

DISSERTATION ZUR ERLANGUNG DES AKADEMISCHEN GRADES
DOCTOR RERUM NATURALIUM TECHNICARUM
(**DR.NAT.TECHN.**)
AN DER UNIVERSITÄT FÜR BODENKULTUR WIEN

**THE STRUCTURE AND ASSEMBLY OF SURFACE
LAYER PROTEINS: A COMBINED APPROACH OF
IN SILICO AND EXPERIMENTAL METHODS**

VON

DI CHRISTINE-MARIA HOREJS

UNTER DER BETREUUNG VON

AO.UNIV.PROF.DR.DIETMAR PUM UND DR.RUPERT TSCHELIESSNIG

DURCHGEFÜHRT AM

DEPARTMENT FÜR NANOBIOTECHNOLOGIE

Wien, Juni 2011

Meiner Familie

Things are only impossible until they're not.
(Jean-Luc Picard)

Danksagung

Ich möchte mich zuallererst bei meinen beiden Doktorvätern Dietmar Pum und Rupert Tscheliessnig bedanken. Dietmar Pum danke ich für seine stetige Unterstützung, seine Rückendeckung in so manch schwieriger Situation, für all die intellektuellen und praktischen Freiheiten, die er mir während meiner Zeit als Doktorandin gegeben hat und für die Möglichkeit, auf internationalen Konferenzen meine Arbeiten zu präsentieren und einen Forschungsaufenthalt in den USA zu absolvieren. Ich danke Rupert Tscheliessnig, von dem ich so vieles gelernt habe, für die Unterstützung, den intellektuellen Input und für all die hilfreichen Diskussionen, die das Entstehen meiner Dissertation überhaupt erst ermöglicht haben. Ich danke auch unserem emeritierten Departmentleiter und geistigem Vater der S-Schicht-Forschung Uwe Sleytr für die zahlreichen Gespräche, die immer wieder neue Aspekte aufgebracht haben und vor allem auch für seine Hilfe und Unterstützung, auf die ich immer zählen konnte.

Murugappan Muthukumar (Muthu) danke ich für die herzliche Aufnahme am *Institute for Polymer Science and Engineering, University of Massachusetts, Amherst, USA*, an dem ich insgesamt 6 Monate verbracht habe. Danke für diese produktive Zusammenarbeit und die vielen Diskussionen, die meine Arbeit wesentlich bereichert haben. Ich möchte auch Mithun Mitra danken für die wirklich tolle Zusammenarbeit und all die C++ Codes, die den Proteinen neues Leben eingehaucht haben.

Herwig Peterlik vom Institut für Dynamik kondensierter Systeme, Fakultät für Physik an der Universität Wien, danke ich für die Möglichkeit die SAXS Messungen bei ihm durchführen zu können und für seine Unterstützung und Hilfe bei den Messungen und der Auswertung der Ergebnisse.

Ich danke der Österreichischen Akademie der Wissenschaften für die Zuerkennung eines DOC-fORTE Stipendiums, das die Finanzierung meines Doktorats und meine Auslandsaufenthalte ermöglicht hat.

Ich bedanke mich bei all meinen Kolleginnen und Kollegen am Department für Nanobiotechnologie, besonders bei Jacqueline Friedmann für die Starthilfe beim AFM, die Unterstützung im Labor und die Hilfe bei so vielen technischen Problemen, bei Andrea Scheberl für ihr umfassendes biochemisches Wissen und ihre Hilfe beim Reinigen der Proteine, bei Birgit Kainz für die vielen Gespräche und Anregungen, bei Robin Ristl für unsere tolle Zusammenarbeit und seine statistischen Fähigkeiten, und bei Dieter Jäger für die Unterstützung bei all meinen finanziellen und bürokratischen Angelegenheiten und dafür, dass er immer ein offenes Ohr für alle kleineren und größeren Probleme hat.

Ich möchte mich bei meiner ewigen Freundin Loni bedanken, die immer auf meiner Seite war und durch so manche alternative Ansicht meine Zeit als Studentin und Doktorandin bereichert hat. Meinem lieben Freund Laurent danke ich für seine stetige Unterstützung und die vielen interessanten und anregenden Diskussionen. Meiner lieben Freundin Alexandra danke ich für ihren ewigen Optimismus und die vielen tollen Gespräche. Und schließlich bedanke ich mich bei meinem *Circle of Trust* für diese wirklich einmalige Zeit.

Ich danke Sergi dafür, dass er mir immer den Rücken gestärkt und mich unterstützt hat. Moltes gràcies!

Ich danke meinen Eltern Ute und Harald, meinen Schwestern Barbara und Angelika und meiner Großmutter Maria für die mentale sowie finanzielle Unterstützung über all die Jahre und für ihre vielen Ratschläge, die mir geholfen haben, so manches "unlösbares" Problem zu lösen.

Abstract

Self-assembly of matter is one of nature's most sophisticated strategies to organize molecules on a large scale and to create order from disorder. Surface (S-)layer proteins self-assemble in a highly reproducible and robust fashion in order to form crystalline layers that completely cover and protect prokaryotic cells. Long conserved during evolution, S-layers constitute a unique model system to study the molecular mechanisms of functional self-assembly, while additionally, they provide a basic matrix for the specific construction of ordered nanostructures. Due to their intrinsic capabilities to self-assemble into two-dimensional crystals, the elucidation of the three-dimensional structure of single S-layer proteins demands an approach beyond conventional structure determination methods. In this work, computer simulations were combined with experimental techniques in order to study the structure and intra- and intermolecular potentials guiding the proteins to self-assemble into lattices with different symmetries. Molecular dynamics, Monte Carlo methods, small-angle X-ray scattering involving a new theoretical description, and AFM-based single-molecule force spectroscopy yield new insights into the three-dimensional structure of S-layer proteins, the location, type and distribution of amino acids in S-layer lattices, the molecular mechanisms behind the self-assembly process, the mechanical stability and adaptive structural conformations that S-layer proteins are able to establish. *In silico* studies - embedded in an adequate experimental and theoretical scaffold - offer the possibility to calculate structural and thermodynamic features of proteins, while this work demonstrates the growing impact of such theoretical techniques in the fascinating field of biophysics at the nano-scale.

Kurzfassung

Selbstorganisation ist eine der raffiniertesten Strategien der Natur, um Moleküle in großem Maßstab zu organisieren und mit wenig Aufwand aus einem ungeordneten ein strukturiertes molekulares System zu schaffen. S-Schicht Proteine organisieren sich selbst zu kristallinen zweidimensionalen Schichten auf der Oberfläche prokaryotischer Zellen, wobei sie die Zelle vollkommen bedecken und sowohl Schutz als auch eine spezifische Barriere zur Außenwelt bilden. Damit stellen S-Schichten ein einzigartiges Modellsystem zur Untersuchung der molekularen Mechanismen dar, die der Selbstorganisation zu Grunde liegen. Die Konstruktion definierter Nanostrukturen basiert in erster Linie auf den Eigenschaften solcher Moleküle, wobei S-Schichten eine hervorragende selbst organisierte Matrix bilden. Diese Eigenschaften erschweren allerdings auch die Aufklärung der dreidimensionalen Struktur einzelner S-Schicht Proteine, weil klassische experimentelle Methoden kaum eingesetzt werden können. In dieser Arbeit werden theoretische und experimentelle Methoden kombiniert, um die Struktur sowie die intra- und intermolekularen Potentiale, die der Selbstorganisation von S-Schichten zu Grunde liegen, zu untersuchen. Molekulardynamische Simulationen, Monte Carlo Methoden, Kleinwinkelröntgenstreuung zusammen mit neuen theoretischen Auswerteverfahren und Einzelmolekül-Kraftspektroskopie ermöglichen neue Einblicke in die dreidimensionale Struktur von S-Schicht Proteinen, die Lage und Verteilung von Aminosäuren im S-Schicht Gitter, die molekularen Mechanismen der Selbstorganisation, die mechanische Stabilität und in die unterschiedlichen Konformationen, die S-Schicht Proteine je nach Bindungspartner einnehmen können. *In silico* Methoden - basierend auf entsprechenden experimentellen und theoretischen Ansätzen - ermöglichen die Berechnung dreidimensionaler Strukturen sowie thermodynamischer Eigenschaften von Proteinen. Diese Arbeit demonstriert die steigende Relevanz solch theoretischer Methoden in der Biophysik.

Contents

Abstract	V
Zusammenfassung	VI
1 Introduction	1
1.1 Surface layers: An outstanding self-assembly system	2
1.1.1 The challenge of determining the atomistic structure of S-layers . .	3
1.1.2 The S-layer proteins SbsB and SbpA	4
1.2 Computer simulations: Beyond experimental structure determination methods	5
1.2.1 The "other" protein folding problem	5
1.2.2 The protein structure prediction process	6
1.2.3 Molecular dynamics	7
1.2.3.1 Theoretical background: Newton and motion	7
1.2.3.2 Interaction potentials and force fields	8
1.2.3.3 Free energy calculation	9
1.2.3.4 Limitations and challenges	12
1.2.4 Langevin and Brownian dynamics	13
1.2.5 Monte Carlo simulations	14
1.2.6 Coarse-graining and statistical potentials	14
1.3 Structure prediction of an S-layer protein: A hybrid approach	16
1.4 Small-angle X-ray scattering	18
1.4.1 Theoretical basics of X-ray scattering	18
1.4.2 Reconstruction of low-resolution structures	23
1.4.3 Instrumentation	26
1.5 Self-assembly in a fractal space: Toward an understanding of S-layers in solution	28
1.5.1 The fractal geometry	28
1.5.2 Scattering of S-layer proteins: Theoretical scaffold	28
1.5.3 Scattering of S-layer proteins: Refinement of the structure	32
1.6 AFM-based single-molecule force spectroscopy	34
1.6.1 Basic principles	34
1.6.2 Mechanical stability can be regulated by ligand binding	36
1.6.3 S-layers: May the force be with them!	37
1.7 The structure of an SbpA unit cell	40
1.8 Molecular mechanisms of S-layer protein self-assembly	43

2	Publications	47
2.1	C. Horejs , D. Pum, U.B. Sleytr, R. Tscheliessnig. Structure prediction of an S-layer protein by the mean force method. (published in the <i>Journal of Chemical Physics</i>)	58
2.2	C. Horejs , D. Pum, U.B. Sleytr, H. Peterlik, A. Jungbauer, R. Tscheliessnig. Surface layer protein characterization by small-angle X-ray scattering and a fractal mean force concept: From protein structure to nanodisk assemblies. (published in the <i>Journal of Chemical Physics</i>)	66
2.3	C. Horejs , H. Gollner, D. Pum, U.B. Sleytr, H. Peterlik, A. Jungbauer, R. Tscheliessnig. Atomistic structure of monomolecular surface layer self-assemblies: Toward functionalized nanostructures. (published in <i>ACS Nano</i>)	77
2.4	C. Horejs , M.K. Mitra, D. Pum, U.B. Sleytr, M. Muthukumar. Monte Carlo study of the molecular mechanisms of surface-layer protein self-assembly. (published in the <i>Journal of Chemical Physics</i>)	91
2.5	C. Horejs , R. Ristl, R. Tscheliessnig, U.B. Sleytr, D. Pum. Single-molecule force spectroscopy reveals the individual mechanical unfolding pathways of a surface layer protein. (revision submitted to the <i>Journal of Biological Chemistry</i>)	104
2.6	U.B. Sleytr, B. Schuster, E.M. Egelseer, D. Pum, C. Horejs , R. Tscheliessnig, N. Ilk. Nanobiotechnology with S-layer proteins as building blocks. Chapter: Molecular modeling and computer simulations. In <i>Progress in Molecular Biology and Translational Science: Molecular Assembly in Natural and Engineered Systems</i> . (Ed. Howorka, S., Elsevier, Amsterdam) (in press)	104
3	Summary and Conclusions	110
4	Bibliography	112
5	Appendix	133
	List of Figures	133
	List of Abbreviations	137
	List of Symbols and Physical Constants	138
	Curriculum vitae	139

1 Introduction

The fascinating process of biological molecules organizing themselves into a variety of ordered large-scale structures has arrested the attention of scientists over decades.¹⁻³ Self-assembly of matter is one of nature's strategies to organize molecules on a large scale and provides the basis for the ordered formation of a variety of biological structures such as DNA,⁴ lipids,^{5,6} viruses,⁷ polymers^{8,9} and self-assembled monolayers.¹⁰ Mimicking nature's strategy, self-assembly provides a promising tool for the controlled bottom-up design of biological material on a very short length scale. Surface (S-)layers represent one remarkable example of a highly robust self-assembly system based on proteins.^{11,12} These crystalline protein layers cover a great variety of prokaryotic cells and are built up by so called S-layer proteins, which self-assemble *in vivo* and *in vitro* into highly ordered structures with defined pores and lattice constants and they provide the outermost barrier between the cells and the environment.¹³⁻¹⁵ Surface layers represent on the one hand a unique model system to investigate the molecular mechanisms of self-assembly,¹⁶ and can serve on the other hand as a basic matrix for the controlled bottom-up design of nanostructures. S-layer proteins have been extensively studied for such nanobiotechnological applications and lots of basic research has been done in order to understand the genetics, the structure, the function, the biochemistry and biophysics of S-layers,¹⁵ but still we don't know the atomistic structure of one unmodified native S-layer protein nor the structure and interactions of self-assembled S-layers at an amino acid level. The aim of this study is the elucidation of atomistic structural details of S-layers and the understanding of this fascinating self-assembly process by means of a combined approach of computer simulations and experimental techniques. In a first step molecular dynamics simulations of protein folding based on pre-modeled protein domains offered the opportunity to obtain a structural model at an atomistic level of one S-layer protein, which could be consecutively refined using small-angle X-ray scattering. The description of self-assembling systems in solution demands a formulation of an adequate theoretical background, which has been done using the concept of a fractal mean force for the intra- and intermolecular interactions of the proteins. Based on this structural model, Monte Carlo simulations were performed to study the self-assembly of S-layers in solution at an amino acid level, which opened up the path to determining the morphology of an S-layer lattice and the kinetics of the self-assembly process. A structural model of the unit cell of another S-layer protein, which self-assembles into more complex structures, has been calculated based on the same approach, but could be additionally refined using three-dimensional modeling by transmission electron microscopy. Finally, one individual S-layer protein has been mechanically unfolded using AFM-based single-molecule force spectroscopy, which revealed not only a high mechanical stability of S-layers, but gave also insight into the different unfolding pathways when the proteins are bound to their specific ligands in the cell membrane. This combination of techniques led to structural models of

two different S-layer proteins, which could fairly well explain experimental findings and observations. New structural details and molecular mechanisms could be revealed, which will help to better understand the concept of self-assembly itself as well as the specific use of S-layers in nanobiotechnology. The successful implementation of *in silico* methods for the modeling of S-layer proteins demonstrates the importance of this technique for the molecular study of proteins - when embedded in an adequate experimental background.

1.1 Surface layers: An outstanding self-assembly system

S-layers constitute the outermost cell component of most archaea and of hundreds of different species of bacteria. They are composed of single protein or glycoprotein species with molecular weights ranging from 40 to 200 kDa. These proteins crystallize on the cell surface into lattices with different symmetries (p1, p2, p3, p4, p6) and completely cover the cell surface.¹³⁻¹⁵ Figure 1 A shows an example of a cell covered by a crystalline S-layer and (C-F) electron micrographs of self-assembled S-layers with different symmetries.

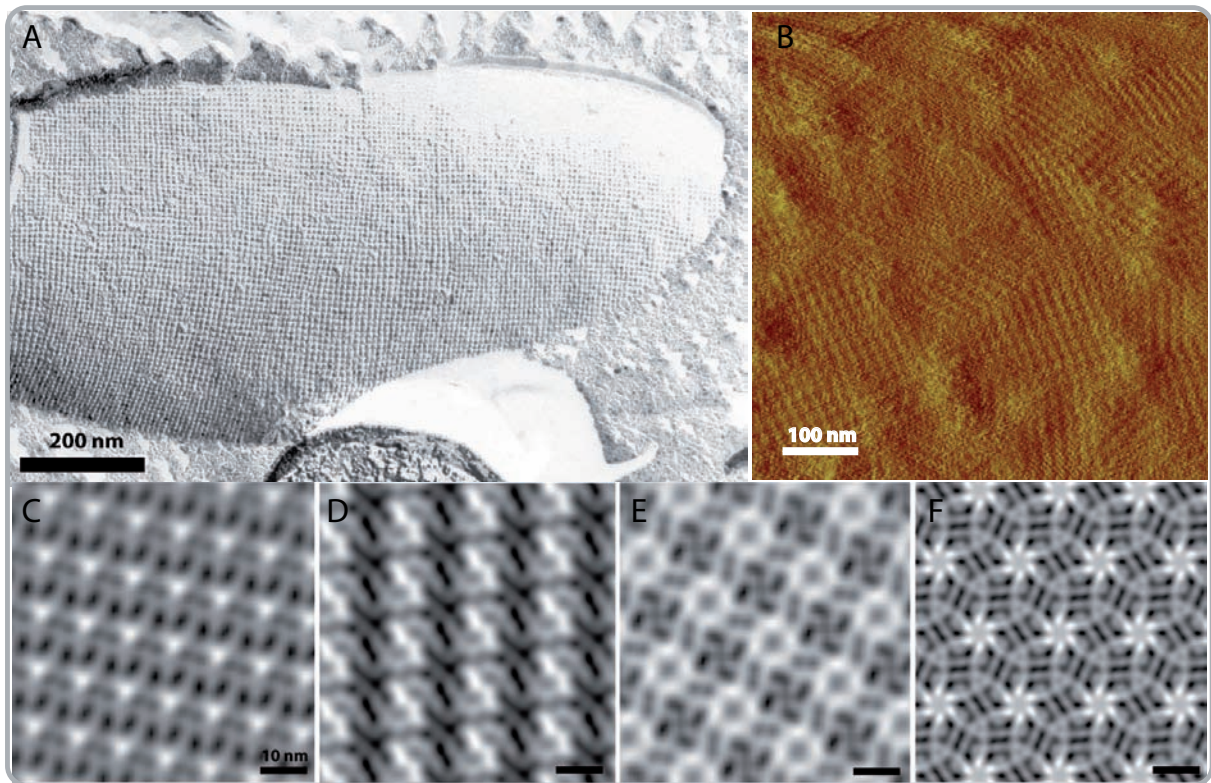


Figure 1: Surface layer proteins. (A) Electron micrograph of a freeze-etched preparation of *Desulfotomaculum nigrificans* showing a regular S-layer lattice.¹⁷ (B) Recrystallized patches of the S-layer protein SbpA from *Lysinibacillus sphaericus* CCM2177 on a silicon solid support visualized by atomic force microscopy. (C-F) Different lattices symmetries: (C) p1, (D) p2, (E) p4, (F) p6.

S-layers are about 4 to 20 nm thick. Due to the lattice symmetries they exhibit pores of identical size and morphology ranging from 2 to 8 nm.¹⁵ The specific S-layer proteins

are linked to each other and to the cell wall by non-covalent interactions. In archaea the S-layer is the only wall component integrated into the plasma membrane, where in gram-positive bacteria the proteins are bound to secondary cell wall polymers (SCWPs), which are covalently linked to the peptidoglycan.^{18,19} For some S-layer protein species certain domains could be identified that are responsible for the binding of SCWPs in the cell wall, which are called S-layer homologous (SLH) domains.^{20,21} The functions of S-layers for the cells might be mainly the determination and maintenance of the cell shape, the protection of the cell and to provide a molecular sieve, where they can also play a pathogenic and adhesion role.^{22,23} The remarkable feature of S-layer proteins lies in their ability not only to self-assemble *in vivo*, but also in solution, on lipids, at interfaces and on various surfaces¹³ (Figure 1 B shows exemplarily a crystallized S-layer on a silicon support and Figure 2 different self-assembled S-layers in solution).

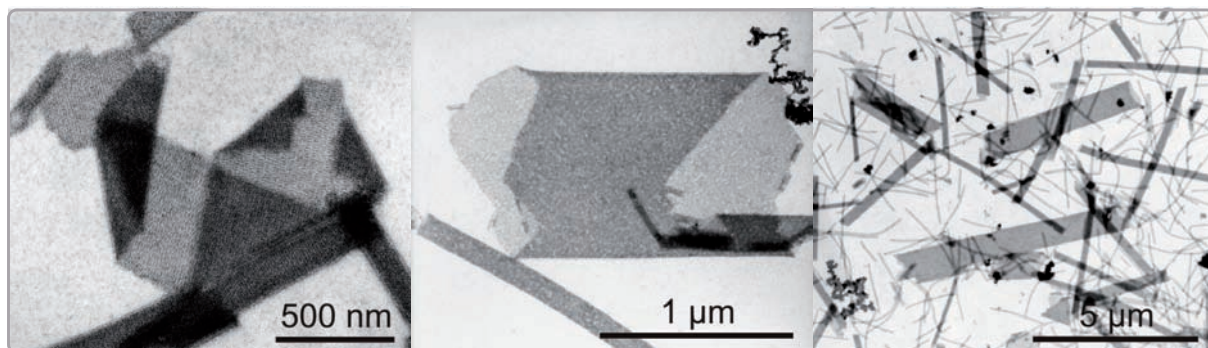


Figure 2: Self-assembly products of negatively stained S-layer proteins in solution visualized by transmission electron microscopy.²⁴

Additionally, due to the robustness of the self-assembly process, S-layer proteins can be genetically modified, truncated and fused with various molecules while still maintaining their intrinsic function. This fact has led to a wide range of nanobiotechnological applications, e.g. for biochip development, vaccine development, liposome coating, diagnostic systems, immobilized biocatalysts, binding of nanoparticles and pH biosensors (a state-of-the-art overview is given in²⁵).

