be reasonable at high salt concentrations where charge screening may reduce the side-chain repulsion of like-charged residues (arginine and lysine in P_5S_3), which was shown to reduce the extended conformation of intrinsically disordered proteins [83,84]. Below 200 nm, samples at high salt concentration were too turbid for CD measurement. Secondary structure estimation by ContinII was hence inappropriate, since it would have required spectral information in the region 190 to 200 nm.

By contrast, P_5S_3-spectra after silica-formation were strongly affected by sodium chloride (Figure 6B). In the absence of sodium chloride, the spectrum of P_5S_3 in silica very much resembled an α-helical spectrum (see also Figure 1), with positive ellipticity below 200 nm and two negative ellipticity peaks at 208 and 228 nm, respectively. At 150 mmol/L sodium chloride, negative ellipticity at 210 - 235 nm was decreased, and this wavelenth region then matched with the spectrum of P_5S_3 in the absence of both silicic acid and sodium chloride (Figure 6A). The spectrum at 150 mmol/L sodium chloride was additionally mostly negative at 190 - 200 nm, though not as negative as the reference in the absence of silicic acid, revealing that still a P_5S_3 silica interaction occured. After increasing the sodium chloride concentration to 1,000 mmol/L, the 210 - 235 nm region remained similar to the spectrum at 150 mmol/L. Around 200 nm, the mean residue ellipticity became more negative (the peak centered at 200 nm decreased from -7 to -9), though was still above the silicic acid-free reference (peak at 199 nm around -19). This clearly demonstrates that still at this high concentration of sodium chloride, the P_5S_3-silica interaction remained.

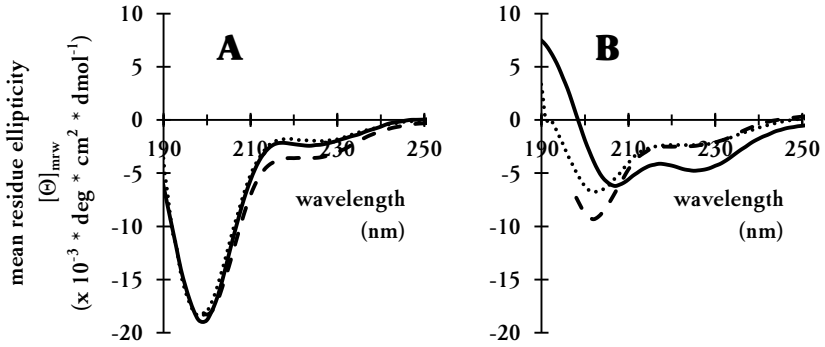

Figure 6. Changes in the P_5S_3 CD-spectrum in the presence of sodium chloride and silicic acid. Spectra were recorded without sodium chloride (*solid line*) or in the presence of 150 mmol/L (*dotted line*) and 1,000 mmol/L (*dashed line*) sodium chloride. The optical density of samples containing 1,000 mmol/L sodium chloride increased rapidly below 198 nm, so that the spectra are only shown to this wavelength. ***A)*** P_5S_3 spectra in buffer at pH 9.5. ***B)*** Spectra at pH 9.5 in the presence of 15 mmol/L silicic acid (silica-precipitation conditions).

Monitoring P_5S_3-induced silica-formation by time-resolved circular dichroism spectroscopy.

The changes in the CD-spectra of P_5S_3 upon silica formation were monitored for 30 minutes at pH 6.5, where silica autopolycondensation is slowed down in comparison with higher pH environments [41,85] but still P_5S_3 induces the formation of silica within 30 minutes ([34] and section III in this thesis). Ellipticity at 197 nm, where the major peak in the P_5S_3 spectrum was located, was recorded for the time-resolved analysis and spectra were compared before and after silica formation (Figure 7). Silicic acid was added either in its monomeric (acidic hydrolysate of TMOS) or pre-condensed (15 minutes at 60 mmol/L silicic acid and pH 9.5) form.

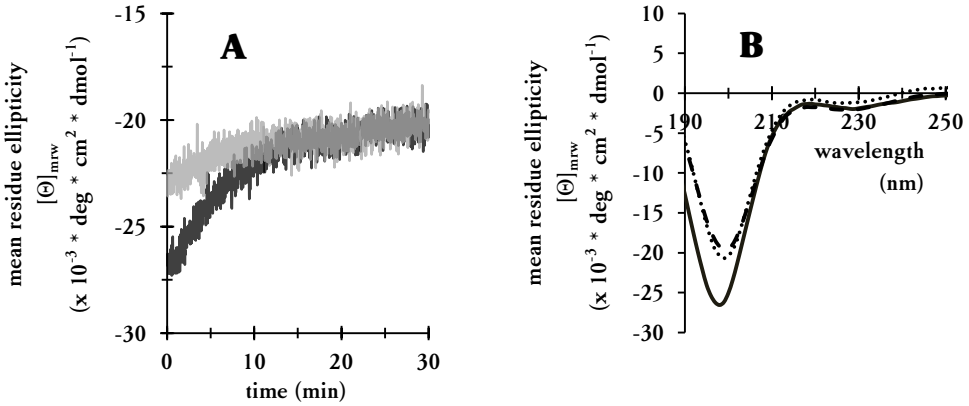

Figure 7. Time resolved changes in the CD-spectrum of P_5S_3 at pH 6.5, in response to interaction with silicic acid. Silicic acid was either added as monosilicic acid or after pre-condensation at 60 mmol/L and pH 9.5. The measurement was at pH 6.5 in the presence of 30 mmol/L silicic acid and 100 µmol/L P_5S_3. ***A)*** After addition of silicic acid (orthosilicic acid: *black line*, pre-condensed silicic acid: *grey line*) ellipticity at 197 nm was recorded over time. ***B)*** Spectra without silicic acid (*solid line*) and after the time-course started with orthosilicic acid (*dotted line*) or pre-condensed silica (*dashed line*).

When monosilicic acid was added, ellipticity at 197 nm started at a mean residue ellipticity around -27, the intensity similar to that in the solution spectrum, indicating that there was no immediate interaction between silicic acid and P_5S_3 leading to a structural transition (Figure 7A). However, with the progress of the measurement, negative ellipticity at 197 nm successively decreased. Within the first ten minutes, this decrease was rapid, subsequently slowed down, and eventually approached an ellipticity around -20. When pre-condensed silica was added, there was an instantaneous reduction of the peak-intensity at 197 nm (starting out around -23). Within the next 15 minutes, intensity was further reduced and reached again -20 as it was seen with monosilicic acid. The steady-state spectra recorded after the time-resolved measurement (Figure 7B) revealed a reduction in the polyproline II conformation upon interaction with silica. The spectra were very similar irrespective of whether monosilicic acid or pre-condensed silica was added. It is hence concluded that the structural changes seen in the CD-spectra were induced by the interaction of P_5S_3 with oligo- or polymeric silicic acid species, but not the monomeric form.

CD-analysis of P_5S_3 in globally undersaturated silicic acid solutions

In globally undersaturated solutions of silicic acid, P_5S_3 has the ability to retard the dissolution of pre-oligomerized silica (A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis). To gain further insight into the behavior of P_5S_3 under this condition, time-resolved CD-monitoring at 197 nm was applied. Silicic acid was allowed to autocondense for 15 minutes at pH 7.5 or 7.0, according to the pH of the subsequent dissolution measurement, and upon mixing with P_5S_3 dilution to undersaturated global silica-concentration (1 mmol/L) was achieved, which allowed the dissolution of silica.

Immediately after addition of silica to P_5S_3 (Figure 8A), the CD-intensity at 197 nm was reduced as expected and started out from -17.5 to -20. The intensity was further decreased to approach a minimum of -10 to -12.5 after around 150 minutes and subsequently increased to -22.5 to -25 after around 450 minutes. The CD spectra of P_5S_3 in the absence of silicic acid and after the silica-dissolution were virtually identical (Figure 8B), demonstrating that the changes in ellipticity at 197 nm were indicative of a P_5S_3-silica interaction.

The reason for the progressive increase within the first 150 minutes remains unclear, though different explanations are possible. A successive increase in bound silica, limited by diffusion, might enhance the change in secondary structure of P_5S_3. Alternatively, re-organization in the P_5S_3-silica adducts may have an impact on the P_5S_3 folding.

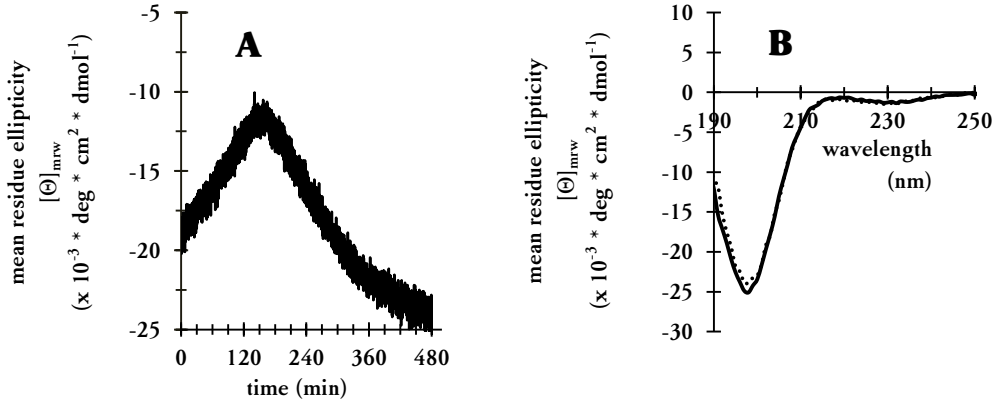

Figure 8. Time resolved changes in the CD-spectrum of P_5S_3 at pH 7.5 during dissolution of silica in undersaturated solution. Silicic acid was pre-condensed at 30 mmol/L and pH 7.5. The dissolution assays contained 1 mmol/L silicic acid (global concentration) and 1 µmol/L P_5S_3 in 0.5 mmol/L Tris-HCl buffer. ***A)*** Time-resolved ellipticity at 197 nm. The measurement was started immediately after addition of pre-condensed silica. ***B)*** Spectra without silicic acid (*solid line*) and after the time-course monitoring (*dotted line*).

Time-resolved CD-monitoring was additionally performed at pH 7.0 (Figure 9A), where silicic acid decondensation is slower than at pH 7.5 [41] and the retardation in the presence of P_5S_3 was less pronounced in molybdenum blue analysis (A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis). After addition of pre-condensed silicic acid, CD-intensity at 197 nm again started with an offset with respect to the spectrum of P_5S_3 in

the absence of silicic acid (Figure 9B) (-8 to -13 in time-resolved measurement versus -14 in the solution state spectrum). Similar to the measurement at pH 7.5, negative ellipticity at 197 nm decreased for 200 minutes, turned, and increased until the end of the measurement at 720 minutes. At this time point, still a deviation of the CD spectra in the presence and absence for silicic acid remained, indicating residual silica interacting with P_5S_3 which was not indicated from the molybdenum blue analysis where all silica was dissolved after 500 minutes (A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis), which might be explained by the higher buffer concentration (and hence higher ionic strength) in that analysis (20 mmol/L Tris-HCl versus 0.5 mmol/L Tris-HCl in the CD analysis).

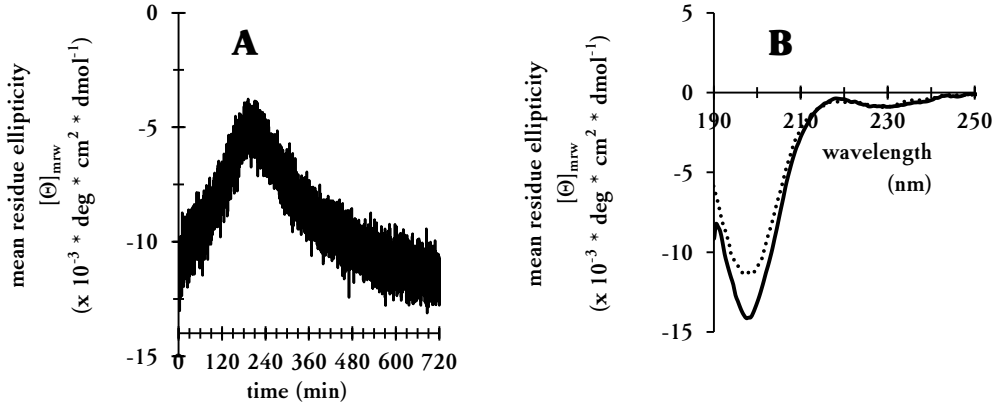

Figure 9. Time resolved changes in the CD-spectrum of P_5S_3 at pH 7.0 during dissolution of silica in undersaturated solution. Silicic acid was pre-condensed at 30 mmol/L and pH 7.0. The dissolution assays contained 1 mmol/L silicic acid (global concentration) and 1 µmol/L P_5S_3 in 0.5 mmol/L Tris-HCl buffer. ***A)*** Time-resolved ellipticity at 197 nm. The measurement was started immediately after addition of pre-condensed silica. ***B)*** Spectra without silicic acid (*solid line*) and after the time-course monitoring (*dotted line*).

Earlier, the retardation of silica-dissolution in the presence of P_5S_3 was proven by measuring the increase in molybdate reactive silica using the well-established molybdenum blue assay for silicic acid quantification (A. Spinthaki, C. Zerfaß, H. Paulsen, S. Hobe, K. Demadis, *in preparation* and section IV of this thesis). The CD-spectroscopic monitoring presented here allowed to analyze the polypeptide behavior during this process. The similarities in the changes of the CD-spectra during silica-formation and -dissolution revealed that in both cases P_5S_3 was bound to oligo- or polymeric silicic acid species and induced either their deposition or stabilized them transiently against dissolution in globally supersaturated or undersaturated silicic acid solution, respectively.

Conclusion

A circular dichroism spectroscopic analysis of the designed silica biomineralizing peptide P_5S_3 is presented. P_5S_3 adopts an extended conformation in solution at pH 4.5 to 9.5, both in an unmodified and phosphorylated state. Under conditions which allow silicic acid polymerization, the spectral features of the extended conformation are successively lost, a feature which can be

used as an indicator for P_5S_3-silica interaction. In the silica matrix, the P_5S_3 structure is more compact, presumably because of compensation of cationic side-chain charges by interaction with acidic and hence anionic silica, which in turn reduces side-chain repulsion. This assumption is supported by the similar effect of SDS above and below its critical micellar concentration. The binding of P_5S_3 to silica is weakened by sodium chloride. However, high salt concentrations do not prevent the polypeptide-silica co-precipitation, even when they suffice for surface desorption. By employing CD-spectroscopy, the interaction between P_5S_3 and silica at high salt conditions is revealed, pointing to the contribution of non-electrostatic forces in the binding. The CD-spectroscopy approach also allows tracing the behavior of P_5S_3 during the dissolution of silica in globally undersaturated silicic acid solutions, with high temporal resolution. The changes in the spectra indicate two different phases for P_5S_3 during silica dissolution, with an increasingly enhanced interaction in the first stage and an eventual release in the second stage. Judged from the CD-analysis, it was found that P_5S_3 interacts with silica in a similar manner both in silica-formation and -dissolution assays, suggesting that the different outcomes (silica-deposition or -dissolution) are a result of the environmental conditions, in particular the global concentration of silicic acid. The reason behind the spectral progression during silica dissolution is not yet elucidated and will be subject of future investigations.

Author contributions

All authors contributed to planning of the work, data analysis and writing of the manuscript.

Acknowledgement

C.Z. is a recipient of a fellowship through the Excellence Initiative (DFG/GSC 266) in the context of the graduate school of excellence "MAINZ" (Materials Science in Mainz).

References

[1] M. Sumper, N. Kröger, J. Mater. Chem. 14 (2004) 2059.

[2] N. Kröger, R. Deutzmann, M. Sumper, Science 286 (1999) 1129.

[3] N. Kröger, R. Deutzmann, M. Sumper, J. Biol. Chem. 276 (2001) 26066.

[4] N. Poulsen, N. Kröger, J. Biol. Chem. 279 (2004) 42993.

[5] M. Nemoto, Y. Maeda, M. Muto, M. Tanaka, T. Yoshino, S. Mayama, T. Tanaka, Mar. Genomics 16 (2014) 39.

[6] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14133.

[7] N. Kröger, M. Sumper, Protist 149 (1998) 213.

[8] N. Kröger, C. Bergsdorf, M. Sumper, EMBO J. 13 (1994) 4676.

[9] N. Kröger, C. Bergsdorf, M. Sumper, Eur. J. Biochem. 239 (1996) 259.