1.1.1 The challenge of determining the atomistic structure of S-layers

S-layers have been successfully investigated and used in nanobiotechnology over decades and thus, lots of information could be gathered regarding the three-dimensional structure of S-layer proteins. The location of certain amino acids on an S-layer could be determined by introducing specific mutations and subsequent screening for the location of these mutations.^{26–28} Truncated recombinant forms were produced to obtain protein fragments that crystallize into isotropic three-dimensional crystals in order to perform X-ray crystallography studies.^{29–31} Electron microscopy was used to model the structure of an S-layer unit cell³² and small-angle X-ray studies of truncated proteins gave some structural

insights into the molecular organization of S-layer proteins.³³ The charge decoration and anisotropy of S-layer lattices could be revealed by various biochemical modifications^{34–38} and biophysical studies addressed the elucidation of structural properties of S-layers in general.^{39–41} All these studies helped to better understand the behavior and properties of S-layers. However, we do not know the full atomistic structure of one single unmodified native S-layer protein. Classical structure determination methods such as X-ray crystallography and nuclear magnetic resonance (NMR) pose problems due to the size and the intrinsic self-assembly capabilities of S-layer proteins.²⁵ S-layers in solution immediately interact in order to nucleate into small oligomeric precursors which grow into monomolecular large-scale structures.⁴² The isotropic crystallization of one individual protein is impossible for proteins having the intrinsic function to self-assemble in solution. Moreover, recent studies on the self-assembly kinetics of one S-layer revealed that this S-layer protein adapts a different conformation in its monomeric state than as part of the S-layer lattice, that is to say that the proteins first condense into an amorphous cluster in a rather extended conformation before they restructure to form a crystal of folded unit cells.^{41,43} The lack of atomistic structural details demands somehow a trial-and-error approach for the directed and specific use of S-layer proteins as a matrix for molecular construction kits and complicates the elucidation of the molecular mechanisms that make these huge proteins self-assemble so reproducibly into highly ordered large-scale layers.

1.1.2 The S-layer proteins SbsB and SbpA

This work focuses on the two S-layer proteins SbsB from *Geobacillus stearothermophilus* pV72/p2⁴⁴ and SbpA from *Lysinibacillus sphaericus* CCM2177.⁴⁵ SbsB consists of 920 amino acids (with 31 N-terminal amino acids comprising the signal peptide) and has a molecular weight of 98 kDa. This S-layer protein self-assembles into lattices with p1 symmetry (one protein per unit cell) with lattice constants of $a=9.91 \pm 0.06$ nm, $b=7.66 \pm 0.05$ nm and $\gamma=81.1 \pm 0.1^\circ$.⁴² The secondary structure distribution, as determined by far-UV CD yielded α -helical fractions of 7-11% and β -sheet fractions of 34-38%. Thermal and chemical unfolding studies revealed a complex unfolding behavior leading to the assumption of a multi-domain protein.³⁹ SbsB consists of three SLH domains, which are located at the N-terminus^{46,47} and are responsible for the specific binding of the SCWP.⁴⁸ The S-layer protein SbpA consists of 1268 amino acids (with 30 amino acids comprising the signal peptide) and self-assembles into lattices with p4 lattice symmetry (four proteins per unit cell). The lattice parameter of the square lattice is $a=13.1$ nm. SbpA also exhibits three N-terminal SLH domains and shows a similar secondary structure distribution as SbsB. This S-layer protein has been extensively studied and used for nanobiotechnological applications due the formation of p4 lattices.^{38,49–56}

1.2 Computer simulations: Beyond experimental structure determination methods

1.2.1 The "other" protein folding problem

Anfinsen and co-workers demonstrated almost exactly 50 years ago that the three-dimensional structures of proteins are completely determined by their amino acid sequences.⁵⁷ Thus, only twenty different amino acid types account for the huge amount of astonishingly diverse protein structures in nature. Based on the experiments by Anfinsen and co-workers it became clear that the native structures of proteins are thermodynamically stable conformations, which represent global minima of the accessible free energy.⁵⁸ Some years later Levinthal addressed the kinetics of protein folding by asking the question, how proteins can fold so fast while having so many possible conformations to randomly search through.⁵⁹ This argument became famous as the Levinthal paradox^{60,61}- how to find the native state by a random search among all possible configurations at such a short timescale? The thermodynamic question of how the amino acid sequence dictates the structure of a protein and the kinetic question of how a protein can fold so fast are together known as the protein folding problem.⁶² The protein folding problem was intensively addressed by means of statistical theories of energy landscapes^{58,63-66} and the simulation of simplified models of proteins⁶⁷⁻⁷¹ and it was referred to as the second half of the genetic code.⁷² Although still object of many scientific debates, a fairly well accepted conclusion has been reached, termed the "new view", as opposite to a variety of different "old" views of the protein folding problem.⁷³ The "new view" of protein folding theory regards the folding process as the progressive organization of an ensemble of partially folded structures through which the protein passes on its way to the native structure downhill the energy landscape.^{71,74} This energy landscape, which is basically the free energy of each conformation as a function of the degrees of freedom, has a rugged funnel-like shape toward the native structure.^{58,75} The concept of a folding funnel together with recent insights into the folding code and the kinetics represent a huge step toward the solution of the protein folding problem,⁶² where much controversy still remains on various topics such as pathway diversity, denaturated states, kinetic and energy barriers and misfolding of proteins.⁷⁶ However, recent experimental results^{77,78} provide strong evidence that there exists an ensemble of collapsed minimum energy states comprising the funneled energy landscape and from this ensemble the native state is acquired through a barrier-limited transition.

Embedded in this theoretical background the prediction of the three-dimensional structure based on the amino acid sequence by means of computational methods including bioinformatic tools as well as different simulation algorithms becomes possible. This challenge was called by Karplus and Kuriyan the "other" protein folding problem,⁷⁹ namely to predict the native conformation based only on the primary sequence. Protein structure prediction schemes can be categorized into three classes: *ab initio* or *de novo*

prediction, fold recognition and homology modeling, where the last two methods are template-dependent and normally not sufficient for a complete prediction of the three-dimensional structure.^{80,81} Although *de novo* predictions still face various problems regarding the accuracy, the expensive cost of time - even the fastest proteins take microseconds to fold, and the reliability of simulated structures due to the difficulty of sampling the whole energy landscape,⁸² all-atom structure prediction and folding simulations have been successfully applied to predict structures and to give insights into protein folding and unfolding.⁸³⁻⁹⁰ Moreover simplified protein models and coarse-graining strategies paved the way for faster simulation procedures and enabled the use of longer time scales. A yearly assessment of various structure prediction methods organized by the protein structure prediction center (www.predictioncenter.org) shows the astonishing progress that has been already achieved by *in silico* methods.⁹¹

1.2.2 The protein structure prediction process

The first step toward the prediction of the tertiary structure of a protein is the identification of possible related primary sequences. Various online databases and tools are available to find relationships between protein sequences.⁹² Additionally, the secondary structure and possible domains can be predicted by various algorithms based on the amino acid sequence (an overview of databases and tools is collected at www.expasy.org). If structural and sequence homologues can be identified, a possible structure might be predicted by using homology or comparative modeling, which is based on the idea that similar protein sequences encode similar three-dimensional structures.⁹³ Thereby one can produce all-atom three-dimensional models based on the alignment of the protein's sequence to one or more related known protein structures,^{94,95} which is referred to as sequence-template alignment.⁹⁴ This method is restricted to sequence alignments with relatively high sequence identity (>30%).⁹⁶ Another possible method to predict the three-dimensional structure of a protein even in the absence of clear homology is called fold recognition or threading.^{96,97} This method works by comparing the target sequence with potential fold templates using potentials and similarity scoring methods. The best structural fit is the one with the lowest energy score. The term threading refers to a method where a sequence is fitted or threaded on to the backbone coordinates of known protein structures.⁹⁶ The applied potentials highly influence the outcome of the modeling process, where various structures were successfully predicted using specific pair-pair interactions.⁹⁸ Both methods strongly rely on the existence of experimentally known protein structures. A method, which is independent of structural homologues and sequence similarities, is called *ab initio* or *de novo* structure prediction. The goal is to predict the three-dimensional structure based entirely on the laws of physics and chemistry and the amino acid sequence. The approach is to simulate the interactions of amino acids, which guide the folding of the protein toward the minimum of the folding funnel that is to say

the global free energy minimum. Various different simulation algorithms and potential functions as well as protein representations ranging from simple lattice to all-atom models can be applied in order to simulate the protein folding process.

1.2.3 Molecular dynamics

1.2.3.1 Theoretical background: Newton and motion "Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare."¹ Newton's first law of motion states that an object will remain at rest or if in motion continues in uniform motion unless an external force acts upon it. This fundamental law of physics, which was already formulated in 1687, provides the basis for molecular dynamics, which is the science of simulating motion of a system of particles by generating successive configurations by integrating Newton's laws of motion.^{99,100} Therefore, molecular dynamics - in contrast to the static, frozen view we can obtain from experimental techniques such as X-ray crystallography - gives us the opportunity to study the dynamics of molecular systems by predicting motion.^{99,101} Newton's second law indicates that force equals the rate of change of momentum $F = ma$ and the third law states that to every action there is an equal and opposite reaction. These laws of motion lead to the following differential equation that describes the motion of a particle of mass m along one coordinate x with the force F acting upon it:¹⁰⁰

$$\frac{d^2x}{dt^2} = \frac{F_x}{m} \quad (1)$$

If we follow the concept of classical mechanics and take into account the principle of least action, matter or in the present case molecules will choose of all possible trajectories the one of least action, where in Newton mechanics the least action is defined by $W = \int_{t_0}^{t_1} \mathcal{L}(t, q_i(t), p_i(t)) dt$ with q_i being generalized coordinates and p_i generalized momenta, which span the phase space of the system containing all possible configurations. $\mathcal{L}(q_i, \dot{q}_i, t) = K - V$ is the Lagrange function with K being the kinetic and V being the potential energy, where the equations of motion might be now written down in the following way:

$$\frac{d}{dt} \frac{\partial \mathcal{L}(q_i, \dot{q}_i)}{\partial \dot{q}_i} = \frac{\partial \mathcal{L}}{\partial q_i} \quad (2)$$

Including the definition of the kinetic energy $\sum_i \frac{m_i \dot{q}_i^2}{2}$, the potential energy $V(q_i)$ and Cartesian coordinates r_i , we obtain Newton's second law of motion $m_i \ddot{r}_i = F_i$, where $F_i = \nabla_{r_i} \mathcal{L} = -\nabla_{r_i} V$ represents the force on the particle i in the system. The generalized momentum can be calculated by $p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}$. Consequently the Hamiltonian equations of motion as a Legendre transformation of the Lagrange function can be written as:

¹Isaac Newton, *Philosophiae naturalis principia mathematica*, 1687

$$\dot{p}_i = -\frac{\partial \mathcal{H}(p_i, q_i)}{\partial q_i} \quad (3)$$

$$\dot{q}_i = \frac{\partial \mathcal{H}(p_i, q_i)}{\partial p_i} \quad (4)$$

where $\mathcal{H}(p_i, q_i) = K(p_i) + V(q_i)$ with assuming a Born-Oppenheimer approximation averaging out the rapid motion of the electrons. Therefore, calculating trajectories involves solving either a system of 3N second-order differential equations according to Newton's equations of motion or 6N first-order differential equations following Hamiltonian mechanics - when having N particles in the system.¹⁰² There are various algorithms for integrating the equation of motions like finite difference methods and predictor-corrector algorithms,¹⁰⁰ which have a main influence on the stability and time resolution of molecular dynamics simulations. Based on Newton's basic three laws and the resulting formulation of equations of motion three key quantities for the simulation of molecules are related: time, configuration and energy.⁹⁹

1.2.3.2 Interaction potentials and force fields The potential energy term in the Hamiltonian contains information regarding inter- and intramolecular interactions between N particles or atoms in a system with the Cartesian coordinates r_i , which can be divided into terms depending on the coordinates of single atoms, atom-pairs, atom-triplets and so on:

$$V(r_i, r_j) = \sum_i v_1(r_i) + \sum_i \sum_{j>i} v_2(r_i, r_j) + \sum_i \sum_{j>i} \sum_{k>j>i} v_3(r_i, r_j, r_k) + \dots \quad (5)$$

The first term v_1 in equation 5 describes the effect of an external field on the potential energy, the second term v_2 represents the pair potential, which only depends on the distance $r_{ij} = |r_i - r_j|$ between the atoms i and j and v_3 includes atom-triplets, which play a major role in fluid systems. The three-body terms in the potential energy function are normally not included in molecular dynamic simulations of biomolecules due to the expensive time cost, but many-body effects can be adequately incorporated by defining an effective pair potential $v^{eff}(r_i r_j)$. In order to describe the energy of a system based on pairwise interactions, empirical model force fields were developed, because the direct calculation by means of quantum chemistry is not quite feasible for large molecules.¹⁰³ Force fields describe an approximation of the potential energy function along with an associated set of parameters, which are normally obtained from experimental and/or quantum mechanical studies of small molecules.¹⁰⁴ The force that every atom in the system experiences is specified by the force field and a simple set of classical pairwise functions is used to describe the potential energy of the system:¹⁰⁰

$$\begin{aligned}
V(r^N) = & \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \\
& \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \sum_{nonbonded} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right)
\end{aligned} \tag{6}$$

The first term in equation 6 models the interactions between bonded atoms by a harmonic potential, the second harmonic potential models angle bending, the third term is a torsional potential that accounts for bond rotation and the fourth term comprises non-bonded interactions, where the first term corresponds to van der Waals interactions, here modeled by a 12-6 Lennard-Jones potential, and the second term to electrostatic interactions modeled by a simple Coulomb potential. The combination of such a potential energy function and the various parameters that go into it constitutes a force field. There is a large number of force fields that have been specifically developed for the simulations of proteins such as AMBER,¹⁰⁵ CHARMM,¹⁰⁶ OPLS,¹⁰⁷ GROMOS^{108,109} and many others.¹⁰³

Long-range interactions in the system can be additionally included using various methods such as Ewald summation¹¹⁰ or the reaction field method,¹¹¹ which both take into account long-range electrostatic interactions.

1.2.3.3 Free energy calculation Statistical mechanics connects microscopic simulations and macroscopic thermodynamic observables and based on statistical mechanics principles the probability of finding a molecular system in one state or another is governed by the difference in the free energy of those states.^{112,113} The differences in the free energy also guide a protein to fold into its native structure and defines the stability of possible protein conformations. Therefore, when simulating the folding or unfolding of a protein the main parameters of interest are the free energies for all possible conformations. The absolute free energy of a system with N atoms is defined by:

$$A = -kT \ln Z \tag{7}$$

where k is the Boltzmann constant, T the temperature and Z the canonical partition function as defined in classical statistical mechanics:

$$Z = \frac{1}{(h^{3N} N!)} \int \int e^{\frac{-\mathcal{H}(p,q)}{kT}} dp dq \tag{8}$$

where h is the Planck constant, \mathcal{H} the Hamiltonian, q the coordinates and p the momenta. The probability of finding the system in a certain configuration is proportional to the Boltzmann factor $e^{\frac{-\mathcal{H}(p,q)}{kT}}$:

$$\pi(p, q) = \frac{e^{\frac{-\mathcal{H}(p, q)}{kT}}}{\int \int e^{\frac{-\mathcal{H}(p, q)}{kT}} dp dq} \quad (9)$$

The ensemble average of a given quantity Q is then given by:

$$\langle Q \rangle = \frac{\int \int Q(p, q) e^{\frac{-\mathcal{H}(p, q)}{kT}} dp dq}{\int \int e^{\frac{-\mathcal{H}(p, q)}{kT}} dp dq} = \int \int Q(p, q) \pi(p, q) dp dq \quad (10)$$

The integral in equation 8 runs over all 6N phase space coordinates and is only computable for very small systems due to insufficient sampling during finite simulations and consequently weak Boltzmann statistics. The integral in equation 10 is easier to evaluate and free energy differences can be expressed by ensemble averages of atomic configurations. Therefore, free energy differences can be calculated using statistical molecular dynamics simulations. Different methods are available for the calculation of solvation free energies, binding free energies and conformational free energies. These methods may be performed either under equilibrium or non-equilibrium conditions and the sampling may be either restrained or unrestrained.^{100,114} However, only a very small part of the total phase space can be sampled in molecular dynamics simulations, which complicates the calculation of the difference in free energy between two states of the system. The simplest method is to count the number of configurations that satisfy a certain objective criterion.¹¹⁵ In the thermodynamic integration technique¹¹⁶ a discrete coupling parameter $0 < \lambda < 1$ is introduced, which can be altered continuously in the simulation, along which the following integral can be determined:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \mathcal{H}(p, q)}{\partial \lambda} \right\rangle d\lambda \quad (11)$$

In free energy perturbation theory⁶⁰ the change of the free energy is calculated from the expectation value of the Boltzmann weighted free energy difference between two states 0 and 1:¹¹⁴

$$e^{-\beta \Delta A} = \langle e^{-\beta \Delta V} \rangle_{\lambda=0} \quad (12)$$

where $\beta = \frac{1}{kT}$ and ΔV is the potential energy of state 1 relative to the reference state 0. These methods hold true for systems under equilibrium conditions. To obtain converged statistics, the conformational space has to be adequately sampled, which is a rather complicated task due to the complex energy landscapes of molecules containing many local minima and maxima. On the one hand λ might be treated as a dynamic coupling parameter in order to chose effective pathways and enhance the sampling efficiency, on the other hand umbrella sampling¹¹⁷ offers the opportunity of imposing a potential energy bias to enforce uniform sampling. Thereby an external potential $V_{bias}(\lambda)$ is included that virtually smoothens the rough free energy landscape, where a reasonable estimation

of V_{bias} poses a great challenge. Another highly effective possibility constitute non-equilibrium methods.

According to the second law of thermodynamics, if one parameter of a system like the coupling parameter λ is changed from λ_0 to λ_t , the average work done on the system cannot be smaller than the difference between the free energy at the initial state at $t = 0$ and the final state at $t = t$:

$$\Delta A \leq \langle W \rangle \quad (13)$$

According to that non-equilibrium processes provide only an upper limit for the free energy difference.¹¹⁸ In 1997 Jarzynski^{119,120} formulated an equality between the change of the free energy and the work for infinitely slow changes of λ . This equality opened up the path for determining free energy differences from non-equilibrium processes and besides experimental impact¹²¹ led to a new method called steered molecular dynamics,¹¹⁸ where time-dependent external forces can be applied to the system.

$$e^{-\beta\Delta A} = \langle e^{-\beta W} \rangle \quad (14)$$

If the coupling parameter λ is now a physical degree of freedom, such as an interatomic distance or another conformational variable, the free energy profile as a function of this coordinate is called a potential of mean force (PMF) and λ is called the reaction coordinate and termed ζ . If the process is infinitely slow or reversible, then the work is the reversible work or the potential of mean force $\omega(\zeta)$ and is therefore related to the mean force by:

$$\Delta\omega[\zeta(\zeta')] = \Delta A[\zeta(\zeta')] = - \int_{\zeta_0}^{\zeta'} \langle F[\zeta(\zeta^*)] \rangle d\zeta^* \quad (15)$$

where $\langle F[\zeta(\zeta^*)] \rangle$ is the mean force acting along the reaction coordinate. Hence, the derivative of the free energy and the average force acting along the reaction coordinate ζ can be written as:

$$\frac{dA(\zeta)}{d\zeta} = - \langle F_\zeta \rangle_\zeta \quad (16)$$

Jarzynski's equality^{119,120} is a relation between equilibrium free energy differences and work done through non-equilibrium processes¹¹⁸ and based on this discovery the free energy landscape of non-equilibrium systems can be sampled, where it still remains a big challenge to sample rare events and overcome free energy barriers that separate thermodynamic states of interest, which is the case for the simulation of macromolecules. In order to calculate the potential of mean force constrained¹²²⁻¹²⁶ or unconstrained^{127,128} simulations can be performed. When using constrained molecular dynamics simulations to calculate $\frac{dA}{d\zeta}$, the reaction coordinate ζ is kept constant by applying a force $-\lambda\nabla\zeta$, where

λ is a Lagrange multiplier associated with $\zeta(x) = \zeta^*$. Thereby the applied constraint force $-\lambda\nabla\zeta$ prevents ζ from changing and represents a direct measurement of the derivative of the free energy. However, a sampling of all possible pathways between two different states seems to be rather impossible using a constraint force. In contrast, unconstrained molecular dynamics methods, where ζ is not constrained, offer a more uniform sampling of the free energy landscape through a biased sampling along ζ . One such method, which was also used in this work, is the adaptive biasing force method (ABF).^{128,129} In the framework of this approach, the force F_ζ is accumulated in small bins of size $\delta\zeta$ and the mean force along the reaction coordinate ζ is estimated from a running average in the respective bin:

$$\frac{dA(\zeta)}{d\zeta} = \left\langle \frac{\partial V(x)}{\partial \zeta} - \frac{1}{\beta} \frac{\partial |J|}{\partial \zeta} \right\rangle = -\langle F_\zeta \rangle_\zeta \quad (17)$$

where $V(x)$ is the potential energy and $|J|$ is the determinant of the Jacobian for the inverse transformation from generalized to Cartesian coordinates. $\langle F_\zeta \rangle_\zeta$ is the average force acting along the reaction coordinate.¹²⁷ Based on this calculation, an external force is applied $-\langle F_\zeta \rangle_\zeta \nabla\zeta$ that opposes the mean force. Therefore, the adaptive biasing force applied along the reaction coordinate ζ equals:

$$F^{ABF} = -\nabla\tilde{A} = -\langle F_\zeta \rangle_\zeta \nabla\zeta \quad (18)$$

where \tilde{A} is the estimate of $\frac{dA(\zeta)}{d\zeta}$ in the respective bin. This biasing force ensures with the evolution of time that the force acting along ζ averages to zero and the system moves by its self-diffusion properties, while the free energy is directly reconstructed from its derivative.