[10] N. Kröger, G. Lehmann, R. Rachel, M. Sumper, Eur. J. Biochem. 250 (1997) 99.

[11] N. Kröger, R. Wetherbee, Protist 151 (2000) 263.

[12] A. Scheffel, N. Poulsen, S. Shian, N. Kröger, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3175.

[13] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, Science 298 (2002) 584.

[14] M. Sumper, Science 295 (2002) 2430.

[15] D.J. Belton, S. V Patwardhan, C.C. Perry, J. Mater. Chem. 15 (2005) 4629.

[16] M.R. Knecht, D.W. Wright, Chem. Commun. (2003) 3038.

[17] A. Roehrich, G. Drobny, Acc. Chem. Res. 46 (2013) 2136.

[18] N. Li, X. Zhang, Q. Wang, F. Wang, P. Shen, RSC Adv. 2 (2012) 3288.

[19] M.M. Tomczak, D.D. Glawe, L.F. Drummy, C.G. Lawrence, M.O. Stone, C.C. Perry, D.J. Pochan, T.J. Deming, R.R. Naik, J. Am. Chem. Soc. 127 (2005) 12577.

[20] L. Kalmar, D. Homola, G. Varga, P. Tompa, Bone 51 (2012) 528.

[21] L. Senior, M.P. Crump, C. Williams, P.J. Booth, S. Mann, A. Periman, P. Curnow, J. Mater. Chem. B 3 (2015) 2607.

[22] D. Eliezer, Curr. Opin. Struct. Biol. 19 (2009) 23.

[23] A.H. Mao, S.L. Crick, A. Vitalis, C.L. Chicoine, R. V Pappu, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 8183.

[24] A.F. Drake, G. Siligardi, W.A. Gibbons, Biophys. Chem. 31 (1988) 143.

[25] A. V Mikhonin, N.S. Myshakina, S. V Bykov, S.A. Asher, J. Am. Chem. Soc. 127 (2005) 7712.

[26] L. Ma, Z. Ahmed, S.A. Asher, J. Phys. Chem. B 115 (2011) 4251.

[27] M.L. Tiffany, S. Krimm, Biopolymers 6 (1968) 1379.

[28] S. V Patwardhan, R. Maheshwari, N. Mukherjee, K.L. Kiick, S.J. Clarson, Biomacromolecules 7 (2006) 491.

[29] P.A. Mirau, J.L. Serres, M. Lyons, Chem. Mater. 20 (2008) 2218.

[30] K.M. Hawkins, S.S.-S. Wang, D.M. Ford, D.F. Shantz, J. Am. Chem. Soc. 126 (2004) 9112.

[31] H. Cui, V. Krikorian, J. Thompson, A.P. Nowak, T.J. Deming, D.J. Pochan, Macromolecules 38 (2005) 7371.

[32] L. Xia, Z. Li, Langmuir 27 (2011) 1116.

[33] E.G. Bellomo, T.J. Deming, J. Am. Chem. Soc. 128 (2006) 2276.

[34] C. Zerfaß, S. Braukmann, S. Nietzsche, S. Hobe, H. Paulsen, Protein Expr. Purif. 108 (2015) 1.

[35] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, Protein Sci. 4 (1995) 2411.

[36] A. Kuliopulos, Proc. Natl. Acad. Sci. 84 (1987) 8893.

[37] E. Gross, B. Witkop, J. Am. Chem. Soc. 83 (1961) 1510.

[38] N. Sreerama, R.W. Woody, J. Mol. Biol. 242 (1994) 497.

[39] N. Sreerama, R.W. Woody, Anal. Biochem. 287 (2000) 252.

[40] R.K. Iler, The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry, Wiley, New York, 1979.

[41] H. Baumann, Kolloid-Zeitschrift 162 (1959) 28.

[42] L.J. Henderson, Am. J. Physiol. 21 (1908) 173.

[43] H.N. Po, N.M. Senozan, J. Chem. Educ. 78 (2001) 1499.

[44] T. Coradin, D. Eglin, J. Livage, Spectroscopy 18 (2004) 567.

[45] L. Jiang, L. He, M. Fountoulakis, J. Chromatogr. A 1023 (2004) 317.

[46] H. Schaegger, G. von Jagow, Anal. Biochem. 166 (1987) 368.

[47] G. von Jagow, H. Schaegger, A Practical Guide to Membrane Protein Purification, Academic Press, San Diego, 1994.

[48] H. Schägger, Nat. Protoc. 1 (2006) 16.

[49] A.A. Adzhubei, M.J.E. Sternberg, A.A. Makarov, J. Mol. Biol. 425 (2013) 2100.

[50] N. Greenfield, G.D. Fasman, Biochemistry 8 (1969) 4108.

[51] N.J. Greenfield, Nat. Protoc. 1 (2006) 2876.

[52] M.L. Tiffany, S. Krimm, Biopolymers 11 (1972) 2309.

[53] A. Dos, V. Schimming, S. Tosoni, H.-H. Limbach, J. Phys. Chem. B 112 (2008) 15604.

[54] Y.P. Myer, Macromolecules 2 (1969) 624.

[55] D.J. Verbaro, D. Mathieu, S.E. Toal, H. Schwalbe, R. Schweitzer-Stenner, J. Phys. Chem. B 116 (2012) 8084.

[56] F. Eker, K. Griebenow, X. Cao, L.A. Nafie, R. Schweitzer-Stenner, Biochemistry 43 (2004) 613.

[57] J.M. Rifkind, Biopolymers 8 (1969) 685.

[58] A. Roque, I. Ponte, P. Suau, J. Phys. Chem. B 113 (2009) 12061.

[59] D.M. Eby, K. Artyushkova, A.K. Paravastu, G.R. Johnson, J. Mater. Chem. 22 (2012) 9875.

[60] K. Artyushkova, P. Atanassov, ChemPhysChem 14 (2013) 2071.

[61] A. Heredia, H.J. van der Strate, I. Delgadillo, V.A. Basiuk, E.G. Vrieling, ChemBioChem 9 (2008) 573.

[62] B.E. Kemp, D.J. Graves, E. Benjamini, E.G. Krebs, J. Biol. Chem. 252 (1977) 4888.

[63] K.-K. Lee, C. Joo, S. Yang, H. Han, M. Cho, J. Chem. Phys. 126 (2007) 235102.

[64] S.-Y. Kim, Y. Jung, G.-S. Hwang, H. Han, M. Cho, Proteins 79 (2011) 3155.

[65] A.A. Bielska, N.J. Zondlo, Biochemistry 45 (2006) 5527.

[66] D. Otzen, Biochim. Biophys. Acta 1814 (2011) 562.

[67] K.B. Chandrababu, K. Dutta, S.B. Lokappa, M. Ndao, J.S. Evans, J. Moradian-Oldak, Biopolymers 101 (2014) 525.

[68] T.S. Ulmer, A. Bax, N.B. Cole, R.L. Nussbaum, J. Biol. Chem. 280 (2005) 9595.

[69] A. Wahlström, L. Hugonin, A. Perálvarez-Marín, J. Jarvet, A. Gräslund, FEBS J. 275 (2008) 5117.

[70] J. Jarvet, J. Zdunek, P. Damberg, A. Gräslund, Biochemistry 36 (1997) 8153.

[71] N.J. Turro, X.-G. Lei, K.P. Ananthapadmanabhan, M. Aronson, Langmuir 11 (1995) 2525.

[72] B.F. Shaw, G.F. Schneider, H. Arthanari, M. Narovlyansky, D. Moustakas, A. Durazo, G. Wagner, G.M. Whitesides, J. Am. Chem. Soc. 133 (2011) 17681.

[73] M.L. Corrin, W.D. Harkins, J. Am. Chem. Soc. 69 (1947) 683.

[74] M.F. Emerson, A. Holtzer, J. Phys. Chem. 71 (1967) 1898.

[75] C. Thévenot, B. Grassl, G. Bastiat, W. Binana, Colloids Surfaces A Physicochem. Eng. Asp. 252 (2005) 105.

[76] S.S. Berr, R.R.M. Jones, Langmuir 4 (1988) 1247.

[77] E. Dutkiewicz, A. Jakubowska, Colloid Polym. Sci. 280 (2002) 1009.

[78] Z. Hong, K. Damodaran, S.A. Asher, J. Phys. Chem. B 118 (2014) 10565.

[79] V. Puddu, C.C. Perry, ACS Nano 6 (2012) 6356.

[80] S. V Patwardhan, F.S. Emami, R.J. Berry, S.E. Jones, R.R. Naik, O. Deschaume, H. Heinz, C.C. Perry, J. Am. Chem. Soc. 134 (2012) 6244.

[81] F. Xie, T. Nylander, L. Piculell, S. Utsel, L. Wagberg, T. Åkesson, J. Forsman, Langmuir 29 (2013) 12421.

[82] N.G. Hoogeveen, M.A.C. Stuart, G.J. Fleer, J. Colloid Interface Sci. 182 (1996) 133.

[83] A. Norouzy, K.I. Assaf, S. Zhang, M.H. Jacob, W.M. Nau, J. Phys. Chem. B 119 (2015) 33.

[84] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, B. Schuler, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14609.

[85] D.J. Belton, O. Deschaume, S.V. Patwardhan, C.C. Perry, J. Phys. Chem. B, 114 (2010) 9947.

Supplementary Information

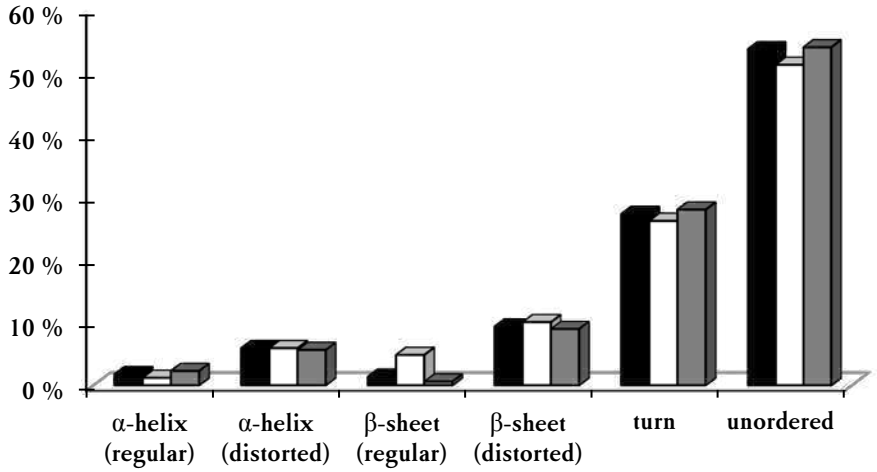

Figure S1. ContinII secondary structure estimation for P_5S_3 CD-spectra at pH 4.5 (see Figure 1). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 30 mmol/L (*white bar*) or 60 mmol/L (*grey bar*) silicic acid.

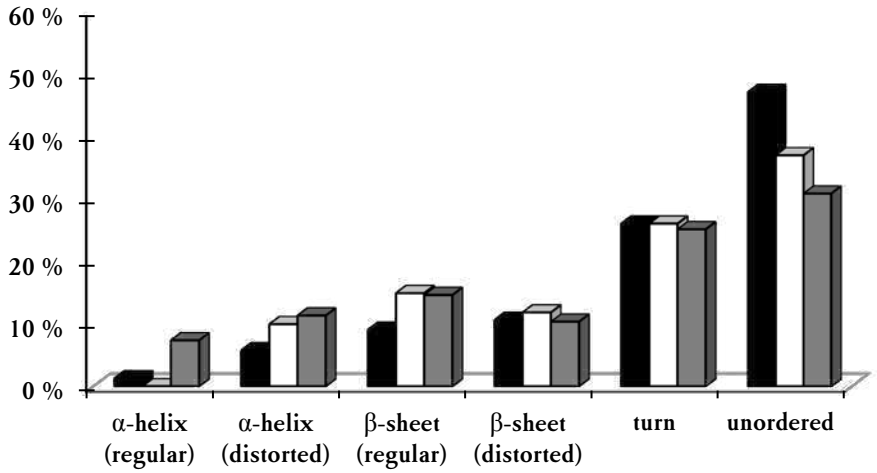

Figure S2. ContinII secondary structure estimation for P_5S_3 CD-spectra at pH 7.5 (see Figure 1). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 15 mmol/L (*white bar*) or 30 mmol/L (*grey bar*) silicic acid.

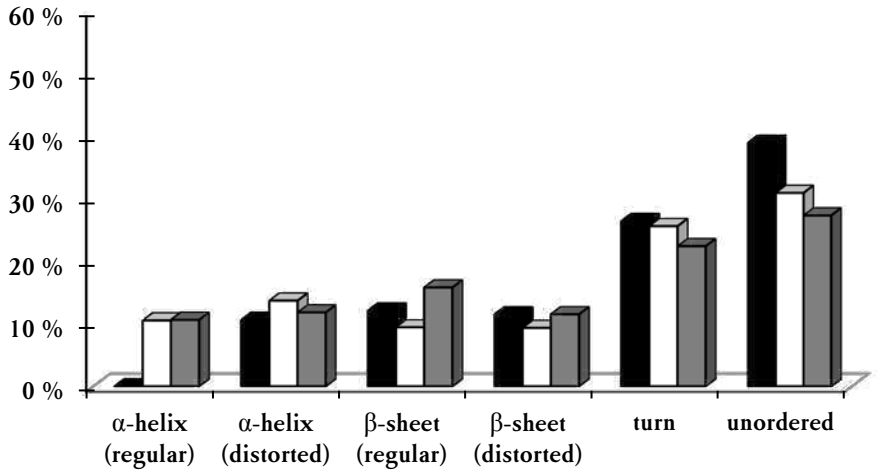

Figure S3. ContinII secondary structure estimation for P_5S_3 CD-spectra at pH 9.5 (see Figure 1). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 7.5 mmol/L (*white bar*) or 15 mmol/L (*grey bar*) silicic acid.

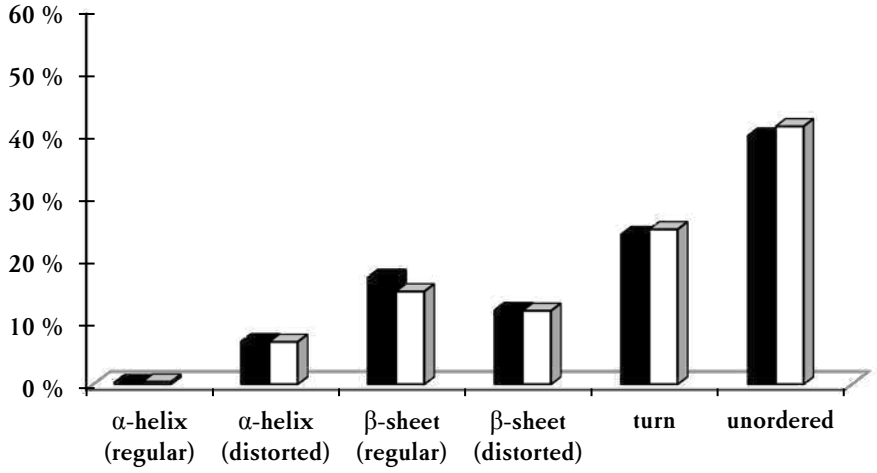

Figure S4. ContinII secondary structure estimation for CD-spectra of phosphorylated P_5S_3 at pH 4.5 (see Figure 2). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 30 mmol/L silicic acid (*white bar*).

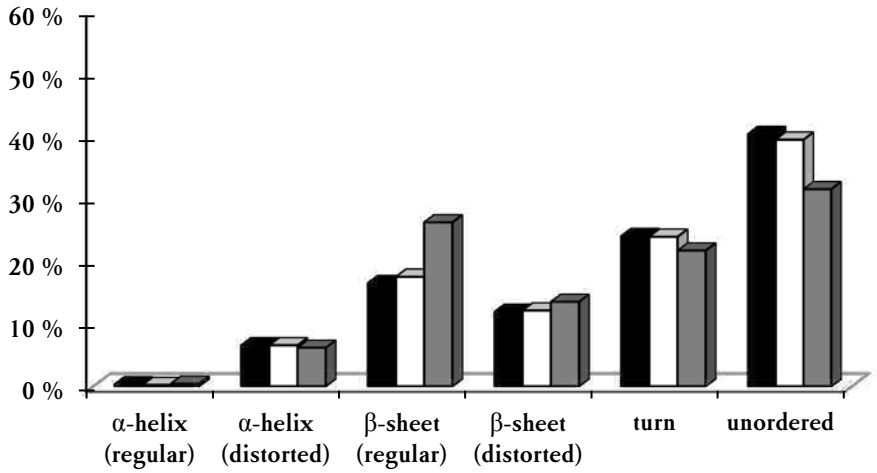

Figure S5. ContinII secondary structure estimation for CD-spectra of phosphorylated P_5S_3 at pH 7.5 (see Figure 2). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 15 mmol/L (*white bar*) or 30 mmol/L (*grey bar*) silicic acid.