1.2.3.4 Limitations and challenges Molecular dynamics simulations of protein folding and consequently, the solution of the "other" protein folding problem, namely to calculate the native three-dimensional conformation of a whole protein, faces in practice still a variety of difficulties.^{82,99} Force fields, although especially designed and parametrized for proteins, are still not able to completely reproduce the potential energy surface of proteins.¹³⁰ Moreover, energy functions, especially regarding the solvent, electrostatic interactions and cross-terms influence the reliability of the results. During the folding process of a protein, long-range water-mediated potentials smooth the underlying folding funnel and there is a lot of research going on to determine the exact role explicit water molecules play in protein folding.^{74,131} As mentioned in chapter 1.2.3.2 three-body terms that account for fluid systems, are normally not included in molecular dynamics simulations due to the expensive time cost and effective pair potentials as well as buried surface area terms should account for hydrophobic effects due to water molecules. These are still poor representations of solvents in simulations and even when explicit rather than implicit water models are used, they cannot capture the important effect of water

for protein folding. The simulation of biomolecules demands timescales of micro- to milliseconds, which is still not possible using computer simulations although a lot of effort is made to develop adequate hardware for the simulation of such large systems.⁸² Another problem arises due to the complex energy landscapes of biomolecules and the related sampling of configurational states to give better Boltzmann statistics. However, if appropriate initial conditions, adequate force fields and analyzing methods are applied, molecular dynamics simulations of protein folding are possible as it has been successfully proved.⁸³⁻⁹⁰ The additional development of coarse-grained models and experimental input help to simulate protein structures and an ongoing effort to enhance sampling methods and force fields promises a huge step toward the solution of the "other" protein folding problem. If one is interested in different protein folding pathways or the whole ensemble of possible collapsed states, thus the detailed structure of the folding funnel, all-atom simulations have to be carried out, which will remain, even with better computational equipment, a great challenge in the future, especially for larger systems. However, three-dimensional structures can be already calculated using coarse-grained models and experimental input, without sampling the whole conformational space, while ensuring that the main paradigms of protein folding are captured in the simulations. Protein folding follows classical physical and chemical principles and therefore, folding simulations based on the same principles - if embedded in an adequate experimental background - become possible in order to predict three-dimensional protein conformations.

1.2.4 Langevin and Brownian dynamics

Langevin and Brownian dynamics are both methods based on molecular dynamics, where additional terms are included in order to better capture the influence of a solvent and diffusion properties of the system. The choice, which kind of algorithm should be used for the simulation, depends upon which contributions are expected to dominate in the motion of the system. Langevin dynamics is based on the Langevin equation,¹³² which describes Brownian motion and represents a stochastic differential equation. The Langevin equation incorporates a frictional term and a random force:

$$m \frac{d^2 r}{dt^2} = F(r) - \xi \frac{dr}{dt} + R \quad (19)$$

with ξ being the friction coefficient and R the random forces experienced by the atom. The random forces R have a Gaussian probability distribution and represent the random impulses of the solvent and solute molecules. By including the friction coefficient ξ the frictional drag of molecules in an implicit solvent can be simulated.

In Brownian dynamics the Langevin equation is numerically integrated while a very large friction coefficient ξ is assumed. Thereby a random walk, describing the motion of a molecule due to the damping effect of the solvent, can be simulated yielding a diffusional trajectory.¹⁰⁴

1.2.5 Monte Carlo simulations

Monte Carlo methods, in contrast to molecular dynamics, represent a stochastic approach in order to generate representative configurations of a system under specific thermodynamic conditions, where the positions of the particles or atoms are randomly changed. In contrast to molecular dynamics, there is no contribution of the momenta of the particles, so the potential energy does not depend on the velocities of the atoms. The Metropolis algorithm¹³³ generates a Markov chain of states implying that the states are not connected in time. Thereby the positions of the particles of the system are changed randomly and the potential energy of the new configuration is calculated. If the energy value of the new configuration is less than of the former configuration, the new one is accepted, if it is greater than for the former configuration, the new one is accepted with a Boltzmann probability $\exp(-\frac{\Delta V_{nm}}{kT})$ with $\Delta V = (\sum_{j=1}^N v(r_{ij}^n) - \sum_{j=1}^N v(r_{ij}^m))$ being the potential energy difference and n and m the two different states, and otherwise rejected based on the probabilities of the two states:

$$\frac{\pi_n}{\pi_m} = \frac{Z_{NVT}^{-1} \exp(-\beta V_n)}{Z_{NVT}^{-1} \exp(-\beta V_m)} = \frac{\exp(-\beta V_n) \exp(-\beta \Delta V_{nm})}{\exp(-\beta V_m)} = \exp(-\beta \Delta V_{nm}) \quad (20)$$

with Z_{NVT} being the canonical partition function. Therefore, a random number is generated between zero and one, which is then compared with $\exp(-\frac{\Delta V_{nm}}{kT})$ and if it is less, the move is accepted. This process is iterated until sufficient statistics are achieved.^{100,102} Monte Carlo methods can serve as an alternative for the calculation of structural and thermodynamic properties of a system especially if the dynamics, that is the time evolution of the system, are not of particular interest. If adequate potential functions and smart coarse-graining procedures are defined, Monte Carlo simulations can be successfully used in order to simulate biomacromolecules with much less computational time-cost as in molecular dynamics.

1.2.6 Coarse-graining and statistical potentials

Questions regarding protein folding, structure and dynamics can be also addressed at a non-atomistic level when using low-resolution representations of proteins, where these simplified models consist of much fewer parameters. The process of reducing the degrees of freedom is called coarse-graining.¹³⁴ Thereby the protein is represented by a chain of beads, where the number of beads per amino acid residue defines the degree of "coarsening" and bonds are either represented as simple lines omitting covalent terms or as elastic springs in order to enable motion of the beads relatively to each other. The beads are normally placed in the center of mass of the amino acid or the side chain if more beads per amino acid are used. The protein can be either simulated off-lattice or the beads can be configured on a two- or three-dimensional spatial lattice, which is then referred to as lattice model. So called "Gō models" represent the first lattice models

that were applied in order to study protein stability and kinetics.^{68,69,135} Gō models are one-bead per amino acid representations of proteins, where the structure is biased toward the native configuration by means of simple attractive or repulsive non-bonded interactions between the beads.^{134,135} Therefore non-physical potentials are used in order to reach the native conformation. The simplest lattice models using physical potentials are HP models, where every bead is assigned either a hydrophobic (H) or a polar (P) flavor.¹³⁶ Each configuration of the bead-chain is a self-avoiding walk on a lattice, where the modeled protein is subject to excluded volume and an HH attraction free energy, i.e. the contacts between hydrophobic beads are favorable. Therefore, the native state is defined as the set of conformations with the largest possible number of HH contacts. The obvious disadvantage of such a simplified model is the loss of resolution, which leads to problems regarding the identification of the correct native state.¹³⁷ However, a rough coarse-graining approach like the use of lattice models opens up the possibility to explore larger conformational changes at much longer timescales. Moreover, lattice models allow the direct enumeration of conformations, which enormously simplifies the computation of a partition function. HP models represent only two types of flavors, which is a rather oversimplified representation of the 20 possible amino acid types. A much more realistic description of a protein can be achieved by applying a unique flavor for every amino acid bead incorporating contact potential terms and charges. These empirical knowledge-based potentials, which represent either solvent-mediated or intramolecular effective residue-residue contacts, are calculated based on contact frequencies of residues in the protein data base.^{138,139} The overall energy is calculated as a sum of pairwise interactions, where for every possible amino acid pair a different contact energy value is applied. Generally, this kind of potentials account for the effect of closely located amino acids due to the fact that they are computed based on experimentally determined native protein conformations. The amino acids interact in a certain cutoff range, where additionally electrostatic interactions can be included. Shortcomings are the overestimation of hydrophobic interactions and the limited applicability to denaturated structures. A variety of different effective statistical potentials has been formulated,^{140–143} where different interaction schemes can lead to different simulation results.^{144–146} However, such simplified protein models offer the opportunity to simulate large biological systems at a long timescale, which makes them an interesting alternative for the simulation of S-layer proteins.

1.3 Structure prediction of an S-layer protein: A hybrid approach

In the framework of my diploma thesis the three-dimensional structure of the S-layer protein SbsB from *Geobacillus stearothermophilus* pV72/p2⁴⁴ was predicted as a first attempt to get information on the atomistic structure of one complete, unmodified S-layer protein. This work was published in 2008 in the Journal of Chemical Physics¹⁴⁷ (see chapter 2.1). The calculated structural model provides the basis for all consecutive structural analyses presented in this work. In order to perform protein folding simulations using molecular dynamics, the protein domains have to be pre-modeled by means of bioinformatics and fold recognition (see chapter 1.2.2) to get an adequate initial configuration for the energy minimization process. Previous experimental studies could identify three S-layer homologous (SLH) domains as well as one additional secondary cell wall binding (SCWP)-binding domain, which together enable the anchoring of the protein to the underlying cell wall^{46,47} (see chapter 1.1 and 1.1.2). Additionally, circular dichroism studies revealed the secondary structure distribution.³⁹ Sequence homology searches for the whole protein sequence showed similarities to other S-layer proteins, especially regarding the location of the SLH domains at the N-terminal part of the protein. Additionally similarities to Ig-like domains could be found. These domains mainly consist of β -sheet like structures and are common motifs for proteins located at the cell surface. Secondary structure predictions revealed a similar distribution of secondary structure elements as the circular dichroism measurements. The N-terminal part mainly consists of α -helices, where the rest of the protein contains β -sheet like structures. Domain predictions, where domain profiles rather than pure sequence similarities are compared, confirmed the existence of similarities to Ig-like and fibronectin-type III domains. Based on these data, the protein is split into eight individual domains. In a first attempt the structure of every single domain is predicted using fold recognition to obtain input coordinates reflecting the possible secondary and tertiary structures. Figure 3 A shows the resulting three-dimensional models. Each pre-folded domain is consequently calibrated in a water sphere, which is shown in figure 3 B. The equilibrated domains are joined and simulated in vacuum (figure 3 C) resulting in a three-dimensional structure as shown in figure 3 D. The resulting structural model of the S-layer protein SbsB is then systematically tested using steered molecular dynamics and the mean force method (see chapter 1.2.3.3). Thereby parts of the protein are reversibly pulled along a chosen reaction coordinate ζ and the protein is deformed to quantify the stability of the structure. The potential of mean force is calculated as a function of the deformation of the protein. Figure 3 F shows exemplarily one plot of the obtained mean force (MF), potential of mean force (PMF) and probability distribution $P[\zeta]$. To calculate the PMF the adaptive biasing method is used as described in chapter 1.2.3.3. Using this approach a fairly detailed structural model of this S-layer protein could be calculated, while the simulated

conformation equals the minimum of the free energy landscape, at least in the one sampled by our method. All simulations were performed using the software NAMD,¹⁴⁸ which generates structure files from the CHARMM force field.¹⁰⁶

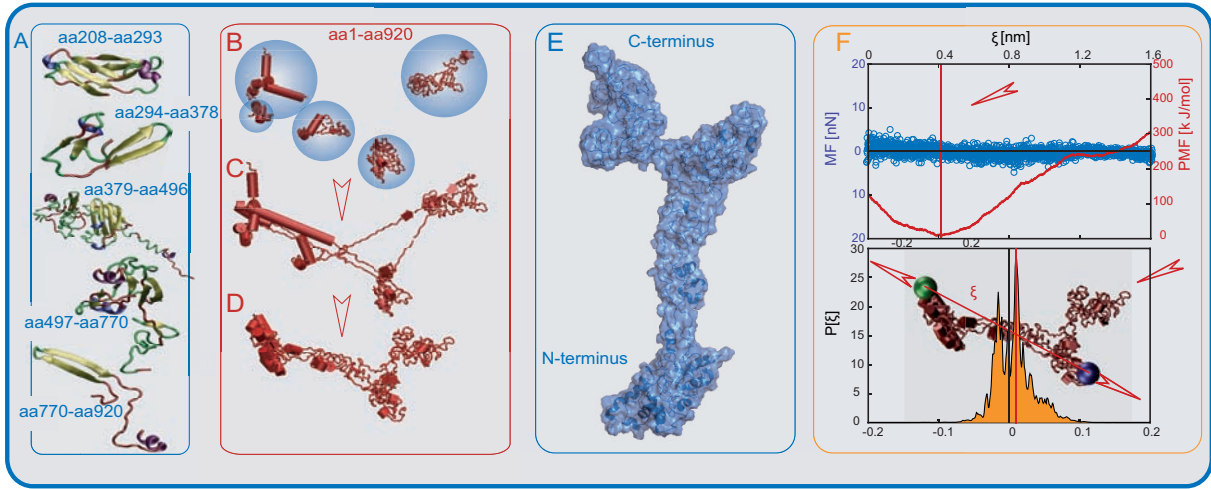


Figure 3: (A) 3D models of the single domains of the S-layer protein SbsB created by fold recognition. Yellow arrows: beta-sheets, violet strands: alpha-helices, green line: turns, and red line: coils. Modeling method: (B) the individual domains were equilibrated in water spheres at 310K, (C) joined in vacuum, and (D) the final structure was obtained by molecular dynamics simulations. (E) Structural model of SbsB. The protein is L-shaped, where the L is formed by the C-terminal domains. The N-terminus contains the SLH domains and is mainly made up of alpha-helices. (F) Structural analysis of the monomer structure by a calculation of the global free energy. The protein was deformed along the reaction coordinate ζ . Mean force values MF are indicated by open blue circles. The red full line gives the potential of mean force PMF, which has a clear minimum at $\zeta = 0$. Orange body gives the local density probability distribution $P[\zeta]$. The model of the protein is given as an insert, the reaction coordinate is indicated, and the green and blue spheres indicate fixed regions.

This model constitutes the first calculated three-dimensional structure of an S-layer protein using various bioinformatic tools, adequate experimental background and protein folding simulations - a hybrid approach using all available structure prediction methods. Although later refined by other experimental methods,⁴² this modeled structure turned out to serve as a good basis for the interpretation and the understanding of experimental findings and S-layer proteins in general.

1.4 Small-angle X-ray scattering

Small angle X-ray scattering (SAXS) belongs to the small-angle scattering (SAS) methods together with small-angle neutron scattering (SANS). These techniques allow the study of the structure and the interactions of biological macromolecules in solution. The outstanding advantage of this approach is the possibility to investigate all sizes of macromolecules in a variety of conditions without the need to crystallize the samples. The main disadvantage is the low resolution, which arises due to the small angles and the random orientation of the molecules in solution lacking any crystalline order. This technique dates back to the 1950s and has been historically mainly used to obtain only a few parameters, such as the radius of gyration, the molecular weight and molecular distances.¹⁴⁹ Today, due to better instrumentation and computational effort, SAXS can be used for *ab initio* shape determination, modeling of macromolecular complexes, conformational transitions, assembly and kinetics of biological systems.^{150–155}

1.4.1 Theoretical basics of X-ray scattering

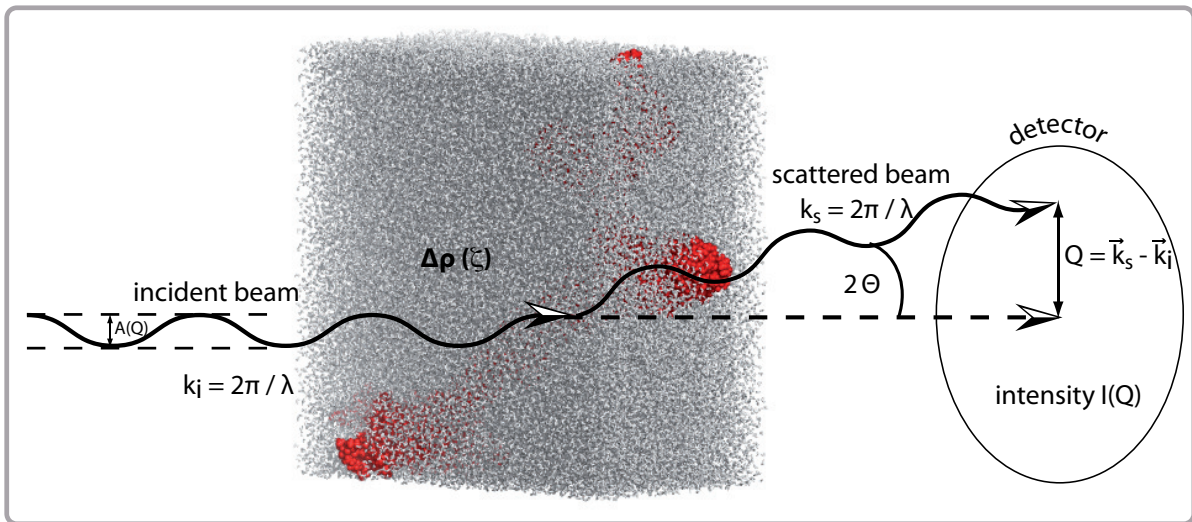


Figure 4: Schematic representation of a SAXS experiment. An incident X-ray beam with a wave vector of $|\vec{k}_i| = \frac{2\pi}{\lambda}$ is elastically scattered by atoms in solution and the scattered beam with $|\vec{k}_s| = \frac{2\pi}{\lambda}$ is recorded as a function of the scattering angle θ . The scattered beam emerges from electrons that resonate with the frequency of the X-rays and consequently emit secondary waves, which interfere with each other. The scattering process is described by a Fourier transformation from the coordinates of the atoms ζ of the molecule to the reciprocal space of the scattering vectors $\vec{Q} = \vec{k}_s - \vec{k}_i$. The scattering amplitudes $A(Q)$ are functions of the momentum transfer $|\vec{Q}| = \frac{4\pi \sin \theta}{\lambda}$ and represent the Fourier transform of the electron density distribution $\Delta\rho(\zeta)$. The phase change $\Delta\varphi = \vec{Q} \cdot \zeta$ depends on the electron coordinates ζ , but is not detectable. Therefore, the detected scattering intensity $I(Q) = |A(Q)|^2$ contains information about the electron density distribution of the sample.

X-ray photons with an energy of about 10 keV have a wavelength λ of about 0.10 – 0.15 nm. The main principle of SAXS is illustrated in figure 4. When X-rays with a wave

vector $|\vec{k}| = \frac{2\pi}{\lambda}$ are elastically interacting with atoms, the electrons start resonating with the same frequency as the X-rays and consequently emit secondary waves with the same frequency $\exp(-iQ\zeta)$, which interfere with each other. Q is the scattering vector or momentum transfer vector $|Q| = \frac{4\pi \sin \theta}{\lambda}$, where $Q = \vec{k}_s - \vec{k}_i$ is the difference between the wave vector of the scattered beam \vec{k}_s and the incident beam \vec{k}_i , and ζ is the position vector of the respective electron under consideration. To observe interference the waves have to be coherent. If a wave plane is scattered by electrons, it changes its direction by an angle θ and a phase shift $\Delta\varphi$ occurs, which is related to the separation of the scatterers $\Delta\varphi = Q \cdot \zeta$.

This scattering process is characterized by a reciprocal law relating real coordinates ζ to the reciprocal space of scattering vectors Q , where the amplitudes are proportional to the Fourier transform of the electron densities. In order to get the full scattering amplitude and due to the fact that the waves are coherent, all waves have to be added up:

$$I(Q) = |\sum b_j \exp(-iQ\zeta)|^2 \quad (21)$$

where b_j is the scattering length of one atom. However, the phase difference as well as the amplitudes are not detectable, where the time-averaged energy flux, i.e. the intensity $I(Q)$ can be measured. The intensity $I(Q)$ is proportional to the number of photons scattered in the direction of Q and is related to the amplitudes by:

$$I(Q) = |A(Q)|^2 \quad (22)$$

If we assume a continuous distribution with the density $\rho(\zeta)$ and approximate the sum by an integral we obtain:

$$A(Q) = \int \rho(\zeta) \exp(-iQ\zeta) d\zeta \quad (23)$$

In SAXS the scatterers can take all orientations in solution. Thus, we assume that the intensity from the entire ensemble is a continuous isotropic function proportional to the scattering from a single particle spherically averaged over all orientations. This is described by Debye's formula:¹⁵⁶

$$\langle \exp(-iQ\zeta) \rangle = \frac{\sin(Q\zeta)}{Q\zeta} \quad (24)$$

Introducing polar coordinates $|Q\zeta| = Q\zeta \cos \theta$, the amplitude equals:

$$A(Q) = 4\pi \int \rho(\zeta) \frac{\sin(Q\zeta)}{Q\zeta} \zeta^2 d\zeta \quad (25)$$

Since the scattering data contains information about the geometry of the sample, the intensity $I(Q)$ is called the form factor $P(Q)$, which is in case of infinitely dilute systems

$P(Q) = |F(Q)F^*(Q)|$, where $F(Q)$ resembles the scattering amplitude.

If we consider molecules in a solvent, adapting random conformations, the density distribution $\rho(\zeta)$ is correctly described by the average excess density $\langle \Delta\rho(\zeta) \rangle = \langle \rho(\zeta) - \rho_s \rangle$, where ρ_s is the solvent density. $\Delta\rho(\zeta)$ is called the contrast. Using this contrast term the form factor can be written as:

$$F(Q) = \left\langle \int_V \Delta\rho(\zeta) e^{-iQ\zeta} d\zeta \right\rangle \quad (26)$$

Thus, if the distance $\zeta = r_1 - r_2$ and accordingly $r_2 = r_1 - \zeta$ between two scatterers is taken into account, the intensity or $|F(Q)|^2$ can be written as:

$$P(Q) = |F(Q)|^2 = F(Q)F^*(Q) = \left\langle \int \int_V \Delta\rho(r_1)\rho(r_1 - \zeta) e^{-iQ\zeta} dr_1 d\zeta \right\rangle \quad (27)$$

with $\gamma(\zeta) = \Delta\rho(\zeta)^2 = \left\langle \int_V \Delta\rho(r_1)\rho(r_1 - \zeta) \right\rangle$ being the spherically averaged autocorrelation¹⁵⁷ or characteristic function¹⁵⁸ of the excess scattering density.

The distance distribution function $p(\zeta)$ corresponds to the distribution of distances between volume elements in the sample and is defined by:

$$p(\zeta) = \gamma(\zeta)\zeta^2 \quad (28)$$

Hence, the scattering curve can be derived from the electron distribution of the sample by combining equation 25 and equation 28:

$$I(Q) = 4\pi \int_0^\infty p(\zeta) \frac{\sin(Q\zeta)}{Q\zeta} d\zeta \quad (29)$$

Consequently, the distance distribution function can be calculated by an inverse Fourier transform of the intensity:¹⁵⁹

$$p(\zeta) = \frac{1}{2\pi^2} \int_0^\infty I(Q) Q\zeta \sin(Q\zeta) dQ \quad (30)$$

In a nutshell, the three-dimensional molecule in real space causes two-dimensional radially averaged scattering data in reciprocal space, which can be transformed into a one-dimensional scattering profile $I(Q)$ *versus* Q . This one-dimensional scattering profile in reciprocal space equals the Fourier transform of the electron distribution of the three-dimensional structure in real space.^{150, 151, 153, 155}

Various common parameters can be defined by SAXS, where different approximations have been proposed. The behavior of the scattering intensity at very small ($Q \rightarrow 0$) and

very large ($Q \rightarrow \infty$) values of Q can be related to overall parameters of the sample. At $Q \rightarrow 0$ the intensity can be described by the Guinier approximation:¹⁴⁹

$$I(Q) = I(0) \exp[-(Q^2 R_G^2)/3] \quad (31)$$

where the Debye factor $\frac{\sin(Q\zeta)}{Q\zeta}$ reduces to $1 - (Q\zeta)^2/3! + \dots$. The radius of gyration R_G and $I(0)$ can be extracted from a plot of $\ln(I(Q))$ against Q^2 , which is called the Guinier plot. The scattering intensity $I(0)$ at zero angle equals the square number of electrons in the scatterer and therefore, is useful for the determination of the molecular weight.¹⁵⁵ The Guinier approximation is only valid for very small angles and arbitrary particle shapes. If the molecules are more elongated, the radius of gyration of the cross-section can be derived. For rather flexible molecules R_G can be calculated using the Debye approximation:¹⁶⁰

$$I(Q)_{average} = \frac{2I(0)}{Q^4 R_G^4} (Q^2 R_G^2 - 1 + e^{Q^2 R_G^2}) \quad (32)$$

where the intensity is averaged over all forms a flexible molecule can acquire with certain probabilities. Finally the radius of gyration can be derived from the distance distribution function $p(\zeta)$:

$$R_G^2 = \frac{\int_0^{D_{max}} p(\zeta) \zeta^2 d\zeta}{\int_0^{D_{max}} p(\zeta) d\zeta} \quad (33)$$

The distance distribution function drops to zero at $\zeta = D_{max}$, where D_{max} is the maximum dimension of the particle.¹⁵⁸ In the high Q regime the scattering intensity decays by Porod's law:¹⁵⁸

$$I(Q) \propto Q^{-4} \quad (34)$$

Porod also defined an invariant B for homogeneous particles in order to determine the surface and volume of the scatterer¹⁵⁸ :

$$B = \int_0^\infty I(Q) Q^2 dQ \quad (35)$$

If equation 35 is combined with the scattering intensity at $Q \rightarrow 0$, the volume can be determined:

$$V = 2\pi^2 \frac{I(0)}{B} \quad (36)$$

and based on the volume V the particle surface is calculated by:

$$\frac{S}{V} = \pi \frac{\lim_{Q \rightarrow \infty} [I(Q)Q^4]}{B} \quad (37)$$

Porod's law assumes uniform density of the scatterer, which is not applicable in the high Q regime due to the contribution of atomic information. A rather general description in the high Q or power law regime can be achieved, if the particles are described in terms of their dimension:

$$I(Q) \propto Q^{-D} \quad (38)$$

where D is the fractal dimension.

In order to account for inter- and intramolecular interactions the description of the scattering intensity has to be extended by the structure factor $S(Q)$:

$$I(Q) = P(Q)S(Q) \quad (39)$$

Form factors $P(Q)$ have been derived for a variety of different geometries.¹⁶¹ The implementation of interactions demands the calculation of the structure factor $S(Q)$. This can be described mathematically as the convolution product of the shape of the particle and the particle distribution and thereby be formulated as the orientational average of the Fourier transform of the spherically averaged autocorrelation function $g(\zeta)$ of the particle distribution or pair correlation function.¹⁶²

$$\langle S(Q) \rangle = S(Q) = 1 + \rho \int_0^\infty 4\pi r^2 (g(\zeta) - 1) \frac{\sin(Q\zeta)}{Q\zeta} d\zeta \quad (40)$$

The pair correlation function $g(\zeta)$ can be calculated using liquid state theory for particles with spherical symmetry interacting with a spherically symmetric potential.^{163,164} The liquid state theory combines the Ornstein-Zernike equation¹⁶⁵ with closure relations in order to relate the interaction potential to the pair correlation function $g(\zeta)$.^{161,163} The total correlation function between two particles can be heuristically described as:¹⁶⁶

$$h(\zeta_{12}) = g(\zeta_{12}) - 1 \quad (41)$$

The Ornstein-Zernike relation introduces the direct correlation function $c(\zeta)$ in terms of the total correlation function $h(\zeta)$:

$$h(\zeta_{12}) = c(\zeta_{12}) + \rho \int d\zeta_3 c(\zeta_{13}) h(\zeta_{23}) \quad (42)$$

where the indirect term $c(\zeta_{13})$ includes the influence of a third particle. In order to solve this equation numerically an approximate closure relation has to be applied to relate

the structure factor to a pair potential. Such closure relations might be integral equations like the mean spherical approximation:^{167,168}

$$\beta v(\zeta) = -c(\zeta) \quad (43)$$

the Percus-Yevick approximation:¹⁶⁹

$$\beta v(\zeta) = \ln[1 - c(\zeta)/g(\zeta)] \quad (44)$$

or the hypernetted chain equation:¹⁶³

$$\beta v(\zeta) = h(\zeta) - c(\zeta) - \ln g(\zeta) \quad (45)$$

where different relations are valid for different model potentials. Mixtures and extensions of closure relations have been shown to give better self-consistent results.^{170–172} These relations are based on the correlation $g(\zeta) \propto \exp[-\beta\omega(\zeta)]$, where $\omega(\zeta)$ is the potential of mean force, which can be simplified by a pair potential $v(\zeta)$ for dilute solutions, where only binary interactions are considered. This simplification is valid in the limit $\rho \rightarrow 0$.¹⁶⁴

1.4.2 Reconstruction of low-resolution structures



Figure 5: Small-angle X-ray scattering experiments provide information about the electron density distribution of a molecule and therefore enable the determination of the shape, the mass and distances (left figure). In order to obtain a more detailed structural information different approaches might be applied possibly leading to the model of an elephant in a snake (right figure). This example just emphasizes the complex process of calculating a unique structure based on X-ray scattering data due to the fact that a variety of different structures may have the same scattering pattern. Figures from: *Le petit prince* (*The little prince*), Antoine de Saint-Exupéry, 1943.

The reconstruction of three-dimensional structures based on SAXS data is limited to low resolutions due to the loss of X-ray phase information as well as the spherical averaging of the intensity of randomly orientated molecules in solution. However, different mathematical methods have been developed in order to reconstruct the shape and possible domains of macromolecules in solution.^{152–154} The spherical averaging legitimates the use of the mathematical apparatus of spherical harmonics, where these form a complete set

of orthogonal functions Y_{lm} on the surface of a unit sphere.^{153,173–175} The underlying basic assumption is that the scattering function does not depend on the orientation of the scatterer, which can therefore be represented including the rotational properties of spherical tensors. Hence, the scattering density distribution $\rho(\zeta)$ of a molecule can be expressed by an expansion as a series of spherical harmonics:

$$\rho(\zeta) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \rho_{lm}(r) Y_{lm}(\omega) \quad (46)$$

with $(\zeta, \omega) = (\zeta, \theta, \varphi)$ being spherical coordinates, Y_{lm} spherical harmonics and ρ_{lm} radial functions:¹⁷³

$$\rho_{lm} = \int_{\omega} \rho(\zeta) Y_{lm}^*(\omega) d\omega \quad (47)$$

The amplitudes can be similarly expressed in reciprocal space:

$$A(Q) = \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}(Q) Y_{lm}(\Omega) \quad (48)$$

and in physical space by a Fourier transform:

$$A(Q) = (2\pi)^{-\frac{2}{3}} \int \rho(\zeta) \exp(iQ\zeta) d^3\zeta \quad (49)$$

The corresponding components of a multipole expansion are connected by Hankel transformations:

$$\rho_{lm}(\zeta) = \sqrt{\frac{2}{\pi}} (-i)^l \int A_{lm}(Q) j_l(Q\zeta) Q^2 dQ \quad (50)$$

$$A_{lm}(Q) = \sqrt{\frac{2}{\pi}} i^l \int \rho_{lm}(\zeta) j_l(Q\zeta) \zeta^2 d\zeta \quad (51)$$

where $j_l(Q\zeta)$ are spherical Bessel functions. Taking into account the definition of the scattering intensity $I(Q) = A(Q)A^*(Q) = |A(Q)|^2$, the intensity can be expressed by:

$$I(Q) = \frac{1}{4\pi} \sum_{l=0}^{\infty} \sum_{m=-l}^l A_{lm}(Q) A_{lm}^*(Q) = \frac{1}{\sqrt{4\pi}} I_{00}(Q) \quad (52)$$

if the scattering intensity is integrated with respect to all orientations of $\rho(\zeta)$. A similar expression for the scattering intensity was derived by Debye:¹⁵⁶

$$I(Q) = (2\pi)^{-3} \int \int \rho(\zeta) \rho(\zeta') \frac{\sin(Q|\zeta - \zeta'|)}{Q|\zeta - \zeta'|} d^3\zeta d^3\zeta' \quad (53)$$

Thus, the scattering intensity is a sum of independent contributions from the substructures corresponding to different spherical harmonics $Y_{lm}(\omega)$, where lower order

harmonics define overall structure parameters and higher order harmonics structural details. Using this mathematical tool kit, it is possible to calculate the scattering function of molecular shapes and structures. However, the obvious shortcoming is that various structures can have the same scattering function making the inverse problem - calculating the unique structure based on the experimental scattering intensity - a great challenge.^{152,153} The two main strategies are the calculation of the shape or molecular envelope^{173,176} on the one hand, and the use of bead models or "dummy atoms"^{177,178} on the other hand, where different minimization algorithms and Monte Carlo procedures can be applied. The former approach is based on the calculation of the angular shape function $F(\omega)$, which describes the shape $\rho(\zeta)$:

$$\rho(\zeta) = \begin{cases} 1 & 0 \leq \zeta < F(\omega) \\ 0 & \zeta \geq F(\omega) \end{cases} \quad (54)$$

where the function $F(\omega)$ can also be developed in a series:

$$F(\omega) = \sum_{l=0}^L \sum_{m=-1}^l f_{lm} Y_{lm}(\omega) \quad (55)$$

with the multipole coefficients

$$f_{lm} = \int_{\omega} F(\omega) Y_{lm}^*(\omega) d\omega \quad (56)$$

where L defines the accuracy of the expansion. If the Bessel function j_l is represented by a Power series, the multipole coefficient of the q^{th} power can be written as:

$$f_{qm} = \int [F(\omega)]^q Y_{lm}^*(\omega) d\omega \quad (57)$$

Hence, the amplitude can be calculated using equation 51 and thereby the scattering intensity can be determined. The *ab initio* envelope determination starts from a spherical shape with the coefficient f_{00} , where the deviations between the experimental and the calculated intensity are minimized.¹⁷⁶ The concept of bead modeling is based on the idea that the maximum dimension of a molecule can be extracted from the scattering intensity, where therefore, the molecule has to fit inside a sphere of this diameter.¹⁷⁷ The maximum diameter D_{max} is the value of ζ at $p(\zeta) = 0$.¹⁵⁵ This sphere is filled with M densely packed beads, where every bead may either belong to the particle (index = 1) or to the solvent (index = 0) and the shape can be completely described by a binary string X of the length M. Starting from a random distribution of ones and zeros, the model is gradually modified using a Monte Carlo algorithm until it fits best the experimental data. Typically, the intensity of the model is calculated using Debye's formula for spheres¹⁵⁶ as shown in equation 53. The Monte Carlo procedure might be either a genetic algorithm¹⁷⁷ or simulated annealing.^{178,179} Apparently, a shape reconstruction based on bead models

yield multiple solutions, which have to be consequently compared and evaluated.

1.4.3 Instrumentation

A SAXS instrument is basically made of an X-ray source, an optical system, which selects appropriate wavelengths and focuses the beam, a sample stage in a vacuum chamber, a beamstop, which blanks the initial unscattered beam, and a two-dimensional detector that detects the intensity of the scattered X-rays.¹⁵³ A schematic representation of such a set-up is shown in figure 6. Different X-ray sources as well as optical systems are available.¹⁵³

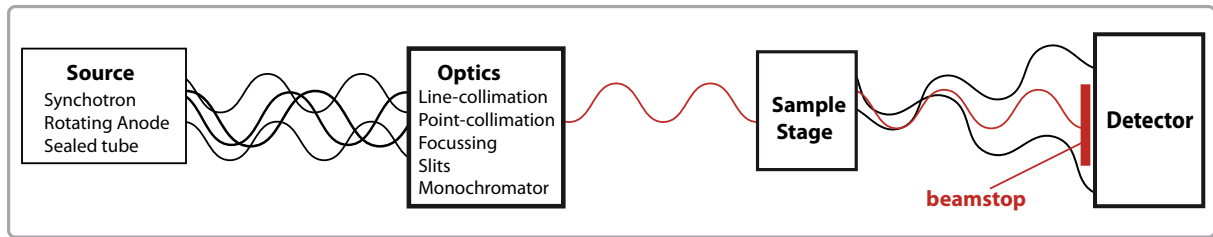


Figure 6: Schematic representation of a small-angle X-ray scattering instrument. An X-ray source, such as a synchrotron, a rotating anode or a sealed tube, produces X-rays of different wavelengths. An optical system selects appropriate wavelengths and focuses the beam on the detector plane. The sample is placed on a sample stage in a vacuum chamber, which can be moved in x- and y-direction. The unscattered initial beam (marked in red) is blanked by the beamstop. The scattered X-rays are detected by a two-dimensional detector.



Figure 7: Small-angle X-ray scattering set-up that was used for the investigation of S-layers in this work. SAXS was performed with Cu K α radiation from a rotating anode generator equipped with a pinhole camera and an area detector. The SAXS measurements were performed at the Institute of Dynamics of Condensed Systems, Faculty of Physics, University of Vienna, Austria under the supervision of Herwig Peterlik.

All SAXS experiments conducted in this work were done at the Institute of Dynamics

of Condensed Systems, Faculty of Physics, University of Vienna, Austria using a laboratory-based SAXS equipment, which is shown in figure 7. SAXS was performed with Cu Ka radiation from a rotating anode generator (Nanostar, Bruker AXS, Germany) equipped with a pinhole camera and an area detector (Vantec 2000, Bruker AXS, Germany). The samples were put into capillaries with 1 mm diameter and 10 μm wall thickness (Hilgenberg) and sealed with polymeric caps. The intensity patterns were taken at a sample to detector distance of 109 cm for 6 h. They were corrected for background scattering and then radially averaged. Finally, the scattering intensity from the protein-free solution was subtracted.

1.5 Self-assembly in a fractal space: Toward an understanding of S-layers in solution

1.5.1 The fractal geometry

Fractal geometry describes the invariance of a structure under a change of scale.^{180,181} This mathematical tool can be used in order to describe irregular structures that lack any "conventional" geometry such as the shape of a cloud, coastlines or generally disordered systems and it has been used to address a variety of physical questions.^{182–184} Fractal geometry provides a rule to extract regularity out of irregularity, namely the fractal dimension, which represents the degree of complexity.¹⁸⁰ The fractal or Hausdorff dimension D is defined as:

$$m(\zeta) = \zeta^D \quad (58)$$

where ζ is some arbitrary self-similar length scale and m is the mass. Therefore the density of a self-similar or fractal object can be described by a power law:

$$\rho(\zeta) \sim \frac{m}{\zeta^3} \sim \zeta^{3-D} \quad (59)$$

From a geometrical point of view proteins are highly irregular structures and Euclidian geometry is a rather poor tool for the description of their tertiary structure.¹⁸⁵ Fractal geometry provides one possibility to analyze and describe the irregularity of protein structures.^{186–188} Hence, a fractal approach can be used in order to investigate irregular structures by X-ray scattering. In this way, aggregation processes, polymer structures and deposition phenomena have been successfully described by fractal geometry,^{189–193} where the X-ray scattering of such systems can be analyzed by a fractal form factor.¹⁹⁴

1.5.2 Scattering of S-layer proteins: Theoretical scaffold

S-layer proteins have the intrinsic property to self-assemble in solution within particularly short timescales. On the one hand, this crystallization behavior opens up the path for a description of S-layers by a fractal approach, while on the other hand, the investigation of S-layer monomers by conventional low-resolution reconstruction procedures as described in chapter 1.4.2 becomes rather complicated. The self-assembly process affects the scattering pattern in the high Q -regime in form of Bragg reflections and in the low Q -regime by the arrangement of the proteins in form of two-dimensional S-layers. Thus, the scattering pattern has to be split into two regimes: one at high Q values and one at low Q values, which are examined separately. In order to analyze the scattering signal produced by S-layers and to reconstruct the structure, the system is described by a fractal geometry. The intensity, as described in equation 27, is the Fourier transform of the autocorrelation function of the system, which is in case of a fractal system the mass correlation (see

equation 59):

$$I(Q) = \mathcal{F}(\langle \rho(\zeta') \rho(|\zeta' - \zeta|) \rangle)[Q] = (\zeta^{3-D})[Q] \quad (60)$$

with ζ being a coordinate and D the fractal dimension. If the fractal autocorrelation function is substituted into equation 29 $I(Q) = \int_0^\infty 4\pi \zeta^2 \frac{\sin(Q\zeta)}{Q\zeta} \zeta^{3-D} d\zeta$, the intensity of the fractal system is proportional to:¹⁸⁴

$$I(Q) \sim Q^{-D} \quad (61)$$

Thus, the fractal dimension D of the system is accessible in a double-logarithmic plot of $I(Q)$ *versus* Q . The specific intra- and intermolecular interactions of S-layer proteins in solution have to be incorporated into the analysis of the scattering function, where due to the fact that a fractal geometry is assumed, the underlying potential also has to depend on the dimensionality of the system, which by the way, is the same for Euclidean geometries. Hence, the sensitivity to geometrical constraints has to be captured in the potential function. The Poisson equation describes the potential for a given charge distribution and therefore, has to be obeyed by a system of interacting proteins in solution. A fractal theory for the solution of the Poisson equation has been already successfully derived.^{195, 196} A linearized Poisson-Boltzmann approach $g(\zeta) - 1 \simeq -\beta\Phi(\zeta)$ and a following coordinate transformation yield a potential containing Bessel functions of the second kind K_ν . However, a shortcoming of this method is the nonlinearity, which makes an analytical Fourier transformation impossible. Another, more direct approach is to solve the Poisson equation by the use of Green's function, if additionally, the known relation between the potential of mean force $\omega(\zeta)$ of the system and local density distributions $\rho(\zeta)$ is taken into account. This relation is based on the reversible work theorem^{112, 116, 197, 198} and Jarzynski's formulation of the equality between free energy changes and the work:^{119, 120}

$$-\beta\omega(\zeta) \approx \ln(\rho(\zeta)) \quad (62)$$

Green's function is a potential of a charge distribution that looks like a Delta function. Delta functions represent point charges and in our case, rotationally symmetric point scatterers constitute the potential:

$$G_D \sim (\zeta, \zeta') = \delta(|\zeta - \zeta'|) \quad (63)$$

Hence, the Poisson equation can be written as:

$$\Delta^D \omega(\zeta) = \langle (\delta(\zeta') \delta(|\zeta' - \zeta|)) \rangle \quad (64)$$

The rotationally symmetric electronic density distribution $\rho(|\zeta - \zeta'|)$ is represented by autocorrelated point scatterers $\langle (\delta(\zeta') \delta(|\zeta' - \zeta|)) \rangle$ by $g(\zeta) = \exp(-\beta\omega(|\zeta' - \zeta|))$. Thus,

the solution of the Poisson equation equals:

$$\omega(\zeta) = \int G_D(\zeta') \langle (\delta(\zeta') \delta(|\zeta' - \zeta|)) \rangle d^D \zeta' \quad (65)$$

The fractal Fourier transform yields the solution for the mean potential, with $G_D(u) = \mathcal{F}_u^{-1}(\hat{G}(u))$ and thus $\hat{G} = \frac{1}{u^2 + s^2}$.¹⁹⁹

$$\mathcal{F}(\omega(\zeta))[Q] \approx \mathcal{F}\left(\int G_D(\zeta') \langle (\delta(\zeta') \delta(|\zeta' - \zeta|)) \rangle [Q] d^D \zeta'\right) \quad (66)$$

$$= \mathcal{F}(\langle G_D(\zeta) \delta(\zeta) \rangle [Q]) \quad (67)$$

$$= \frac{1}{(2\pi)^D} \int_0^\infty du^D \hat{G}(u) \mathcal{F}_\zeta(\delta(\zeta - u)) [Q] \quad (68)$$

$$= \frac{1}{(2\pi)^D} \int_0^\infty du u^{D-1} \frac{1}{u^2 + \kappa^2} \frac{J_{D/2-1}(Qu)}{(Qu)^{D/2-1}} \quad (69)$$

$$= \frac{1}{(2\pi)^D} \kappa^{D/2-1} \frac{K_{D/2-1}(Q\kappa)}{(Q\kappa)^{D/2-1}} \quad (70)$$

Equation 70 represents the analytical form factor $P^D[Q]$ of the proteins, where J are Bessel functions of the first kind. In a three-dimensional space the form factor simplifies to:

$$P^3[Q] = \frac{1}{(2\pi)^3} \kappa^{1/2} \frac{K_{1/2}(Q\kappa)}{(Q\kappa)^{1/2}} \quad (71)$$

Numerical reconstructions of the scattering intensities P_{num} can be done by using the Debye equation¹⁵⁶ as shown in chapter 1.4.2. Equation 70 represents an extension of the reconstruction process derived by Debye, where the fractal dimension is put to $D = 3$, the weight factor $\frac{1}{u^2 + \kappa^2}$ is put to 1 and the integral is replaced by a sum.