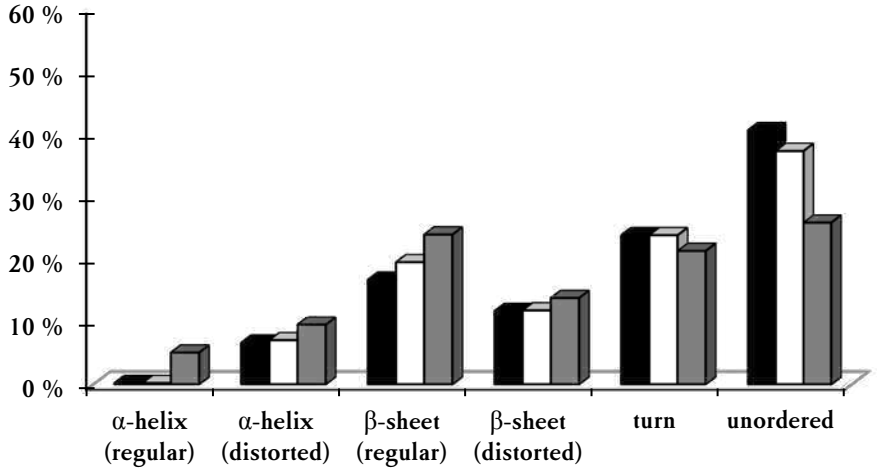

Figure S6. ContinII secondary structure estimation for CD-spectra of phosphorylated P_5S_3 at pH 9.5 (see Figure 2). Relative fractions of secondary structure motifs are presented for P_5S_3 in buffer (*black bar*) and in the presence of 7.5 mmol/L (*white bar*) or 15 mmol/L (*grey bar*) silicic acid.

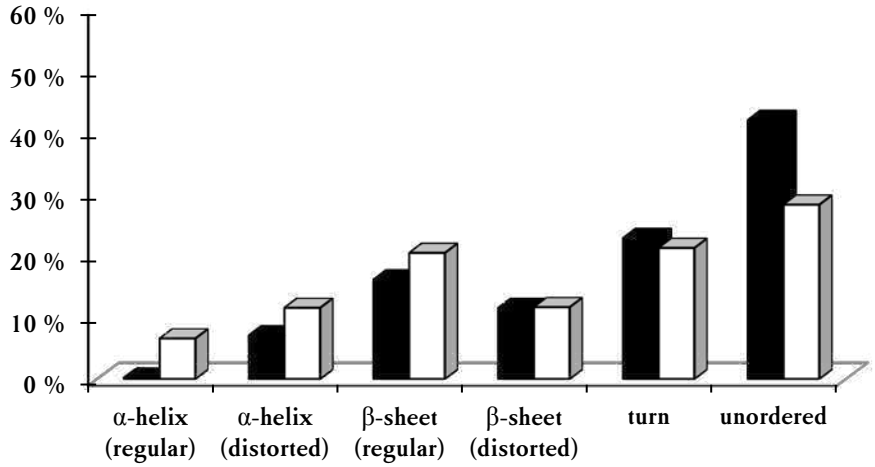

Figure S7. ContinII secondary structure estimation for P_5S_3 CD-spectra in presence of SDS above the critical micellar concentration (see Figure 3A). Relative fractions of secondary structure motifs are presented for 100 µmol/L P_5S_3 in the absence of detergent (*black bar*) and in the presence of 34.7 mmol/L (1 % w/v) SDS (*white bar*).

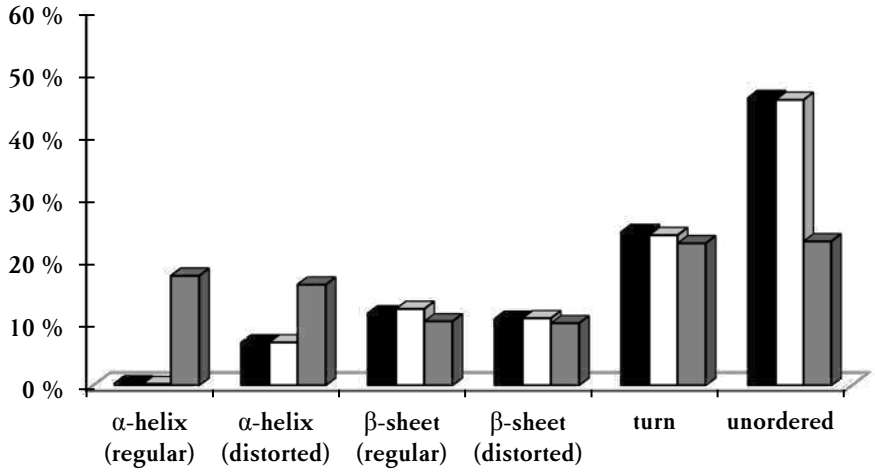

Figure S8. ContinII secondary structure estimation for P_5S_3 CD-spectra in the presence SDS or DM at 1 mmol/L (see Figure 3B). Relative fractions of secondary structure motifs are presented for 1 µmol/L P_5S_3 in the absence of detergent (*black bar*) and in the presence of 1 mmol/L dodecyl maltoside (DM) (*white bar*) or 1 mmol/L SDS (*grey bar*).

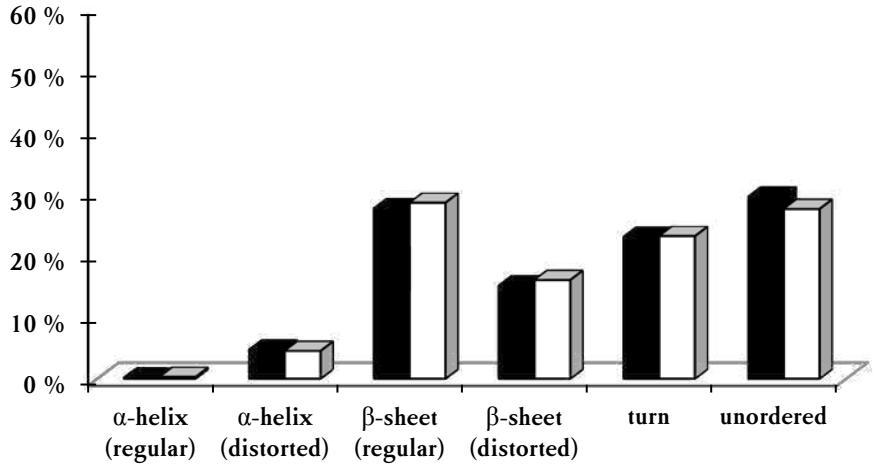

Figure S9. ContinII secondary structure estimation for CD-spectra of phosphorylated P_5S_3 in the presence of SDS at 1 mmol/L (see Figure 3C). Relative fractions of secondary structure motifs are presented for 1 µmol/L P_5S_3 in the absence of detergent (*black bar*) and in the presence of 1 mmol/L SDS (*white bar*).