The fractal description of the form factor takes into account inter- and intramolecular interactions, implements a mean potential that is constraint to a fractal geometry and assumes point scatterers rather than spherical or elliptical form factors as normally used for a structure reconstruction process. However, the self-assembly of S-layer proteins into two-dimensional sheets demands an additional structure factor, which accounts for the interactions between these layers. One possible description of two-dimensional S-layers is the form factor for a disk with a specific height H_{\gg} and radius R_{\gg} ,¹⁴⁹ where the symbol \gg indicates that these dimensions are much larger than the protein monomers. To be consistent, the disk-like structure and the interactions between the disks are also described by a fractal geometry. Hence, the disks are described as fractal spaces that are orthogonal in respect to each other. The volume element of a D -dimensional space equals:²⁰⁰

$$d\tau_{\perp} = \zeta^{D_{\perp}-1} \sin^{D_{\perp}-2} \theta \sin^{D_{\perp}-3}(\Phi_1) \dots \sin(\Phi_{D_{\perp}-3}) d\zeta d\theta d\Phi_i^{D_{\perp}-2} d\Phi_i \quad (72)$$

The parallel fractal complement is:

$$d\tau_{||} = \zeta^{D_{||}-1} \cos^{D_{||}-2} \theta \sin^{D_{||}-3}(\Phi_1) \dots \sin(\Phi_{D_{||}-3}) d\zeta d\theta \Pi_i^{D_{||}-2} d\Phi_i \quad (73)$$

The disk-like S-layers can be reconstructed by a product of a Delta function $\delta(\zeta_{\perp}^2 - H_{\gg}^2)$ and a Heaviside function $\Theta(R_{\gg} - \zeta_{||})$:

$$\rho_{\gg} = \delta(\zeta_{\perp}^2 - H_{\gg}^2) \Theta(R_{\gg} - \zeta_{||}) \quad (74)$$

with $\zeta_{\perp} = \zeta \sin \theta$ and $\zeta_{||} = \zeta \cos \theta$. ρ_{\gg} are the local densities, where the Fourier transformation of ρ_{\gg} gives the structure factor S_{\gg} for the interactions of S-layer sheets in solution:

$$S_{\gg}[Q] = \mathcal{F}(\rho_{\gg}(\zeta))[Q] \quad (75)$$

$$= \mathcal{F}_{\perp}(\delta(\zeta_{\perp}^2 - H_{\gg}^2))[Q] \mathcal{F}_{||}(\Theta(R_{\gg} - \zeta_{||}))[Q] \quad (76)$$

$$= C_{\Gamma,D} \int_0^{\frac{\pi}{2}} \pi(\theta^*) \times \frac{J_{D_{\perp}/2-1}(QH_{\gg} \cos \theta^*)}{(QH_{\gg} \cos \theta^*)^{D_{\perp}/2-1}} \quad (77)$$

$$\times \frac{J_{D_{||}/2}(QR_{\gg} \sin \theta^*)}{(QR_{\gg} \sin \theta^*)^{D_{||}/2}} d(-\cos \theta^*) \quad (78)$$

with $C_{\Gamma,D}$ being a normalization constant. In order to solve the integral, two different representations of the Bessel functions J are applied:

$$J_v(\zeta) = \pi^{-\frac{1}{2}} \Gamma^{-1}(v + \frac{1}{2}) \left(\frac{\zeta}{2}\right)^v \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(\zeta \sin u) \cos^{2v} u du \quad (79)$$

$$J_v(\zeta) = \pi^{-\frac{1}{2}} \Gamma^{-1}\left(\frac{\zeta}{2}\right)^v \int_0^{\pi} \sin(\zeta \sin u) \cos^{2v} u du \quad (80)$$

Thus, the structure factor S_{\gg} describes S-layers as disks with a height H_{\gg} and a radius R_{\gg} that are arranged with an angle distribution $\pi(\theta)$ with respect to each other. Based on this theoretical background and the fact that mean potentials are independent and additive and thus superpose, the scattering signal, produced by interacting S-layer proteins in solution, is described by:

$$I(Q) = c_1 P^D(Q) \times S_{>}(Q) \times S_{\gg}(Q) + c_2 \quad (81)$$

where c_2 is a small constant from a parasitic background. The form factor P^D describes the structure of S-layer monomers, the structure factor $S_{>}$ accounts for the interactions

between the individual proteins and the structure factor S_{\gg} specifies the large-scale behavior of self-assembled S-layers:

$$P^D[Q] = \sum_i \frac{J_{D/2-1}(Qu_i)}{(Qu_i)^{D/2-1}} \quad (82)$$

$$S_{>}[Q] = \left(\frac{1}{2\pi}\right)^D \int d^D u u^{D-1} \hat{G}_{u,\kappa\rho_{>}}(u) \frac{J_{D/2-1}(Qu)}{(Qu)^{D/2-1}} \quad (83)$$

$$S_{\gg}[Q] = \left(\frac{1}{2\pi}\right)^D \int d^D R_{\gg} R_{\gg}^{D-1} \hat{G}_{R_{\gg},\kappa\mathcal{F}(\rho_{\gg}(\zeta))}[Q] \quad (84)$$

This fractal theoretical scaffold allows the analysis of small-angle X-ray scattering of S-layer proteins in solution that self-assemble into two-dimensional layers. The interactions between the proteins are taken into account by relating a fractal local density distribution to a fractal mean potential that fulfills the Poisson equation. Thus, a low-resolution reconstruction of the protein shape as well as the disk-like shape becomes possible.

1.5.3 Scattering of S-layer proteins: Refinement of the structure

In order to reconstruct the shape of the S-layer protein SbsB two different samples were investigated: one sample containing mainly S-layer monomers and a second sample containing self-assembled two-dimensional sheets. Both samples were analyzed the same way. The fractal dimension of the monomer equals $D = 2.4$ and of the self-assemblies $D = 2.9$, respectively. The SAXS intensity data was fitted analytically as well as numerically based on the theoretical background as described in chapter 1.5.2. The resulting fitting curves are shown in figure 8. The reconstruction of the shape was done by generating scattering patterns using a numerical algorithm. In order to restrict the problem, the fitting procedure was closely connected to the structural model as calculated by molecular dynamics simulations (see chapter 1.3), which serves as starting configuration for the calculation of electron densities. Based on a sphere that includes the entire atomistic model, a numerical fitting algorithm changes position of point scatterers with the restriction that their position is within a size limit of 2 nm with respect to the model. The resulting electron densities are shown in figure 9. Thus, the simulated model could be refined and the shape verified by the SAXS data, where it served as a permanent corrective of the numerical reconstruction procedure.

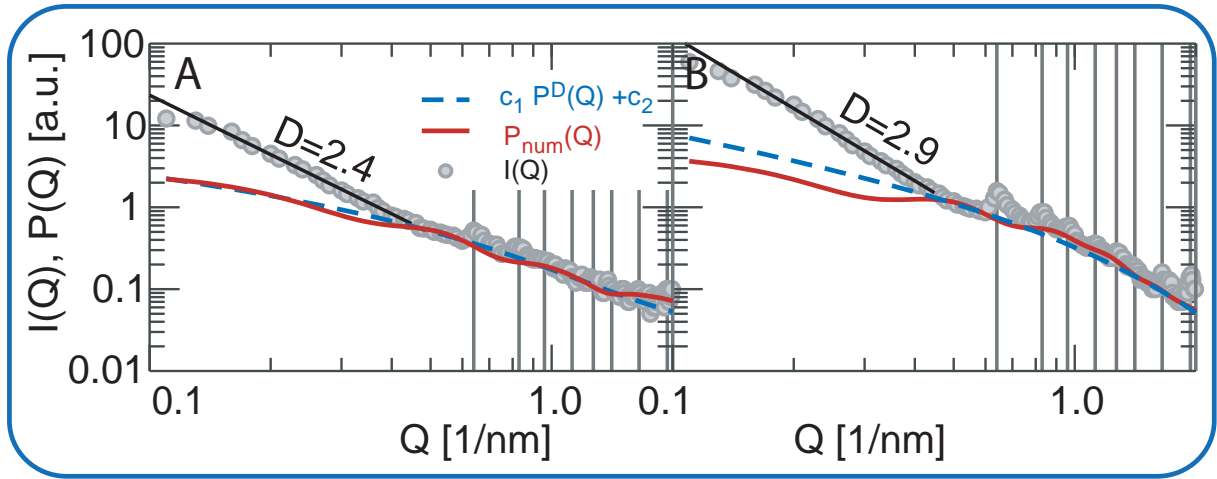


Figure 8: SAXS data of (A) the monomeric solution and (B) the self-assemblies are given by gray filled circles and slopes for the fractal dimension are inserted as black lines: (A) $D=2.4$ and (B) $D=2.9$. The dashed blue line is the analytical fit P^D for (A) $P^{2.4}$ and (B) $P^{2.9}$. The full red line is the complete numerical reconstruction using a Monte Carlo algorithm generalized to fractal dimensions.

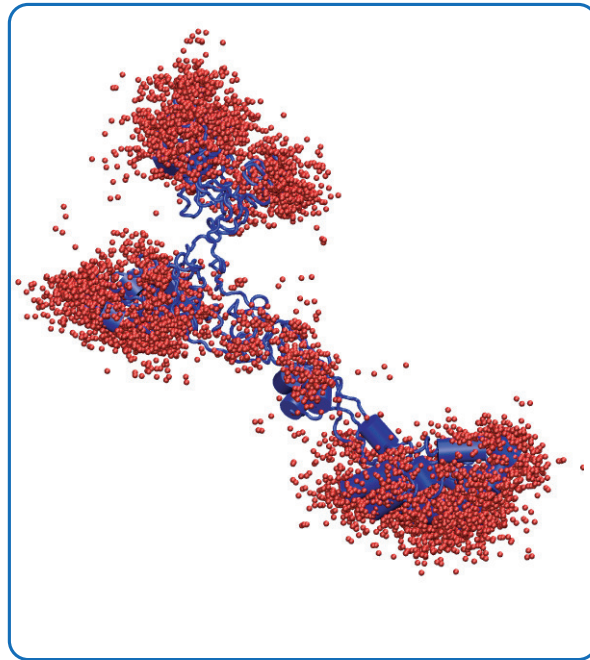


Figure 9: Numerical reconstruction of the SAXS scattering intensities of the monomeric sample. High electron densities are indicated by red beads. The structural model as calculated by molecular dynamics simulations is shown in blue.

1.6 AFM-based single-molecule force spectroscopy

1.6.1 Basic principles

The atomic force microscope offers the opportunity to manipulate proteins at the single-molecule level by applying mechanical forces onto individual molecules and consequently stretch and mechanically unfold them. The force required to unfold single protein domains is highly dependent on the topology of the fold and thus, gives an insight into the structural and mechanical architecture of the protein.^{201–204} Figure 10 A shows schematically the principle of an atomic force microscope in the force spectroscopy mode. The sample, containing single molecules in solution, is placed on a piezoelectric scanner. A sharp tip, which is mounted on a flexible cantilever, is brought into contact with the sample surface, where single proteins are either adsorbed on the tip or attached by the formation of specific bonds. When the distance between the tip and the sample surface is increased, the single molecule extends, which generates a restoring force that causes the cantilever to bend. A laser beam that is directed toward the cantilever, changes its deflection due to the bending of the cantilever, which is detected by a photodiode. If the elastic properties - the force constant - of the cantilever are known, the output of the photodiode can be related to the applied force. The mechanical unfolding of multiple protein domains (figure 10 B) at constant velocity typically results in a saw-tooth like force-extension pattern as shown in figure 10 C, where the number of peaks corresponds to the number of unfolded structural units that exhibit mechanical stability. The elastic behavior of proteins can be modeled by entropic elasticity models that describe the entropy reduction accompanying the force-induced extension of polymer chains. Among the most prominent models describing such force-extension behavior is the worm-like chain model (WLC),²⁰⁵ which has been extensively used in order to describe the mechanical unfolding of proteins:

$$F(x) = \frac{k_B T}{p} \left[\frac{1}{4} \left(1 - \frac{x}{L_c} \right)^2 - \frac{1}{4} + \frac{x}{L_c} \right] \quad (85)$$

where L_c is the contour length, which represents the length of the polymer at the maximal physically possible extension, and p the persistence length, which quantifies the stiffness of a polymer. The smaller the persistence length the greater the entropy of the polymer and the greater the resistance to mechanical extension.²⁰³ Thus, such models can be used to describe the saw-tooth like unfolding pattern and allow the calculation of the specific contour length L_c of unfolded structural units, which can be related to the protein structure.

In order to determine the specific unfolding peaks that are related to the protein under investigation, polyprotein engineering offers the opportunity of an internal standard for the experiment and avoids the misinterpretation of force peaks that might arise due to non-specific interactions between the tip and the sample surface. Multiple copies of protein domains are linked at the DNA-level and expressed together as polyproteins (figure 10 B).

Such methodology was inspired by the first AFM-based mechanical unfolding experiments using naturally occurring polyproteins, namely different domains of the muscle protein titin and the extracellular protein tenascin.^{206–210}

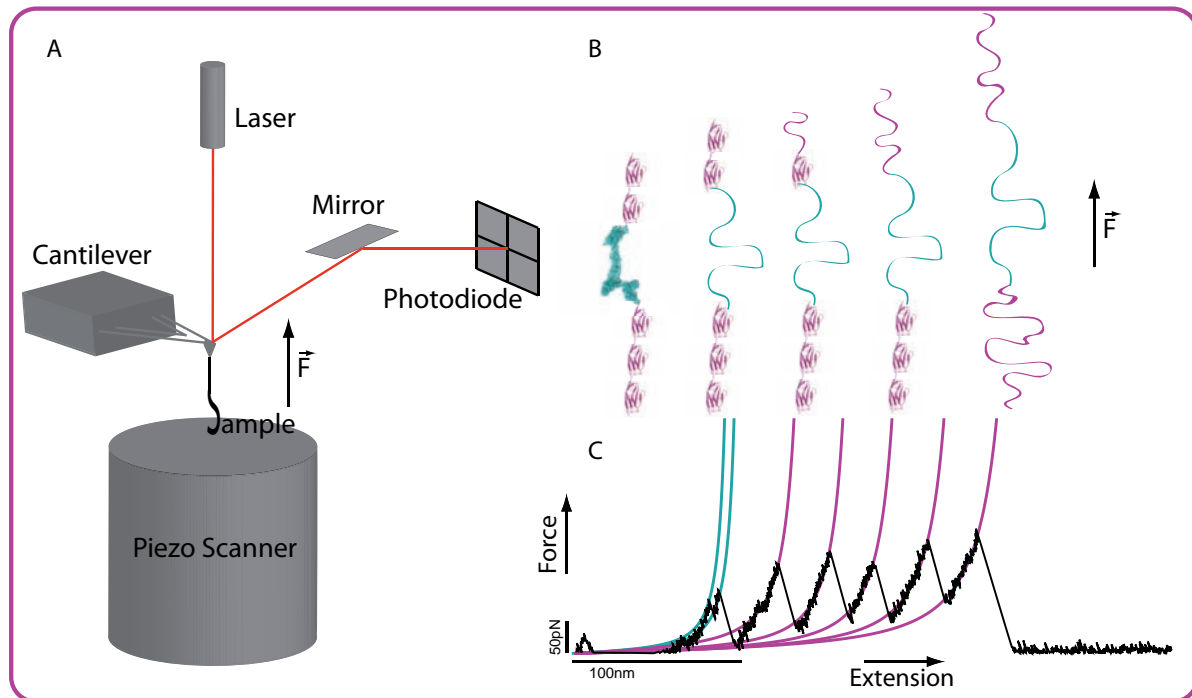


Figure 10: (A) Schematic of the principle of an atomic force microscope. A sharp tip is mounted at the end of a flexible cantilever. Forces acting between the tip and the sample cause a deflection of the cantilever, which is detected by a laser beam. A piezo scanner enables a movement in x-, y- and z-direction. (B) Principle of AFM-based single-molecule force spectroscopy. A protein construct - here an S-layer protein embedded within five titin I27 domains is exemplarily shown - is stretched through the adsorption of one terminus to the AFM tip and consequent pulling. By exerting mechanical force the polyprotein is stochastically unfolded resulting in force-extension curves as shown in (C). The mechanical unfolding of proteins or protein domains at constant velocity causes a saw-tooth like force-extension pattern revealing a wealth of information regarding the structure, mechanical stability and unfolding pathways, where the number of peaks corresponds to the number of unfolded domains. Such characteristic response to mechanical force can be described by different entropic elasticity models, where here the worm-like chain model has been applied.²⁰⁵

The mechanical unfolding pattern sheds light on the structural composition and the mechanical stability of a protein. The magnitude of the required unfolding forces allows for the analysis of possible structural elements, since β -sheets are known to bear higher mechanical stress than α -helices or coils.^{206,207,210–215} Moreover, folding and unfolding pathways can be described by the identification of structural intermediates, which helps to better understand the energy landscape of the respective protein - an important tool on the road to solve the protein folding problem.^{77,208,216–226} There are two basic protocols of single-molecule force spectroscopy experiments. The constant-velocity protocol provides control over the end-to-end distance and is therefore well applicable for the determination of structural intermediates enabling a very accurate

length resolution. Force-clamp protocols use feedback loops in order to control the applied force and thus, are better qualified for kinetic studies associated with protein folding/unfolding and chemical reactions.^{216,227,228} The extraction of thermodynamic and kinetic data is a highly complex issue due to the fact that statistical mechanics can be only hardly applied to experiments involving only one single molecule.^{229,230} Different approaches, such as Jarzynski’s equality concerning the calculation of the potential of mean force^{121,231–234} or Crook’s fluctuation theorem^{235,236} have been proposed in order to calculate thermodynamic parameters and to relate nonequilibrium experiments at the single-molecule level to bulk experiments involving millions of molecules. However, force influences the energy landscape of a molecule²²⁹ and single-molecule force spectroscopy might not be the method of choice for thermodynamic questions. In order to address the kinetics of single-molecule force spectroscopy experiments, a modified model of Arrhenius’ equation for the temperature dependence of the rate of a chemical reaction²³⁷ has been firstly proposed by Bell.²³⁸

$$k(F) = k_0 \exp \left(\frac{F \Delta x}{k_B T} \right) \quad (86)$$

with $k(F)$ being the reaction rate as a function of the force F and Δx the distance to the transition state beyond which the bond will fail. Such single-exponential dependence has been successfully used in order to describe the kinetics in different single-molecule force spectroscopy experiments,^{239–242} where the approach based on Arrhenius’ equation could be also generalized.^{243–245} Only recently, non-exponential models were applied in order to describe the multiple reaction pathways of a protein.²⁴⁶ To sum it up, single-molecule force spectroscopy experiments can be applied to better understand the force response of proteins, especially for molecules that play a role in biomolecular mechanics, functional and structural architectures and energy landscapes of protein folding and unfolding.

1.6.2 Mechanical stability can be regulated by ligand binding

Force spectroscopy offers the possibility to study functional and structural consequences of the binding of specific ligands to proteins at the single-molecule level. Moreover, the crucial role mechanical force might play for mechano-sensitive proteins in order to provide conformational states that enable the binding of a ligand can be studied. In most cases the binding of a ligand enhances the mechanical stability, where this is not necessarily related to a change of the protein structure.^{247–252} The binding might take place at locations far away from those regions that are critical for mechanical stability. However, for mechano-sensitive proteins, which are able to adopt different conformations when exposed to mechanical stress, ligand-binding can consequently alter the unfolding pathway due to a change of the structure, which can be observed by a different force-extension pattern.^{253–257}

1.6.3 S-layers: May the force be with them!

S-layer proteins perform their intrinsic function under a constant mechanical stress due to the fact that they completely cover whole cells during all stages of the cell growth and division cycle.^{258,259} Their mechanical stability has been successfully demonstrated by using force spectroscopy, where single proteins were unzipped out of the layer. Unfolding forces ranging from about 50 pN up to 300 pN could be measured.^{260–262} However, in this work, for the first time, the mechanical unfolding of one individual S-layer protein has been investigated, where a polyprotein approach as shown in figure 11 B was applied. The known mechanical features of the I27 domain of the muscle protein titin were used in order to provide a fingerprint for the unfolding events of a single S-layer monomer.^{206–208,211,220} Two independent experiments were performed. First, the polyprotein was mechanically unfolded in order to gain insight into the unfolding pathway, mechanical stability and structural architecture of one S-layer protein. Second, secondary cell wall polymer (SCWP) was added and the polyprotein, bound to SCWP, was again mechanically unfolded under the same conditions. All experiments were performed using the JPK NanoWizard I atomic force microscope as shown in figure 11 A in a constant-velocity mode at 400 nm/s.

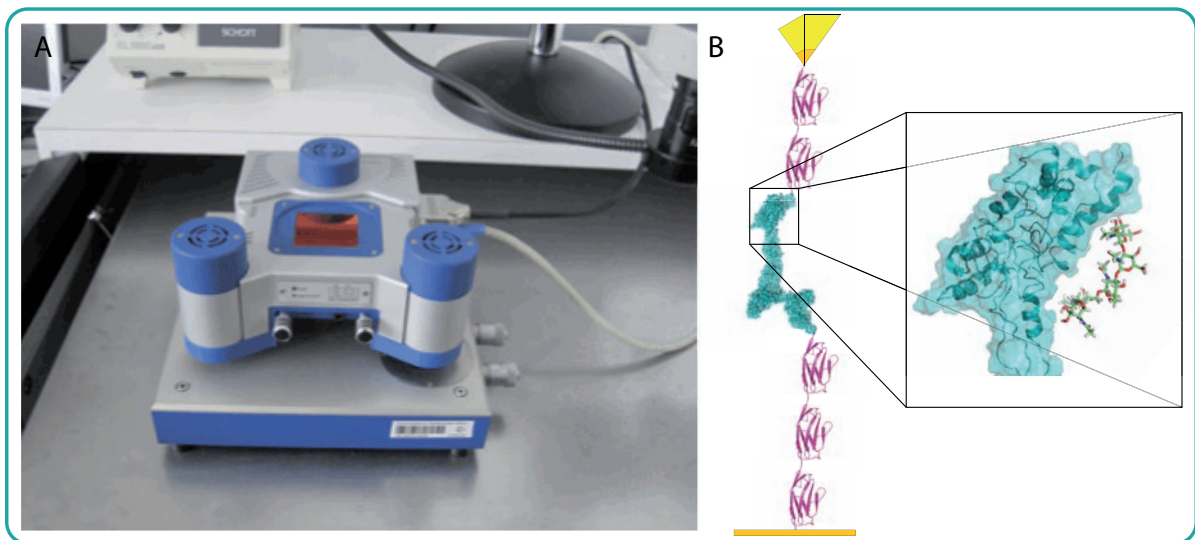


Figure 11: (A) JPK Nano Wizard I set-up. This atomic force microscope was used for all single-molecule force spectroscopy experiments presented in this work. (B) Polyprotein ($I27$)₂ – *SbsB* – ($I27$)₃ set-up: the S-layer protein *SbsB* is embedded within five I27 domains of the human muscle protein titin. The enlarged view shows the proposed interaction site between the N-terminal region of the S-layer protein and a secondary cell wall polymer (SCWP). SCWPs represent the specific binding partners for S-layer proteins within the peptidoglycan layer of the cell wall and anchor the proteins to the cell surface. Experiments were conducted independently using first the polyprotein alone and second, the protein bound to SCWP, in a constant-velocity mode.

Figure 12 shows exemplarily two typical force-extension traces for the unbound protein (figure 12 A) and the protein bound to SCWP (figure 12 B). The S-layer protein is

mechanically unfolded through distinct well defined intermediates revealed by a double force peak, where the unfolding force is about 90 pN for the unbound and about 110 pN for the ligand-bound protein. This data indicates that the mechanical stability increases through the binding of SCWP. Interestingly, a different unfolding pathway is observed when the S-layer protein interacts with SCWP, which is assumed due to the appearance of a third peak corresponding to SbsB (figure 12 B, blue star). This is a direct indication for a conformational change of the protein induced by the binding of SCWP.

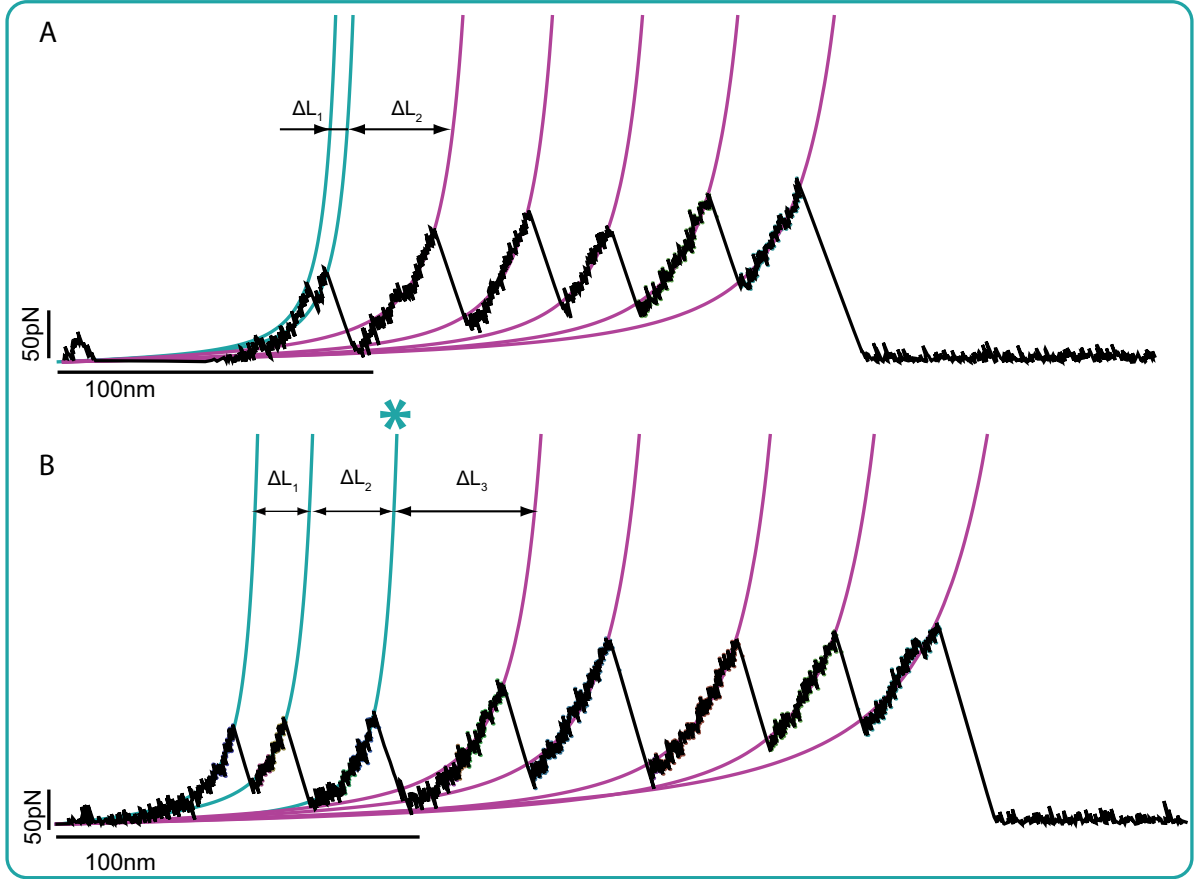


Figure 12: Typical force-extension traces for the unfolding of the polyprotein $(I27)_2 - SbsB - (I27)_3$ for the unbound protein (A) and the protein bound to its ligand SCWP (B). (A) The force-extension trace shows five consecutive unfolding peaks for the five I27 domains (WLC fit in purple) and one double peak for the S-layer protein SbsB (WLC fit in blue). The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 12.1 \pm 4.9$ nm and between the second peak of the double peak and the first I27 peak $\Delta L_2 = 36.3 \pm 5.2$ nm. The mean unfolding force of the first peak is 86.8 ± 12.5 pN and 87.6 ± 10.8 pN, respectively ($n = 90$). The mean contour length increment between the consecutive I27 domains is 28.14 ± 2.2 nm and the mean unfolding force 197.89 ± 16.2 pN for $n=400$. (B) An additional peak is observed when the S-layer protein is bound to SCWP (marked by a blue star). The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 17.3 \pm 4.2$ nm, between the second peak of the double peak and the third peak $\Delta L_2 = 44.9 \pm 6.5$ nm and between the third peak and the first I27 peak $\Delta L_3 = 32.8 \pm 7.2$ nm. The unfolding forces are higher than for the unbound protein, namely $F_1 = 110.1 \pm 9.1$ pN, $F_2 = 109.9 \pm 9.8$ pN and $F_3 = 117.4 \pm 9.7$ pN.

It was recently suggested that some S-layer proteins might be able to adopt different

conformations depending if they are in a monomeric or polymeric state.^{41–43} Apparently, the interaction with SCWP also causes a structural change, which might be an interesting finding for the understanding of the translocation of S-layer proteins through the peptidoglycan-containing rigid cell wall matrix, which has to occur in a rather unfolded conformation. Potentially, the binding of SCWP induces a crucial conformational change that marks an essential step for the proteins on their way to the cell surface, where they self-assemble into S-layers. The lack of an initial extension might be related to the formation of highly stable, compact and mechanically resilient conformations. The same phenomenon has been previously observed for the mechanical unfolding of polyQ chains, where it was suggested that this ensemble of mechanically resilient conformations might play a role in the assembly process of polyQ chains.²⁶³ In a nutshell, the single-molecule force spectroscopy experiments, shown in this work, demonstrate the mechanical unfolding through distinct well-defined intermediates and a mechanical stability comparable with other extracellular proteins, which identifies S-layers as mechanically stable protein layers able to bear the mechanical stress that prokaryotic cells permanently experience in nature. Apparently, an ensemble of mechanically resilient conformations is responsible for the lack of initial extension of the polyprotein. These experiments provide the first demonstration of a change of the mechanical unfolding pathway due to the binding of a specific ligand prior to a force-induced extension of the protein, which serves as a direct indication of a conformational change of the protein. Finally, an increase in mechanical stability is observed when the individual proteins are bound to SCWPs. The observed conformational change of an S-layer protein together with experimental findings related to the behavior of S-layers within the peptidoglycan layer,²⁶⁴ provide a possible explanation of how the complex translocation of individual proteins through the cell wall might be controlled.

1.7 The structure of an SbpA unit cell

The unit cell of the S-layer protein SbpA consists of four monomers as described in chapter 1.1.2. In order to model the three-dimensional structure at an amino acid level of one SbpA unit cell, a combined approach of molecular dynamics simulations, small-angle X-ray scattering and transmission electron microscopy was used. Following the same road as for SbsB, the tertiary structure of an SbpA monomer was initially modeled using *de novo* structure prediction and molecular dynamics (see chapter 1.3). The resulting intermediate structure is shown in figure 13. The calculation of a whole unit cell demands a coarse-graining of the protein due to its tremendous size (figure 13 right figure), where every amino acid is represented by a single bead placed in the center of mass. Consecutive beads are linked by harmonic potentials.

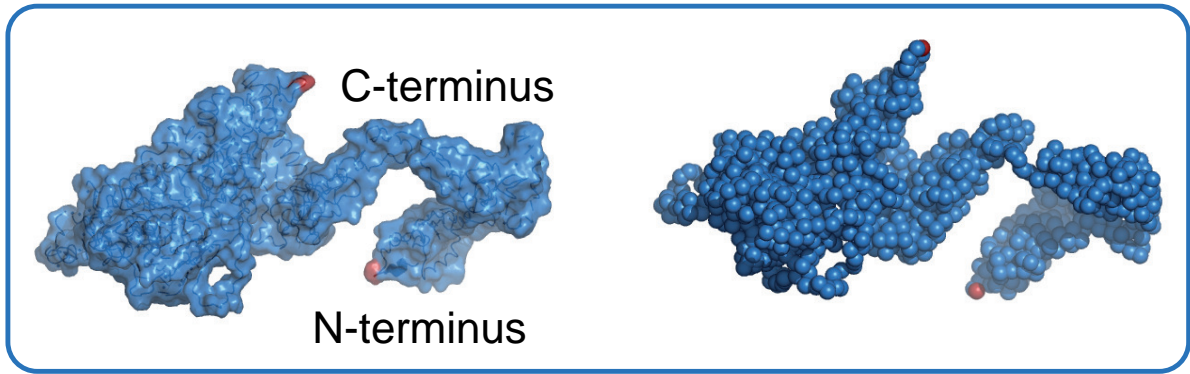


Figure 13: Atomistic and coarse-grained structure of the S-layer protein SbpA as obtained by molecular dynamics simulations. The termini are marked in red. In the coarse-grained model every amino acid is represented by a single bead, where consecutive beads are linked by an harmonic potential.

In contrast to the structure prediction approach used for the S-layer protein SbsB (see chapter 1.3), three-dimensional density distribution data were available for SbpA. These experiments were performed by Harald Gollner²⁶⁵ using a transmission electron microscope. Therefore, tilting studies of negatively stained self-assembly products of SbpA were conducted resulting in single sections of the whole unit cell. An inverse Fourier transform of these sections and a proximate superposition yielded a density distribution. The respective sections and an electron micrograph of the unit cell are shown in figure 14. Consequently, the coarse-grained protein model was fit into the three-dimensional density model by arranging four monomers in a way that the whole unit cell is covered and overlaps are avoided. This intermediate unit cell was then minimized by a cooling process ranging from 300 K down to 30 K, where the beads interact *via* attractive screened Coulomb and associative Gaussian pair-pair potentials. Finally, the resulting structural model was used to refine the small-angle X-ray scattering signal produced by SbpA self-

assemblies in solution as described in chapters 1.5.2 and 1.5.3, where slight modifications were introduced regarding the description of two-dimensional layers in solution.

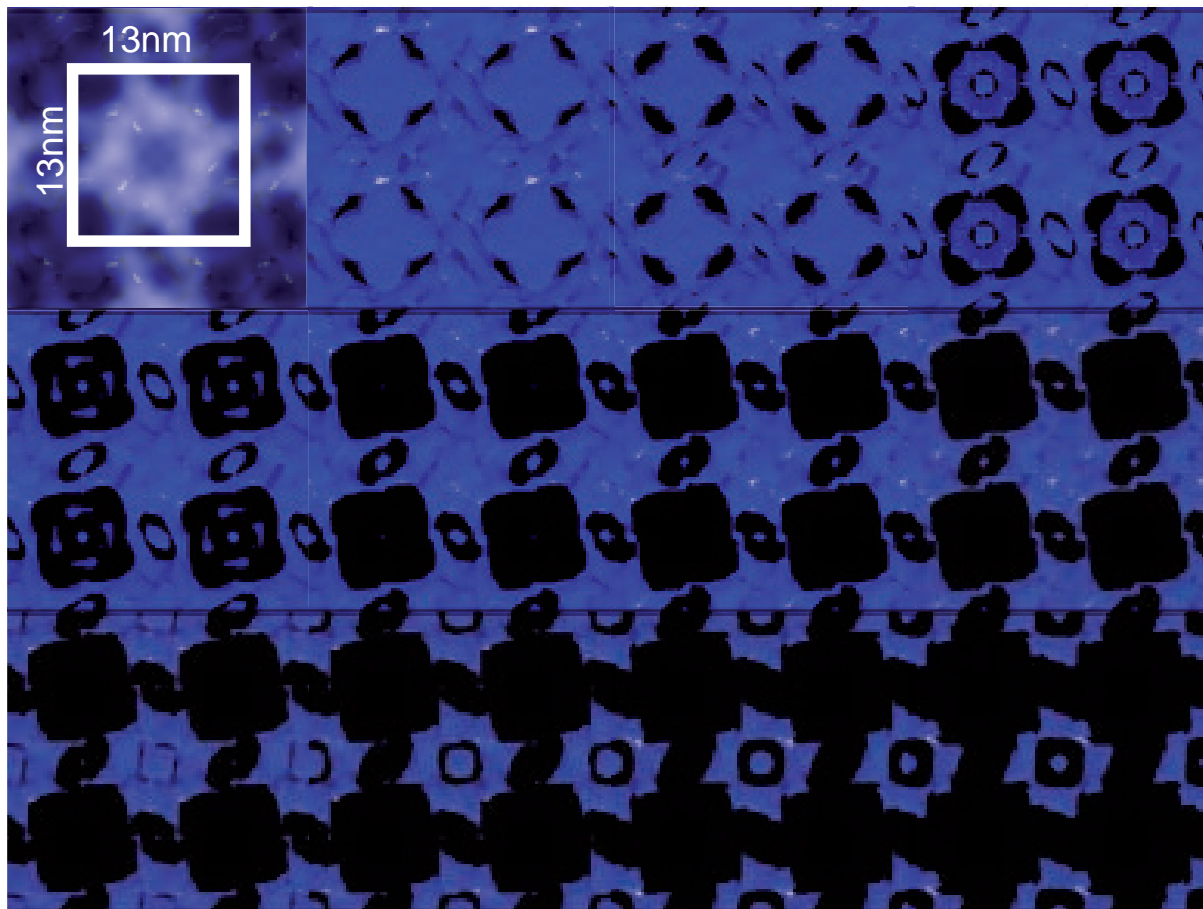


Figure 14: Reconstruction of the three-dimensional density distribution of one SbpA unit cell calculated by inverse Fourier transform of single sections of SbpA self-assembly products.²⁶⁵ A transmission electron microscope was used to obtain sections of the unit cell by performing tilting measurements. In the top left corner, an electron micrograph of one unit cell is shown. The density boundaries are systematically increased (starting in the top left corner). A superposition of all sections results in a three-dimensional unit cell.

The resulting structural model of an SbpA unit cell is shown in figure 15. The monomers are interlocked into one another, where both termini are accessible at either surface (magnified beads in figure 15). Figure 15 D shows the electron density contrasts as calculated by small-angle X-ray scattering. Apparently, higher electron density contrasts accumulate in those domains that are not part of direct overlaps in the unit cell, which shows that electron density contrasts are diminished if interactions take place. Finally, the unit cell model was compared with experimental studies addressing the self-assembly capability of various truncated forms of the protein. A fairly good consistency between the model and experimental results as well as the SAXS results strongly substantiates the structural model of the SbpA unit cell.

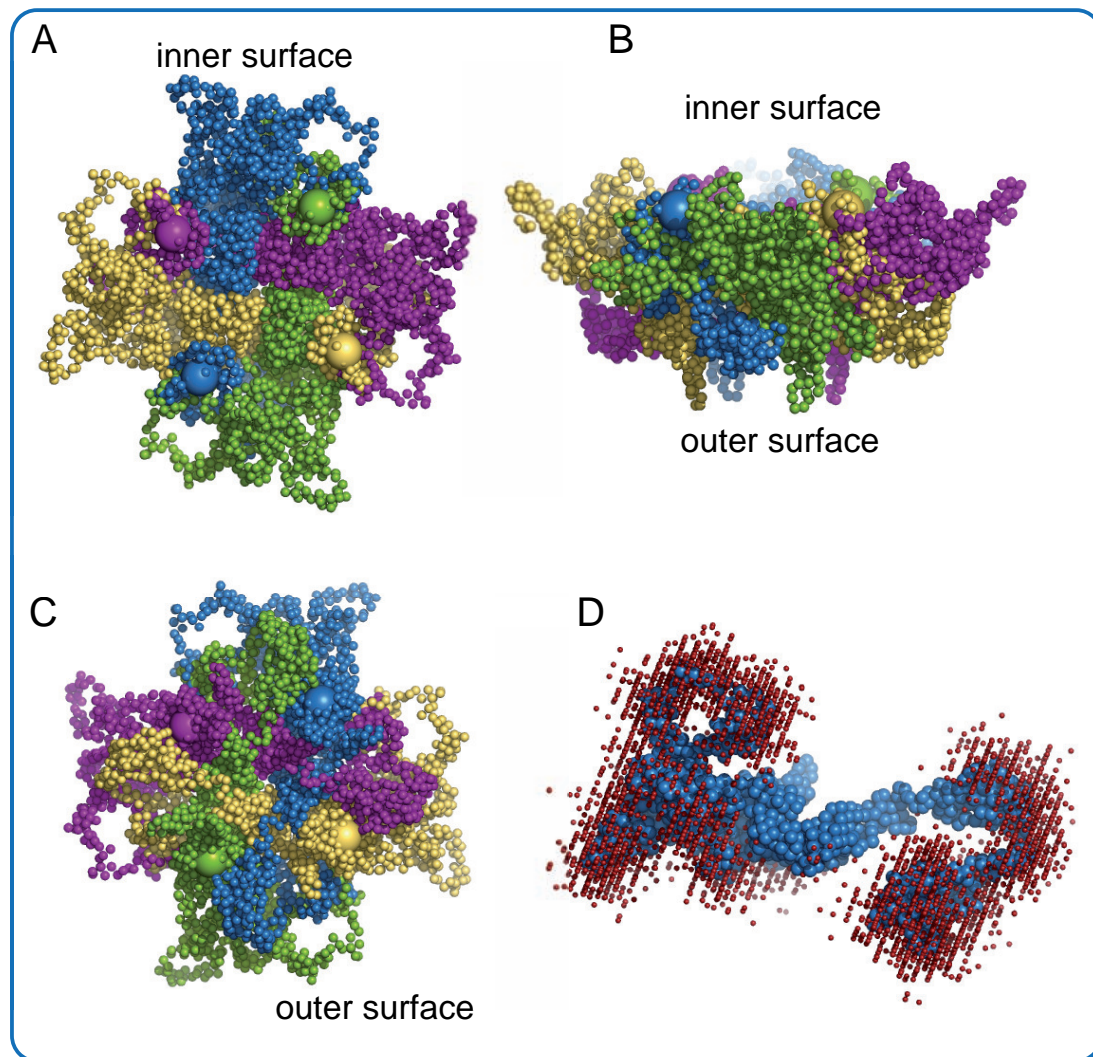


Figure 15: Three-dimensional structure of the SbpA unit cell. Every monomer in the tetramer is illustrated in a different color. The proteins are interlocked into each other. (A) Inner surface of the tetramer, which anchors the protein to the cell surface. The N-termini are represented by magnified beads and are accessible on the surface. (B) Side view of the unit cell. (C) Outer surface of the tetramer, which is exposed to the surroundings of the cell. The C-termini are also accessible and marked as beads. (D) Scattering clusters (red beads) of one SbpA monomer as determined by SAXS and a Monte Carlo algorithm. The scattering clusters represent regions of high electronic contrast, where those domains in the protein, which do not show high electronic contrast, are related to interacting or overlapping parts in the tetramer.