List of Figures and Tables

Section I

Section II

Section III

Section IV

Section V

Appendix

Sequences of expressed constructs (A1 - A23), listed in Table 1 (Section I)

Sequence A1: SilL-(FXa)Sx3

Sequencing file: Xpau0056

Sequencing primer: T7PrompET22b(+)fw

>anionic leader from Sil1p

```
ATGGTGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCC  < 60
M  V  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A
         10         20         30         40         50
GCCTCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTG  < 120
A  S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L
         70         80         90         100        110
GAATCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCC  < 180
E  S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S
         130        140        150        160        170
                                   >FXa-Site      >Sx3
GAAGAGGAAGAACTTCGTATCCCCGAGAATTCCATCGAGGGTCGCTCCATGGGATCCAAG  < 240
E  E  E  E  L  R  I  P  E  N  S  I  E  G  R  S  M  G  S  K
         190        200        210        220        230
GGTTCCAAGCGTCGCATCTTGTCCTCCAAAAAATCCGGATCCTACTCGGGATCCAAGGGC  < 300
G  S  K  R  R  I  L  S  S  K  K  S  G  S  Y  S  G  S  K  G
         250        260        270        280        290
TCCAAGCGTCGCAACTTGTCCTCCAAGAAATCCGGATCCTACTCGGGATCCAAGGGTTCC  < 360
S  K  R  R  N  L  S  S  K  K  S  G  S  Y  S  G  S  K  G  S
         310        320        330        340        350
AAGCGTCGCATCTTGTCCGGGGGTCTCGCGAGCTCCAAGCGTCGCAACTTGTCCTCCAAG  < 420
K  R  R  I  L  S  G  G  L  A  S  S  K  R  R  N  L  S  S  K
         370        380        390        400        410
AAATCCGGATCCTACTCGGGATCCAAGGGTTCCAAGCGTCGCATCTTGTCCGGGGGTAAG  < 480
K  S  G  S  Y  S  G  S  K  G  S  K  R  R  I  L  S  G  G  K
         430        440        450        460        470
         >His6
CTTCTCGAGCACCACCACCACCACCACTGA  < 510
L  L  E  H  H  H  H  H  H  *
         490        500
```

Sequence A2: SilL-P_5S_3(-Y,H)

Sequencing file: Xpau0069

Sequencing primer: T7PrompET22b(+)fw

>anionic leader from SilIp

```
ATGGTGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCC  < 60
M  V  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A
         10         20         30         40         50
GCCTCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTG  < 120
A  S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L
         70         80         90         100        110
GAATCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCC  < 180
E  S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S
         130        140        150        160        170
                                        >HRV3C-Site
GAAGAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 240
E  E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A
         190        200        210        220        230
>P5S3(-Y,H)
ATGGGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA  < 300
M  G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         250        260        270        280        290
GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCT  < 360
G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A
         310        320        330        340        350
                                        >His6
AGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 414
S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370        380        390        400        410
```

Sequence A3: pelB-SilL-CysP_2S_0(-Y,H)

Sequencing file: Xpau0522

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                >CysP2S0(-Y,H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
                                                   >His6
GGATCCTGTCGACGTGCTTCCTTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCAC  < 360
G  S  C  R  R  A  S  L  G  K  L  R  R  A  S  L  E  H  H  H
         310       320       330       340       350
CACCACCACTGA  < 372
H  H  H  *
         370
```

Sequence A4: pelB-SilL-CysP_3S_1(-Y,H)

Sequencing file: Xpau0501

Sequencing primer: T7PrompET22b(+)fw

>pelB leader sequence

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site >$CysP_3S_1$(-Y,H)

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTGTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  C  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
```

>His_6

```
AAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 405
K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A5: pelB-SilL-CysP_5S_3(-Y,H)

Sequencing file: Xpau0512

Sequencing primer: T7PrompET22b(+)fw

>pelB leader sequence

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site >CysP_5S_3(-Y,H)

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTGTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  C  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 420
K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         370       380       390       400       410
```

>His_6

```
CTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA   < 471
L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         430       440       450       460       470
```

Sequence A6: pelB-SilL-P_2S_0(-Y,H)

Sequencing file: Xpau0541

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P2S0(-Y,H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
                                                   >His6
GGATCCTCTCGACGTGCTTCCTTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCAC  < 360
G  S  S  R  R  A  S  L  G  K  L  R  R  A  S  L  E  H  H  H
         310       320       330       340       350
CACCACCACTGA  < 372
H  H  H  *
         370
```

Sequence A7: pelB*(X)*-SilL-CysP_2S_0(-Y,H)

Sequencing file: Xpau0534

Sequencing primer: T7PrompET22b(+)fw

```
>(duplication & mutation)          >pelB(X) leader sequence
ATGGAATACCTGCTGCCGACCGGAGATATACATATGGAATACCTGCTGCCGACCGCTGCT  < 60
M  E  Y  L  L  P  T  G  D  I  H  M  E  Y  L  L  P  T  A  A
         10        20        30        40        50
                                          >anionic leader from Sil1p
GCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCGATGGCCAGCGACTCCTCGGATGACGCA  < 120
A  G  L  L  L  L  A  A  Q  P  A  M  A  S  D  S  S  D  D  A
         70        80        90        100       110
TCTGATTCCTCTGTCGAGTCTGTCGACGCCGCCTCCTCTGACGTCTCTGGTTCCTCTGTC  < 180
S  D  S  S  V  E  S  V  D  A  A  S  S  D  V  S  G  S  S  V
         130       140       150       160       170
GAATCTGTCGACGTCTCTGGTTCCTCTCTGGAATCCGTTGACGTCTCTGGTTCCTCTCTG  < 240
E  S  V  D  V  S  G  S  S  L  E  S  V  D  V  S  G  S  S  L
         190       200       210       220       230
GAGTCCGTCGACGACTCCAGTGAGGACTCCGAAGAGGAAGAACTTCGTATCCCCGAGAAT  < 300
E  S  V  D  D  S  S  E  D  S  E  E  E  E  L  R  I  P  E  N
         250       260       270       280       290
   >HRV3C-Site                    >CysP2S0(-Y,H)
TCCCTGGAAGTTCTGTTCCAGGGGCCCGGCCATGGGATCCTGTCGACGTGCTTCCTTAGGG  < 360
S  L  E  V  L  F  Q  G  P  A  M  G  S  C  R  R  A  S  L  G
         310       320       330       340       350
                              >His6
AAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 405
K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A8: pelB-SilL-CysP_2S_0(-Y)

Sequencing file: Xpau0535

Sequencing primer: T7PrompET22b(+)fw

>pelB leader sequence

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10         20         30         40         50
```

>anionic leader from Sil1p

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70         80         90         100        110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130        140        150        160        170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190        200        210        220        230
```

>HRV3C-Site ***>CysP$_2$S$_0$(-Y)***

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250        260        270        280        290
GGATCCTGTCGACGTGCTTCCTTAGGGAAGCTGCGTCGCGCTAGCCTCTAGCACCACCAC  < 360
G  S  C  R  R  A  S  L  G  K  L  R  R  A  S  L  *  H  H  H
         310        320        330        340        350
```

>one His-codon deleted unintentionally, but not in translated region

```
CACCACTGA  < 369
H  H  *
```

Sequence A9: pelB-SilL-P_3S_1(-Y,H)

Sequencing file: Xpau0504

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from Sil1p
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P3S1(-Y,H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
                          >His6
AAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 405
K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A10: pelB*(X)*-SilL-P_3S_1(-Y,H)

Sequencing file: Xpau0532

Sequencing primer: T7PrompET22b(+)fw

>pelB(X) leader sequence

```
ATGGAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  E  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site ***>P_3S_1(-Y,H)***

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
```

>His_6

```
AAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 405
K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A11: pelB-SilL-P_3S_1(-Y)

Sequencing file: Xpau0533

Sequencing primer: T7PrompET22b(+)fw

>pelB leader sequence

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site ***>P_3S_1(-Y)***

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAGCTGCGTCGCGCTAGCCTCTAGCACCACCACCACCACCACTGA  < 405
K  L  R  R  A  S  L  *  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A12: pelB-SilL-P_5S_3(-Y,H)

Sequencing file: Xpau0094

Sequencing primer: T7PrompET22b(+)fw

>pelB leader sequence

```
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site >P_5S_3(-Y,H)

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 420
K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         370       380       390       400       410
```

>His_6

```
CTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA   < 471
L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         430       440       450       460       470
```

Sequence A13: pelB*(X)*-SilL-P_5S_3(-Y,H)

Sequencing file: Xpau0516

Sequencing primer: T7PrompET22b(+)fw

>pelB(X) leader sequence

```
ATGGAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  E  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
```

>anionic leader from SilIp

```
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
```

>HRV3C-Site ***>P_5S_3(-Y,H)***

```
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 420
K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         370       380       390       400       410
```

>His_6

```
CTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA   < 471
L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         430       440       450       460       470
```

Sequence A14: pelB-SilL-P_5S_3(-Y)