1.8 Molecular mechanisms of S-layer protein self-assembly

The correlation between the structure of an S-layer protein and the morphology of an S-layer lattice can give an insight into the molecular mechanisms guiding the self-assembly of proteins into functional large-scale structures. Here, the structural model of SbsB (see chapter 1.3) was used in order to simulate the self-assembly of this protein into a lattice exhibiting p1 symmetry using Monte Carlo simulations (see chapter 1.2.5). The simulation of the self-assembly process demands a coarse-graining of the protein structure in order to reduce the interaction sites and therefore save computational costs. Figure 16 shows the coarse-grained protein model. Every amino acid residue is represented by a single bead, which is located in the center of mass of the whole amino acid. The protein is treated as a rigid body. This coarse-graining approach provides for the shape of the protein as well as the charge decoration (figure 16 colored beads) and the amino acid composition.

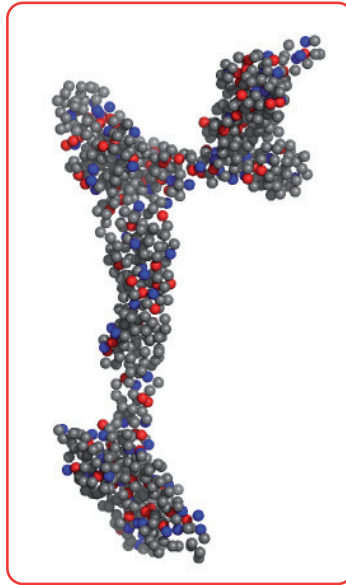


Figure 16: Coarse-grained model of the S-layer protein SbsB. Each amino acid residue is represented by a single bead of 0.65 nm diameter, which is located at the center of mass. Negatively charged beads are colored in red and positively charged beads are colored in blue. The protein is treated as a rigid body.

In a first attempt, the interaction between two monomers was investigated using continuum Monte Carlo simulations. The interaction potential is based on a statistical potential (see chapter 1.2.6) formulated by Miyazawa and Jernigan¹⁴² M_{ij} , which was smoothed by a sigmoidal function S_{ij} over an interaction range defined by a lower bound r_{LB} and an upper bound r_{UB} :

$$U_{contact} = \sum_{i \in A} \sum_{j \in B} M_{ij} S_{ij} \quad (87)$$

$$S_{ij} = \frac{\frac{1}{1+e^{r_{ij}}} - \frac{1}{1+e^{r_{UB}}}}{\frac{1}{1+e^{r_{LB}}} - \frac{1}{1+e^{r_{UB}}}} \quad (88)$$

The electrostatic interactions are taken into account by a Debye-Hückel potential:

$$U_{electrostatic} = \sum_{i \in A} \sum_{j \in B} \frac{q_i q_j \ell_B \exp(-r_{ij}/\kappa)}{r_{ij}} \quad (89)$$

with $\kappa = 1$ being the Debye-length, $\ell_B = 1$ being the Bjerrum length, and q_{ij} the respective charges of the amino acid side chain (+1, 0, -1). In order to prevent unphysical interpenetration of interacting beads, a cutoff potential is added, which accounts for the fact that the spherical nature of the bead does not correctly represent the extended conformation of an amino acid sidechain:

$$U_{cutoff} = \begin{cases} 0 & r_{CM}^A - r_{CM}^B > \eta L \\ \infty & r_{CM}^A - r_{CM}^B < \eta L \end{cases} \quad (90)$$

where r_{CM} is the center of mass of the respective amino acid and η is the cutoff parameter, which was chosen to be 0.15 after evaluating different possibilities. The final potential energy is given by:

$$U_{total} = U_{contact} + U_{electrostatic} + U_{cutoff} \quad (91)$$

Figure 17 shows the result of the simulation of two monomers. Two stable conformations exhibiting the lowest contact energy values could be identified, where the two monomers are interacting along the y- and the x-axis respectively (figure 17 A, C). Additionally, various different intermediate states emerged, which are structurally closely related to the two ground state conformations. These intermediate states might play a crucial role during the self-assembly of the lattice, but have not been included in the following large-scale simulations in order to reduce simulation time. The energy values as well as the interaction directions were consequently used in order to perform lattice Monte Carlo simulations, where the proteins are represented as unit cubes (figure 17 B, D). The unit cubes can move randomly within a cube of linear size 50 in units of the size of each monomer. Two different simulations were performed, where firstly, a single seed was placed in the center of the cube, which serves as the nucleation point for the self-assembly process, which allows the calculation of the growth rate, and secondly, random nucleation could take place in order to get an insight into the cluster size distribution and the growth mechanism. Finally, the modeled S-layer lattice was compared with experimental studies addressing the accessibility of various amino acid residues in the

lattice and the monomer, respectively. A good agreement between the results of the simulations and the experiments support the proposed model of the lattice. The molecular mechanisms guiding the proteins to self-assemble into such regular structures are mainly driven by hydrophobic interactions, which could be also shown by Monte Carlo simulations of simplified coarse-grained models of the protein. Such a simple approach including a simplified coarse-grained model of the protein and different Monte Carlo simulations, offers the possibility to study a highly complex protein system such as S-layers at an amino acid level, where fundamental experimental results can be reproduced and molecular mechanisms can be fairly well explained.

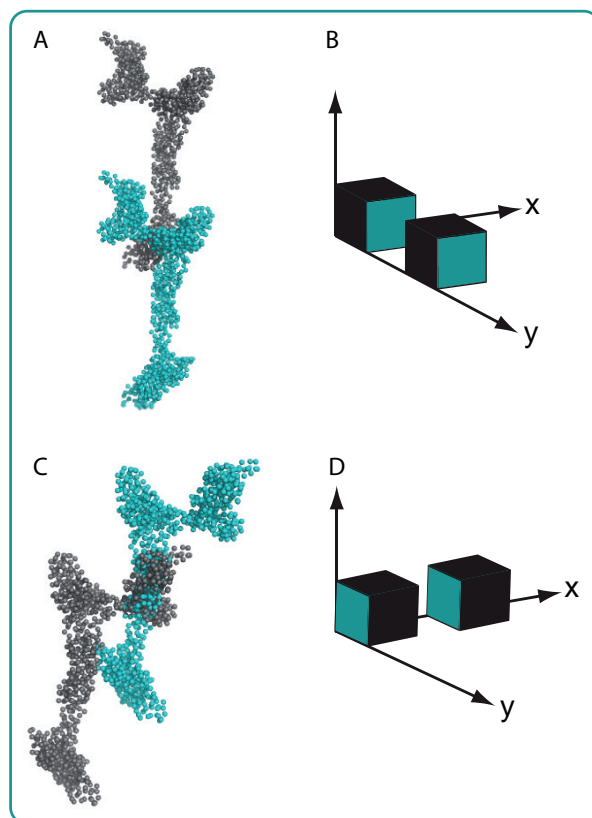


Figure 17: Results of the Monte Carlo simulation of the interaction of two monomers. (A) (C) Two stable dimer conformations could be identified, where they are interacting in (A) y-direction and (C) x-direction, respectively. (B) (D) Corresponding dimer conformations, where the proteins are represented by unit cubes to be used in the lattice Monte Carlo simulation of the large-scale assembly process. Blue colored faces represent the interaction directions.

Based on these two Monte Carlo approaches, the morphology of a self-assembled p1 S-layer lattice could be calculated. Figure 18 shows the coarse-grained structural model of the S-layer sheet. The architecture of the lattice exhibits identical pores, which mainly consist of hydrophilic residues (magnified view in figure 18). The lattice parameters are in good agreement with experimentally determined values and an anisotropy of the lattice - already discovered by various experimental methods - can be explained by the model. The N- and C-terminal regions of the single monomers stick out of the lattice (figure 18

B), which leads to charge and hydrophobicity differences of the inner and outer surface of the lattice. This phenomenon also highly influences the behavior of S-layer sheets in solution.

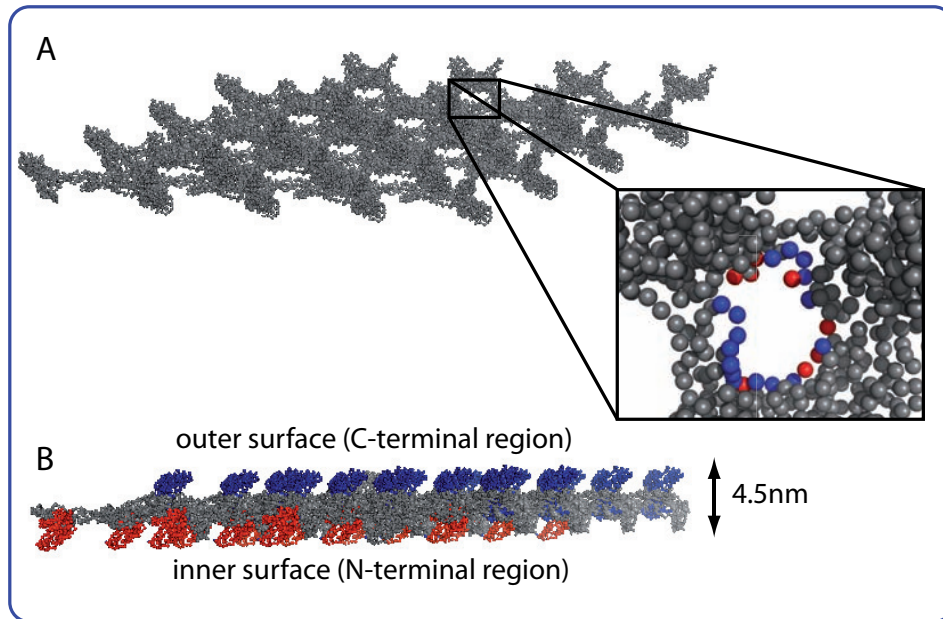


Figure 18: Morphology of the self-assembled S-layer lattice as obtained by lattice Monte Carlo simulations. (A) Lattice with p1 symmetry made up of 16 monomers, represented by the coarse-grained model. The magnified view shows the details of one pore. Hydrophilic residues are colored in blue, hydrophobic ones in red. (B) Cross section of the lattice. The outer surface is composed of the C-terminal region (blue), the inner surface of the N-terminal region (red).

Structure prediction of an S-layer protein by the mean force method

C. Horejs, D. Pum, and U. B. Sleytr

Center for Nanobiotechnology, BOKU University of Natural Resources and Applied Life Sciences,
Gregor Mendel-Straße 33, A-1180 Vienna, AustriaR. Tscheliessnig^{a)}Institute for Applied Microbiology, BOKU University of Natural Resources and Applied Life Sciences,
Muthgasse 18, A-1190 Vienna, Austria and Institute for Chemical and Energy Engineering, BOKU University
of Natural Resources and Applied Life Sciences, Muthgasse 107, A-1190 Vienna, Austria

(Received 22 October 2007; accepted 27 November 2007; published online 12 February 2008)

S-layer proteins have a wide range of application potential due to their characteristic features concerning self-assembling, assembling on various surfaces, and forming of isoporous structures with functional groups located on the surface in an identical position and orientation. Although considerable knowledge has been experimentally accumulated on the structure, biochemistry, assemble characteristics, and genetics of S-layer proteins, no structural model at atomic resolution has been available so far. Therefore, neither the overall folding of the S-layer proteins—their tertiary structure—nor the exact amino acid or domain allocations in the lattices are known. In this paper, we describe the tertiary structure prediction for the S-layer protein SbsB from *Geobacillus stearothermophilus* PV72/p2. This calculation was based on its amino acid sequence using the mean force method (MF method) achieved by performing molecular dynamic simulations. This method includes mainly the thermodynamic aspects of protein folding as well as steric constraints of the amino acids and is therefore independent of experimental structure analysis problems resulting from biochemical properties of the S-layer proteins. Molecular dynamic simulations were performed in vacuum using the simulation software NAMD. The obtained tertiary structure of SbsB was systematically analyzed by using the mean force method, whereas the verification of the structure is based on calculating the global free energy minimum of the whole system. This corresponds to the potential of mean force, which is the thermodynamically most favorable conformation of the protein. Finally, an S-layer lattice was modeled graphically using CINEMA4D and compared with scanning force microscopy data down to a resolution of 1 nm. The results show that this approach leads to a thermodynamically favorable atomic model of the tertiary structure of the protein, which could be verified by both the MF Method and the lattice model. © 2008 American Institute of Physics. [DOI: 10.1063/1.2826375]

I. INTRODUCTION

Two-dimensional bacterial cell surface layer protein crystals (S-layers) are the most commonly observed cell surface structure in bacteria and archaea. These surface layer crystals are made up of identical protein units called S-layer proteins, which self-assemble into monomolecular crystalline arrays in suspension and on various solid substrates.¹ In gram-positive bacteria and in archaea such S-layer lattices assemble on the surface of the wall matrix, and in gram-negative bacteria the S-layer is attached to the lipopolysaccharide component of the outer membrane. Transmission electron microscopic and atomic force microscopic studies showed² that S-layer lattices are 5–20 nm thick, often have a smooth outer surface and a more corrugated inner surface, and the protein subunits are organized in lattices with oblique ($p1, p2$), square ($p4$), or hexagonal ($p3, p6$) symmetry with a center-to-center spacing of approximately 3–35 nm. S-layer lattices are very porous structures, and due to the fact that they are composed of identical species of subunits, they

exhibit pores of identical size and morphology with diameters in the range of 2–8 nm.³ These characteristic features, the assembling of the subunits into monomolecular arrays in suspension, on surfaces or interfaces, and the forming of isoporous structures with functional groups located on the surface in an identical position and orientation, have led to the applications of S-layers as ultrafiltration membranes, immobilization matrices for functional molecules, conjugate vaccines, carriers for Langmuir-Blodgett films, and as patterning elements in molecular nanotechnology.⁴ S-layer lattices are composed of single protein or glycoprotein species with a molecular mass ranging from 40 to 200 kDa. The secondary structure of these S-layer proteins has been identified by circular dichroism (CD) or has been derived from the amino acid sequence^{5,7,8} and it mainly contains β -sheets (40%) and α -helices (20%).¹ By sequence comparison, so-called S-layer homologous (SLH) domains were identified on the N-terminal part of several S-layer proteins, and these SLH domains were suggested to anchor the proteins to the cell surface.⁶

Until now no structural model of an S-layer protein at atomic resolution has been available, whereas this fact may

^{a)}Electronic mail: rupert.tscheliessnig@boku.ac.at.

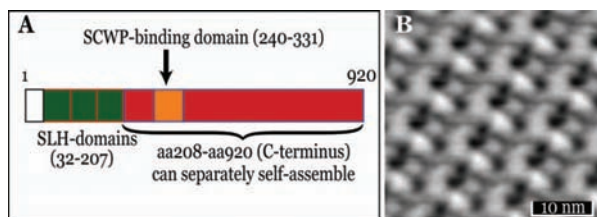


FIG. 1. (Color) (A) Schematic picture of the domain distribution of SbsB. Green: N terminus (aa32–aa207) with three SLH domains. White: signal peptide. Red represents the C terminus mainly consisting of β -sheets, and the SCWP-binding domain is indicated yellow. (B) Projection map of native SbsB derived by scanning for microscopy.

be explained by the molecular mass of the protein subunits being too large for structural studies by NMR (nuclear magnetic resonance) as well as by the property of the proteins to self-assemble into two-dimensional lattices and thereby avoiding the formation of isotropic three-dimensional crystals, which are required for x-ray crystallography.^{1,9} Regarding the wide range of the application potential of S-layer proteins, the importance of clarifying the tertiary structure and, therefore, the exact distribution of the amino acids in the lattices is obvious.

These experimental structure analysis problems resulting from biochemical properties of the S-layer proteins can be avoided by using molecular dynamic (MD) simulations to calculate the folded form of the proteins, which has been applied to the S-layer protein SbsB from *Geobacillus stearothermophilus* PV72/p2, a thermophilic gram-positive bacterium.⁷ This protein consists of 920 amino acids, has a molecular mass of 98 kDa, and assembles in an oblique ($p1$) lattice type. SbsB does not contain any cysteine, five methionine residues, and 35.4% polar, 11.8% acidic, 10.8% basic, and 49% hydrophobic residues. Valine, threonine, and alanine are the most frequently occurring residues. The isoelectric point (pI) is 5.7. Secondary structure determination by far-UV CD yielded α -helical fractions of 7%–11% and β -sheet fractions of 34%–38%.⁷ Structure-function relationship studies by producing various truncated forms of SbsB (Refs. 7 and 9) indicated that rSbsB_{208–920} retains the ability to self-assemble, thus the N terminus of the protein (aa1–aa207) is not essential for the self-assembling process. On the N terminus, two binding domains could be identified,⁸ one for peptidoglycan and another for a secondary cell wall polymer (SCWP). By structure comparison, three SLH domains were encoded, that are responsible for attaching exoproteins of gram-positive bacteria to the underlying cell wall, whereas the peptidoglycan-binding domain lies inside this SLH region. The SLH domains mainly consist of α -helices according to secondary structure predictions and CD measurements.⁹ The SCWP-binding domain is supposed to be aa240–aa331.⁸ The other part of the protein (aa208–aa920) forms mainly β -sheets and is therefore responsible for the nonglobular structure of SbsB, that is responsible for lattice formation [Fig. 1(A)]. Digital image reconstructions of rSbsB forms revealed that the lattice constants are $a = 10.4$ nm, $b = 7.9$ nm, and $\gamma = 81^\circ$ [Fig. 1(B)].⁹ Predicting the tertiary structure of SbsB by molecular dynamic simulations, thus by thermodynamic principles, requires a predeter-

mination of structurally meaningful parts of the protein that can be separately simulated and modeled, because the length of SbsB avoids a conformation search including the whole protein. The conformation space is too large and the protein would stay in local energy minima, whereby the global energy minimum could not be calculated. Molecular dynamic simulation processes require smaller molecular systems to determine energy differences and thermodynamic parameters. Therefore, the modeling of the tertiary structure of SbsB started with an analysis of the amino acid sequence. The first step was to identify if the target sequence of interest is homologous to other known sequences stored in databases. This could be clarified by sequence homology searching. A domain prediction was used to identify conserved motifs that are structurally or functionally important. Based on previous information and domain, as well as secondary structure predictions, SbsB was split into eight different parts, that were supposed to be structurally independent, whereas the separation of N and C terminus was obvious. To obtain three-dimensional coordinates of the single parts, a premodeling by fold recognition¹⁰ was performed. With these parts, molecular dynamic simulations could be processed and the free energy of the various conformations could be calculated. Finally, the obtained structure of SbsB could be analyzed by using the mean force method to verify that the calculated conformation is located in the global free energy minimum.

II. THEORY

A vast amount of literature exists in how to sample rare events, events that happen *once in a blue moon*.¹¹ Adoptions thereof were developed to investigate protein-protein interactions,¹² complexation,¹³ adsorption,^{14–17} or reaction rates.^{18,19} A detailed recent review on blue moon ensembles and the concept of the potential of mean force (PMF) and methods to calculate either is given in (Ref. 20). For the sake of brevity, we shortly draft the most important working equations.

In this paper protein configuration shall be given by Cartesian coordinates of the individual N sites, \mathbf{x}^N . Its potential energy resembles $V(\mathbf{x}^N)$. The probability to find a protein in a certain configuration, quantified by a reaction coordinate $\zeta(\mathbf{x}^N) = \zeta'$, is given by

$$\pi[\zeta(\zeta')] = \langle \delta(\zeta(\mathbf{r}^N) - \zeta') \rangle = C \exp(-\beta w[\zeta(\zeta')]). \quad (1)$$

Therein, $\langle \cdots \rangle$ denotes a canonical ensemble average, $\beta = (k_B T)^{-1}$, with k_B as the Boltzmann constant and T the temperature of the system. The quantity $w[\zeta(\zeta')]$ equals the potential of mean force. The constant C is accessible by the normalization condition of the probability $\pi[\zeta(\zeta')]$.

According to Jarzynski,^{21,22} the work $W[\zeta(\zeta')]$, and the change of free energy $A[\zeta(\zeta')]$ resemble in general an inequality: $\exp(-\beta \Delta A[\zeta(\zeta')]) = \langle \exp(-W[\zeta(\zeta')]) \rangle$, but are equal for infinitely slow changes of $\zeta(\zeta')$. This equality, aside experimental impact,^{23,24} led to steered molecular dynamics,^{25,26} and are used to test a full atomistic protein model from the *scratch*.

The potential of mean force, $w[\zeta(\zeta')]$, and the mean force $\langle F[\zeta(\zeta')] \rangle$ are related to one another by a thermodynamic integration,

$$\Delta w[\zeta(\zeta')] = \Delta A[\zeta(\zeta')] = - \int_{\zeta_0}^{\zeta'} \langle F[\zeta(\zeta')] \rangle d\zeta'. \quad (2)$$

The mean force may be calculated either by a conditional or constrained MD simulation.^{20,12,28,19,29} In the second case, it is calculated by constraining the reaction coordinate $\zeta(\zeta')$,^{30,31}

$$\begin{aligned} \langle F[\zeta(\zeta')] \rangle_{\zeta'} = & - \frac{\langle \partial_{\zeta} \mathbf{x}_i \cdot \partial_{\mathbf{x}_i} V(\mathbf{x}^N) Z^{-1/2} \rangle_{\zeta'}}{\langle Z^{-1/2} \rangle_{\zeta'}} \\ & + \frac{1}{2\beta} \cdot \frac{\langle \{ \sum_{i=1}^{N_c} m_i^{-1} \nabla_i \zeta \cdot \nabla_i Z \} Z^{-5/2} \rangle_{\zeta'}}{\langle Z^{-1/2} \rangle_{\zeta'}}. \end{aligned} \quad (3)$$

Note that the second summand is a consequence of the metric correction, and has been introduced by Fixman.²⁷ The factor Z is defined by $Z = \sum_{i=1}^{N_c} (1/m_i) (\partial_{\mathbf{x}_i} \zeta(\mathbf{x}^N))^2$.