Sequencing file: Xpau0519

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P5S3(-Y)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGG  < 360
G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L  G
         310       320       330       340       350
AAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 420
K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         370       380       390       400       410
CTAGGGAAGCTGCGTCGCGCTAGCCTCTAGCACCACCACCACCACCACTGA   < 471
L  G  K  L  R  R  A  S  L  *  H  H  H  H  H  H  *
         430       440       450       460       470
```

Sequence A15: pelB-SilL-P_2S_0(H)

Sequencing file: Xpau0542

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P2S0(H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
                                                      >His6
GGATCCTATTCTCGACGTGCTTCCTTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCAC  < 360
G  S  Y  S  R  R  A  S  L  G  K  L  R  R  A  S  L  E  H  H
         310       320       330       340       350
CACCACCACCACTGA  < 375
H  H  H  H  *
         370
```

Sequence A16: pelB-SilL-P_3S_1(H)

Sequencing file: Xpau0510

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P3S1(H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAAGCTGCGTCGCGCTAGCCTA  < 360
G  S  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         310       320       330       340       350
                           >His6
GGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 408
G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         370       380       390       400
```

Sequence A17: pelB-SilL-P_5S_3(H)

Sequencing file: Xpau0298

Sequencing primer: T7PrompET22b(+)fw

```
>pelB leader sequence
ATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCCAGCCGGCG  < 60
M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A
         10        20        30        40        50
   >anionic leader from SilIp
ATGGCCAGCGACTCCTCGGATGACGCATCTGATTCCTCTGTCGAGTCTGTCGACGCCGCC  < 120
M  A  S  D  S  S  D  D  A  S  D  S  S  V  E  S  V  D  A  A
         70        80        90        100       110
TCCTCTGACGTCTCTGGTTCCTCTGTCGAATCTGTCGACGTCTCTGGTTCCTCTCTGGAA  < 180
S  S  D  V  S  G  S  S  V  E  S  V  D  V  S  G  S  S  L  E
         130       140       150       160       170
TCCGTTGACGTCTCTGGTTCCTCTCTGGAGTCCGTCGACGACTCCAGTGAGGACTCCGAA  < 240
S  V  D  V  S  G  S  S  L  E  S  V  D  D  S  S  E  D  S  E
         190       200       210       220       230
                                >HRV3C-Site                 >P5S3(H)
GAGGAAGAACTTCGTATCCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCCATG  < 300
E  E  E  L  R  I  P  E  N  S  L  E  V  L  F  Q  G  P  A  M
         250       260       270       280       290
GGATCCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA  < 360
G  S  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         310       320       330       340       350
GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCT  < 420
G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A
         370       380       390       400       410
                                   >His6
AGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 474
S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         430       440       450       460       470
```

Sequence A18: KSI-P_3S_1(-Y,H)

Sequencing file: Xpau0545

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10         20         30         40         50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70         80         90         100        110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130        140        150        160        170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190        200        210        220        230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250        260        270        280        290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310        320        330        340        350
                                        >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370        380        390        400        410
   >P3S1(-Y,H)
ATGGGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAAGCTGCGTCGCGCTAGCCTA  < 480
M  G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         430        440        450        460        470
                                  >His6
GGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 528
G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         490        500        510        520
```

Sequence A19: KSI-P_5S_3(-Y,H)

Sequencing file: Xpau0537

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10         20         30         40         50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70         80         90         100        110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130        140        150        160        170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190        200        210        220        230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250        260        270        280        290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310        320        330        340        350
                                        >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370        380        390        400        410
   >P5S3(-Y,H)
ATGGGATCCTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTA  < 480
M  G  S  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S  L
         430        440        450        460        470
GGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGCGCT  < 540
G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R  A
         490        500        510        520        530
                                        >His6
AGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 594
S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         550        560        570        580        590
```

Sequence A20: KSI-P_3S_1(H)

Sequencing file: Xpau0547

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10         20         30         40         50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70         80         90         100        110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130        140        150        160        170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190        200        210        220        230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250        260        270        280        290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310        320        330        340        350
                                        >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370        380        390        400        410
   >P3S1(H)
ATGGGATCCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 480
M  G  S  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         430        440        450        460        470
                                     >His6
CTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA  < 531
L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         490        500        510        520        530
```

Sequence A21: KSI-P_5S_3(H)

Sequencing file: Xpau0548

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10        20        30        40        50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70        80        90        100       110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130       140       150       160       170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190       200       210       220       230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250       260       270       280       290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310       320       330       340       350
                                        >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370       380       390       400       410
   >P5S3(H)
ATGGGATCCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 480
M  G  S  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         430       440       450       460       470
CTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGC  < 540
L  G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R
         490       500       510       520       530
                                        >His6
GCTAGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA   < 597
A  S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         550       560       570       580       590
```

Sequence A22: KSI-CysP_5S_3(H)

Sequencing file: Xpau0648

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10        20        30        40        50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70        80        90        100       110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130       140       150       160       170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190       200       210       220       230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250       260       270       280       290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310       320       330       340       350
                                    >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370       380       390       400       410
   >CysP5S3(H)
ATGGGATGCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGC  < 480
M  G  C  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         430       440       450       460       470
CTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGC  < 540
L  G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R
         490       500       510       520       530
                                          >His6
GCTAGCCTAGGGAAGCTGCGTCGCGCTAGCCTCGAGCACCACCACCACCACCACTGA   < 597
A  S  L  G  K  L  R  R  A  S  L  E  H  H  H  H  H  H  *
         550       560       570       580       590
```

Sequence A23: KSI-P_5S_3

Sequencing file: Xpau0578

Sequencing primer: pET22b(+)rev

>KSI

```
ATGCATACCCCAGAACACATCACCGCCGTGGTACAGCGCTTTGTGGCTGCGCTCAATGCC  < 60
M  H  T  P  E  H  I  T  A  V  V  Q  R  F  V  A  A  L  N  A
         10        20        30        40        50
GGCGATCTGGACGGCATCGTCGCGCTGTTTGCCGATGACGCCACGGTGGAAGACCCCGTG  < 120
G  D  L  D  G  I  V  A  L  F  A  D  D  A  T  V  E  D  P  V
         70        80        90        100       110
GGTTCCGAGCCCAGGTCCGGTACGGCTGCGATTCGTGAGTTTTACGCCAACTCGCTCAAA  < 180
G  S  E  P  R  S  G  T  A  A  I  R  E  F  Y  A  N  S  L  K
         130       140       150       160       170
CTGCCTTTGGCGGTGGAGCTGACGCAGGAGGTACGCGCGGTCGCCAACGAAGCGGCCTTC  < 240
L  P  L  A  V  E  L  T  Q  E  V  R  A  V  A  N  E  A  A  F
         190       200       210       220       230
GCTTTCACCGTCAGCTTCGAGTATCAGGGCCGCAAGACCGTAGTTGCGCCCATCGATCAC  < 300
A  F  T  V  S  F  E  Y  Q  G  R  K  T  V  V  A  P  I  D  H
         250       260       270       280       290
TTTCGCTTCAATGGCGCCGGCAAGGTGGTGAGCATCCGCGCCTTGTTTGGCGAGAAGAAT  < 360
F  R  F  N  G  A  G  K  V  V  S  I  R  A  L  F  G  E  K  N
         310       320       330       340       350
                                    >HRV3C-Site
ATTCACGCATGCCAGATGCTGCCCGAGAATTCCCTGGAAGTTCTGTTCCAGGGGCCGGCC  < 420
I  H  A  C  Q  M  L  P  E  N  S  L  E  V  L  F  Q  G  P  A
         370       380       390       400       410
```

>P_5S_3

```
ATGGGATCCTATTCTCGACGTGCTTCCTTAGGGAAATCTAAAAAAGCTGCGTCGCGCTAGC  < 480
M  G  S  Y  S  R  R  A  S  L  G  K  S  K  K  L  R  R  A  S
         430       440       450       460       470
CTAGGGAAATCTAAAAAGCTGCGTCGCGCTAGCCTAGGGAAATCTAAAAAGCTGCGTCGC  < 540
L  G  K  S  K  K  L  R  R  A  S  L  G  K  S  K  K  L  R  R
         490       500       510       520       530
GCTAGCCTAGGGAAGCTGCGTCGCGCTAGCCTCTAGCACCACCACCACCACCACTGA   < 597
A  S  L  G  K  L  R  R  A  S  L  *  H  H  H  H  H  H  *
         550       560       570       580       590
```

Concluding Declaration

Ich erkläre, dass ich die vorgelegte Thesis selbständig, ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt habe, die ich in der Thesis angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der Satzung der Johannes Gutenberg-Universität Mainz zur Sicherung guter wissenschaftlicher Praxis niedergelegt sind, eingehalten.

Mainz, 27. April 2015

_____

Christian Zerfaß

Acknowledgements

I wold like to thank my advisors, [redacted] for planning this thesis with me and guiding me through the PhD research. I very much appreciate the valuable discussions about the laboratory work, and the support in e.g. planning the visit of conferences or working in collaborations. I would also like to thank [redacted] to evaluate my thesis as third referee.

Alike, I would like to thank all people who joined and supported this research within collaborations. I thank [redacted] (University of Crete) for evaluating the ability of my polyeptides to stabilize soluble silicic acid, [redacted] (Pacific Northwestern Laboratories) for their efforts to resolve the polypeptides in silica by NMR and being my host during my USA visit in 2014, [redacted] (Johannes Gutenberg University Mainz) for visualizing the polypeptide assemblies in TEM, [redacted] (University Hospital Jena) for visualizing silica-polypeptide co-precipitates by SEM, and [redacted] for advice on DLS and ITC and the possibility to conduct measurements in her laboratory. I would like to acknowledge the interesting discussions with all partners and the possibility to present my research to foreign groups and to receive their precious feedback on the progress of my research.

The **Graduate School Materials Science in Mainz** (MAINZ) is acknowledged for supporting me by a stipend and setting a framework for the PhD time, supplemented by the organization of transferable skill courses and the offer to receive informal advice on any occasion.

I wish to thank my mentor [redacted], who engaged in counselling me in the framework of MAINZ, for helpful advice in terms of carreer planning and as well being interested in my research and discussing the progress with me.

My deepest gratitude is devoted to **all my family**, in particular my parents, brothers and grandparents, for supporting me through all my life, making especially the work on my PhD-thesis an enjoyable time knowing to be bolstered and secured in all occasions. I sincerely appreciate your patientness with me and for being there when something went wrong. I would like to extend this appreciation to **all my friends** inside and outside the university.

Within a several year PhD study, many colleagues and fellows come and go, but all leave a lasting memory. To acknowledge my colleagues, I would first like to thank three ladies who joined in my research for doing their Bachelor theses, which are [redacted]. Their excitement for and engagement in doing research gave me precious results in topics associated with my own thesis. I would like to thank [redacted] for many conversations on biological topics far away from my thesis, providing me interesting literature on these, and sharing his expertise on DNA modification with me; [redacted], somehow my predecessor as PhD student in the biomineralization topic, for valuable discussions

both on research, and on political issues; my longest term PhD fellow students [redacted] for working and discussing together, for their open-mindedness, for discussions on literally all topics occurring along the way, for mutual supportiveness during the work, and (without detailed breakdown) for reminding me that life is not serious anytime (#yolo); [redacted] for interesting discussions, support and advice; [redacted] for all support in and around the laboratory as well as interesting conversations at the institute. I would furthermore like to acknowledge all the agreeable students who came around during my time at the institute for theses and practicals and made the time enjoyable, and I apologize to not acknowledging everyone in particular, though they will be remembered.

I would like to close with expressing my appreciation to all of those I met on the way occasionally and who actively shared their experiences with me on how to succeed through life. Although some of these friends may be only met once or occasionally, I will keep their precious advices in mind, and even though I may not be able to thank them in particular, I want to acknowledge their open-minded attitude and supportiveness.

Curriculum Vitae

Christian Zerfaß

[redacted]

christian.zerfass@yahoo.com

Personal information

- Date and place of birth: 7th March 1988, in Bad Kreuznach (Germany)
- Nationality: German

Education

11/2011 - Present

PhD-studies on biomineralization
Supervisor: [redacted]

10/2010 - 11/2011

Master studies in preparation for Fast Track PhD-program (not concluded with MSc. diploma in favor of earlier PhD)

2009 - 2011

Recurrent project-work at Max Planck Institute for Polymer Research, Mainz, working group [redacted]
(Current address: [redacted])

04/2008 - 10/2010

Bachelor studies "Molecular Biology"
Concluded with BSc diploma (with [redacted], now PhD-advisor, please see address above)
Overall Grade: Very Good

Thesis: "Characterization of synthetic, silaffin-related peptides and their

silica-precipitation activity"

- Peptide synthesis through Merrifield approach
- Product purification (HPLC) and analysis (NMR, mass spectrometry)
- Silica precipitation assays and colorimetric silica quantification

1998 - 03/2007 — Lina Hilger Gymnasium Bad Kreuznach
accomplished with "Abitur" (higher education entrance qualification)
final grade: 1.8

Fellowships

10/2011 - Present — Stipend of the Graduate School of Excellence "Materials Science in Mainz" for PhD-studies (www.mainz.uni-mainz.de)

02/2011 - 09/2011 — Stipend of the "Studienstiftung des Deutschen Volkes" (German National Academic Foundation, www.studienstiftung.de/en/)

Research experience

PhD-studies, Working Group of [redacted]

- Recombinant expression of highly cationic peptides in *Escherichia coli*
- Analyses of silicic acid precipitation activity
- Establishment of Circular Dichroism spectroscopy based approach to study the peptide-silica interaction dynamically

Undergraduate research assistance, Working Group of [redacted]

- Cell-free expression systems for production of seven transmembrane proteins (e.g. Odorant Receptor R5)
- Studies on oriented insertion into small unilamellar vesicles
- Radioactive protein labelling for gel electrophoretic analysis and radio blotting

Teaching experience

Co-Supervision of three BSc candidates

Supervision of two students in advanced practical course

University services

2011 - 2014	Member of the Gutenberg Teaching Council (GTC) (www.glk.uni-mainz.de) Including representation during the university's application within the Excellence Initiative of the German Research Foundation (DFG)
2008 - 2011	Member of the Student Representative Council in Biology Including representation in the faculty's Committee on Budgets as well as the university's central Student Representative Council

Technical practice

Broad experiences	• PCR-amplification of DNA • Cloning techniques (plasmid-based genetic modification of bacteria) • Recombinant protein expression in bacterial host • Protein fusions in expression of demanding sequences • Protein purification approaches • Mineral precipitation assays and colorimetric silica quantification • *In vitro* protein expression systems • Radioactive labeling of *in vitro* synthesized proteins • Gel electrophoresis and staining/blotting *via* coomassie, Western blotting and radio blotting • Circular Dichroism spectroscopy
Basic experiences	• Eukaryotic cell cultures within gel matrices • Quantitative Real-Time PCR • High Pressure Liquid Chromatography • Chemical peptide synthesis (Merrifield)
Rudimentary experiences	• Atomic Force Microscopy • Electron Microscopy • Surface Plasmon Resonance • Nuclear Magnetic Resonance Spectroscopy • Dynamic Light Scattering

Memberships in scientific societies

Member of the Society of German Chemists (Gesellschaft deutscher Chemiker, GDCh)

Member of the Royal Society of Chemistry (RSC)

Publications

1.) Sandra Ritz, Michael Hulko, Christian Zerfaß, Sylvia May, Ingeborg Hospach, Nadejda Krasteva, Gabriele Nelles, Eva-Kathrin Sinner
„Cell-free expression of a mammalian olfactory receptor and unidirectional insertion into small unilamellar vesicles (SUVs)"
Biochimie **2013**, 95 (10), 1909 - 1916

2.) Christian Zerfaß, Sandra Braukmann, Sandor Nietzsche, Harald Paulsen, Stephan Hobe
„High yield recombinant production of a self-assembling polycationic peptide for silica biomineralization"
Protein Expression and Purification **2015**, 108, 1-8

Conference Participation

1. 11th International Conference on Materials Chemistry (MC11)
Hosted by the Royal Society of Chemistry (RSC)
8th - 11th of July 2013, at University of Warwick (United Kingdom)
Oral Presentation: "Engineering Silaffins: Recombinant, self-assembling peptides with high cationic charge density for silica biomineralization"

2. 12th International Symposium in Biomineralization (Biomin12)
27th - 30th of August 2013, in Freiberg (Germany)
Oral Presentation "Recombinant Silaffin-related Peptides: Expression, Assembly and Silica Precipitation"

3. ISACS 15 Challenges in Nanoscience
Hosted by the Royal Society of Chemistry (RSC)
17th - 20th of August 2014, in San Diego (CA, United States of America)
Poster Presentation "Probing a biomineralizing peptide with circular dichroism spectroscopy"