The local probability distribution $\pi[\zeta]$ and the MF are correlated by the equation

$$\log \left(\frac{\pi[\zeta]}{\pi[\zeta_0]} \right) = -\beta \int_{\zeta_0}^{\zeta} \langle F[\zeta(\zeta')] \rangle_{\zeta'} d\zeta', \quad (4)$$

$$\pi[\zeta] = \pi[\zeta_0] \exp \left(\beta \int_{\zeta_0}^{\zeta} \langle \partial_{\zeta} \mathbf{x}_i V(\mathbf{x}^N) \rangle_{\zeta'} d\zeta' \right). \quad (5)$$

From the probability, we calculate the average value of the reaction coordinate $\bar{\zeta}$ by $\bar{\zeta} = \int_{\Delta} \zeta \pi[\zeta] d\zeta$. Therein, Δ indicates an interval, and for the local probability distribution $\pi[\zeta]$, the normalization $\int_{\Delta} \pi[\zeta] d\zeta = 1$ holds.

In this paper, however, $\zeta(\zeta')$, is always a convenient generalized coordinate. It describes the distance of two subunits, e.g., domains, of a protein, and thus is a linear function of the Cartesian coordinates. For the present system, the reaction coordinate is $\zeta(\mathbf{x}^N) = |\alpha_i^A \mathbf{x}_i^A - \alpha_j^B \mathbf{x}_j^B|$, with $\alpha_j^{A/B} = m_j/m_i$, m_j gives the mass of the j th site and m_i the total mass of the protein. It is trivial then that $Z = \sum_{i=1}^{N_c} (\alpha_i^A)^2 / m_i + \sum_{i=1}^{N_c} (\alpha_i^B)^2 / m_i$, it is a constant and does neither contribute to the PMF nor to $\pi[\zeta]$.

III. METHODS

Molecular dynamic simulations were performed using the software NAMD. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign³² and is free to download and use at www.ks.uiuc.edu/Research/namd/. All visualizations of the simulation results were made with the molecular graphics program VMD,³⁴ which can be downloaded at www.ks.uiuc.edu/Research/vmd/. NAMD simulation software generates structure files from the CHARMM force field, and uses periodic boundary conditions and the particle-mesh Ewald method³³ (PME) for long-range interactions. The equations of motion, that is, the time evolution of the Hamiltonian system, are integrated by the Verlet method.³⁵ NAMD

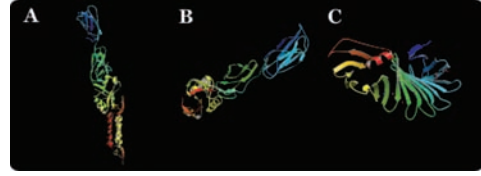


FIG. 2. (Color) (A) Crystal structure of a C-terminal 282-residue fragment of intimin in complex with translocated intimin receptor intimin-binding domain. (B) Crystal structure of a C-terminal 282-residue fragment of enteropathogenic *E. coli* intimin (Ref. 40). (C) Crystal structure of an OspA mutant (Ref. 41).

offers the possibility of steered molecular dynamics,³² whereby external forces can be applied to an atom or to a set of atoms and, therefore, the atoms can be pulled along a vector and conformation studies can be accomplished. Within this method, thermodynamic properties such as free energy data can be calculated using the adaptive biasing force (ABF) method.³⁸ VMD was used to visualize simulation results and to calculate the root mean square deviation (rmsd) of the system. All simulations were computed at the SUN cluster Phoenix at phoenix.zserv.tuwien.ac.at. The pdb files for the N-terminal domains were created using the SWISS-PDBVIEWER.³⁶ The C-terminal domains were premodeled via fold recognition¹⁰ using PHYRE at www.sbg.bio.ic.ac.uk/phyre/, that is the successor of 3D-PSSM.³⁷ Simulations were in water at ambient conditions and in vacuum.

IV. RESULTS AND DISCUSSION

A. Sequence similarity search

Sequence homology searches of the N and the C terminus were separately performed using BLAST (Ref. 39) at <http://www.ncbi.nlm.nih.gov/BLAST/>. Obviously, sequence homologies could be found to other S-layer proteins, of that no tertiary structure has been clarified neither. Regarding the sequence similarities to such a lot of other S-layer proteins, it is conjecturable that there are conserved domains within the S-layer protein family. In the first instance, these conserved parts are the SLH domains, and also other structurally related parts may occur. Concerning the N terminus, the sequence-based homologies mainly do not consist of other proteins that are not S-layer proteins. Figure 2 shows the results for the C terminus (aa208–aa920). Again, the homologies mainly consist of other S-layer proteins, but there are also sequence similarities to bacterial Ig-like group 2 domains. Proteins that contain this domain are found in a variety of bacterial and phage surface proteins such as intimins [Figs. 2(A) and 2(B)], whereas intimin is a bacterial cell-adhesion molecule that mediates the intimate bacterial host-cell interaction. These domains are mostly composed of antiparallel β -sheets, that are linked by turns and coils. BLAST also found homologies to the protein OspA [Fig. 2(C)], whose structure is also mainly made of β -sheets, turns, and coils.

B. Secondary structure prediction

Secondary structure predictions were made using all available algorithms at <http://www.expasy.org/tools/>,

whereas there was a clear consensus concerning the secondary structure elements of the N and the C terminus. For the N terminus (aa1–aa207), the clear consensus were α -helices and coils, and for the C terminus (aa208–aa920), β -sheets and coils. α -helices and β -sheets are two completely different secondary structure elements, not only regarding the structure itself, but also the thermodynamics of formation and stability. Apparently, the two main structural domains of SbsB are the N and the C terminus, distinguishable from each other through different secondary structure elements, varying conserved domains, and, consequently, variable functions. As mentioned, experimental studies support this theory that SbsB consists of two independently acting domains concerning self-assembly and anchoring the protein on the cell surface.

C. Domain prediction and choice of structurally meaningful parts

Domain predictions were performed using MYHITS (Ref. 42) at <http://myhits.isb-sib.ch/cgi-bin/index>, SMART (Ref. 43) at <http://smart.embl-heidelberg.de/>, and CDART (Ref. 44) at <http://www.ncbi.nlm.nih.gov/>, whereas CDART uses the database PFAM,⁴⁵ thus the results of CDART will be indicated by “Pfam” in the following. The N and the C terminus have each been analyzed by the three domain prediction servers. For the N terminus (aa1–aa207), all servers predicted three SLH domains at different amino acid positions (Fig. 3). Regarding the prediction generated by PFAM, the first SLH domain ranges from aa3 to aa43, but the signal peptide is aa1–aa31 and, therefore, this prediction is not reasonable, because the signal peptide cannot contain an essential protein domain. MYHITS and SMART both predict three SLH domains inside the functional region of SbsB, whereas MYHITS proposes continuous domains, which is in line with the sequence similarity search and plausible regarding the distribution of the α -helices resulting from the secondary structure prediction, if the adoption that there are two α -helices per SLH domain is presumed.^{8,9} Consequently, the three SLH domains predicted by MYHITS are used for dividing the N terminus of SbsB in three structurally meaningful parts, which are aa1–aa91 (signal peptide and first SLH domain), aa92–aa144 (second SLH domain), and aa145–aa207 (third SLH domain). The signal peptide has to be included in the simulation of the folding process, because it is also present under native conditions and can therefore have an influence on the formation of the adjacent amino acids. For the C terminus (aa208–aa920), only few domains were predicted by the three servers, which is explainable by the rare sequence similarities to known proteins, but again, bacterial Ig-like group 2 domains could be detected by all servers, which corresponds with the BLAST results. Figure 3(A) shows an example of such a domain proposed by PFAM. The servers predict two bacterial Ig-like group 2 domains, whereas SMART and MYHITS are exactly at the same amino acid positions. Moreover, PFAM predicts a fibronectin type III domain, that is illustrated in Fig. 3(B). This structural domain is supposed to play a role in cell surface binding of fibronectins and is also mainly composed of β -sheets and coils and, therefore, structurally related to the bacterial Ig-like group 2 domain. Regarding the

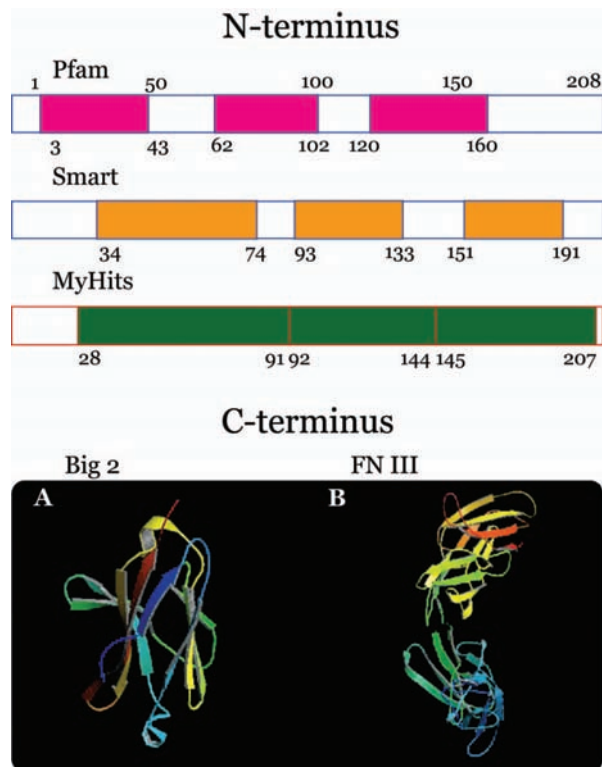


FIG. 3. (Color) Schema of the domain prediction results for the N terminus (aa1–aa207) by PFAM (Ref. 44), SMART (Ref. 43), and MYHITS (Ref. 42). The colored boxes represent the predicted SLH domains. (A) shows an example of a bacterial Ig-like group 2 domain (Ref. 46), while (B) shows an example of a fibronectin type III domain (Ref. 47, both examples were predicted by PFAM.

secondary structure prediction that the C terminus mostly consists of β -sheets and coils, these two predicted domains are likely to be part of SbsB. Based on the secondary structure prediction and these domain predictions, the C terminus is split into the following parts: aa208–aa293 (possible fibronectin type III domain), aa294–aa378 (possible bacterial Ig-like group 2 domain), aa379–aa496 (possible bacterial Ig-like group 2 domain), aa497–aa770 (unknown domain), and aa771–aa920 (unknown domain).

D. Fold recognition

Ultimately, SbsB has been divided into eight structurally independent domains for the simulation of the folding, three domains of the N terminus, and five domains of the C terminus, whereas the five C terminal domains mainly consist of β -sheets and the N-terminal domains of α -helices. Regarding the simulation process itself, that is based on the energy landscape of the protein, β -sheet formation, in contrast to α -helices, poses a problem, because the net change in conformational free energy is not sufficient to overcome the loss in configurational entropy that accompanies sheet formation.⁴⁸ That is to say, that a molecular dynamic simulation based on interactions defined by a force field of a protein domain, that is supposed to be composed of β -sheets, which is the case concerning the five domains of the C terminus, would not result in the formation of β -sheets, because

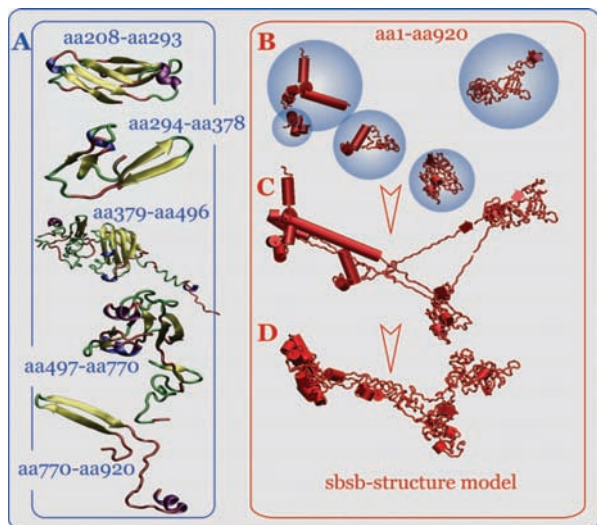


FIG. 4. (Color) (A) 3D models of the five C-terminal domains of SbsB created by PHYRE Ref. 31. Yellow arrows: β -sheets, violet strands: α -helices, green line: turns, and red line: coils. Modeling method: (B) Sbsb for the individual domains were equilibrated in water spheres at 310 K, (C) jointed in vacuum, and (D) SbsB structure obtained by MD.

the interaction possibilities for the atoms relating to the stabilizing forces for sheet formation, stored in the force field, cannot be greater than $T\Delta S$ to overcome the loss of conformational entropy. Another complication concerning β -sheets is that these secondary structure elements may involve interactions between distal regions of the amino acid sequence and are expected to be context dependent. Considering that the folding of the C terminus has to be proceeded piecewisely, because of the length of the protein, this characteristic of β -sheet formation is problematic. Therefore, the development of predetermined three-dimensional coordinates, that tend to form β -sheets rather than other secondary structure elements, is required, which is achieved by modeling the tertiary structure of the five C terminal domains using fold recognition.¹⁰ Fold Recognition was performed for the C-terminal parts using PHYRE at www.sbg.bio.ic.ac.uk/phyre/, that is the successor of 3D-PSSM.³⁷ Figure 4(A) shows the three-dimensional models of the five C terminal parts of SbsB created by PHYRE. All models consist of β -sheets, coils, and turns as secondary structure elements, which conforms to the secondary structure prediction and the domain prediction. The three-dimensional (3D) model for domain aa208–aa293 is made of antiparallel β -sheets, which is typical for fibronectin type III domains. PHYRE used these domains to model the SbsB domain with an estimated precision of 75%, which is in line with the domain prediction for this amino acid section. The two domains aa294–aa378 and aa379–aa496 are also both composed of antiparallel β -sheets, whereat the second domain consists of much more sheets. Both domains were regarded to be bacterial Ig-like group 2 domains according to the domain prediction. PHYRE modeled these two domains based on intimin cell-adhesion fragments, which typically contain these Ig-like domains. The last two domains aa497–aa770 and aa771–aa920, that could not be predefined by the domain prediction, also con-

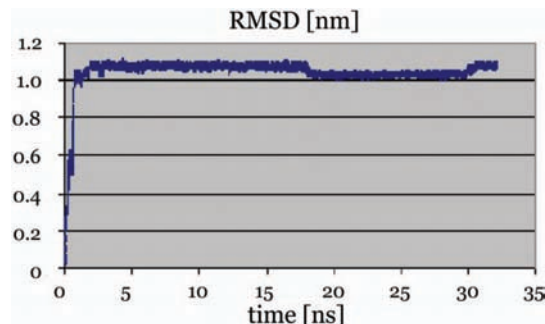


FIG. 5. (Color) RMSD developing of SbsB during the simulation. RMSD in nanometer (nm), time in nanoseconds (ns).

tain β -sheets and coils. These two domains were also modeled based on Ig-like domains, which is in line with the secondary structure prediction.

E. Creation of pdb files of the C-terminal parts

The created models of the C-terminal domains are composed of different amino acids than the original SbsB sequence, because the amino acids were matched through the modeling process to fit the predicted secondary structure. Therefore, to build up pdb files for the five domains for the simulation, the new coordinates have to be merged to the amino acids of the single parts by comparing the sequence of the model with the original sequence, and if the modeled amino acid fits the original one, the coordinates of the model are taken over, if not, random coordinates are included in the new pdb file.

F. Molecular dynamic simulations

The folding of the eight domains of SbsB was computed until the parts were equilibrated, which was verified by calculating the rmsd development during the simulation process. The overall folding of SbsB was calculated by joining all equilibrated parts together by combining the pdb files of the single simulation results. With the new pdb file, simulation runs were performed in vacuum for 30 ns. Figure 5 illustrates the developing of the rmsd value during the simulation. The protein is equilibrated after a few nanoseconds and the rmsd does not change over time. The main folding process takes place in the very first time of the simulation, and this conformation stays constant over the whole simulation time. A fact that goes in line with the native conditions of the protein folding process in the cell, that lasts only a few nanoseconds until the native form of a protein is reached.⁴⁹ Figure 6 shows the results of the simulation. The tertiary structure of SbsB is L shaped, whereas the L is formed by the C terminus. The three SLH domains are arranged closely at the N terminus of the protein and are represented by the purple part in Fig. 6. Most part of the α -helical content of SbsB is located at the N terminus, and these α -helices form a pocket enabled by turns, that link the helices together and facilitate the flexibility of the N-terminal domains (Fig. 7). The C terminus contains three main domains, whereas one is linear and two are formed as globular domains that are

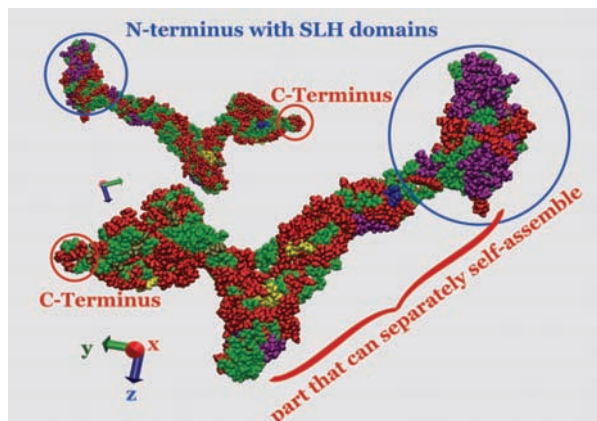


FIG. 6. (Color) Tertiary structure of SbsB obtained via simulations in vacuum after 30 ns. Blue: β -bridge, green: turn, red: coil, yellow: β -sheet, and purple: α -helix.

linked together by turns and coils. These C-terminal domains contain β -sheets, β -bridges, coils, and turns. The protein is composed of many pockets and niches, mainly at the N terminus made up of the SLH domains and the last part of the C terminus.

1. N terminus

The N terminus of SbsB (aa1–aa207) consists of six α -helices that are linked by turns and coils. According to the secondary structure prediction and the sequence similarity search, SbsB has three SLH domains, with every domain made up of two α -helices. The first SLH domain is composed of the two α -helices aa46–aa48 and aa69–aa73, the second of aa77–aa81 and aa105–aa113, and the third domain of aa125–aa134 and aa155–aa158 (Fig. 8). These six α -helices together with the turns form a binding pocket for anchoring the protein on the cell surface. Within this binding pocket there are two more helices located: aa313–aa319 and aa339–aa342. Therefore, the SLH region seems to be accompanied by an additional α -helical region, that is supposed to be the SCWP-binding domain.⁸ The simulation results indicate that the SLH region and the SCWP-binding domain establish one cell-binding domain by generating a binding pocket, that can flexibly move and alter its domain positions

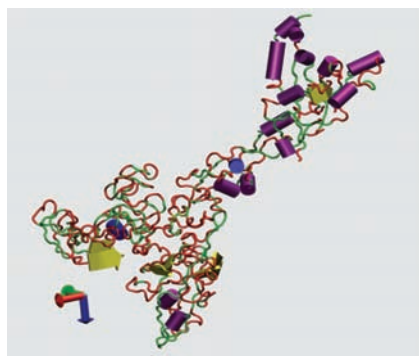


FIG. 7. (Color) Ribbon illustration of the tertiary structure of SbsB. Blue: β -bridge, green: turn, red: coil, yellow: β -sheet, and purple: α -helix.

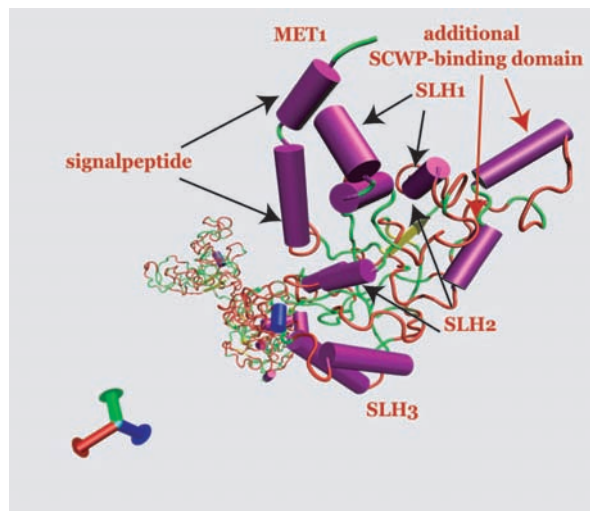


FIG. 8. (Color) Detail view of the N terminus (aa1–aa207). Blue: β -bridge, green: turn, red: coil, yellow: β -sheet, and purple: α -helix.

depending on the arrangement of the secondary cell wall polymers. The SCWP-binding domain is composed of two α -helices that are linked together by turns, whereas these turns enable the domain to move flexibly. The binding pocket of the N-terminus consists of one hydrophilic and one hydrophobic side, whereas the SCWP-binding domain lies inside the hydrophilic part. Moreover, there are much more basic amino acid side chains located within the binding pocket (arginine, lysine) than acidic side chains.

2. C terminus

The C terminus of SbsB accounts for the main part of the protein and is made up of three main domains (Fig. 9). The first domain reaches from amino acid 208 to amino acid 486 and is linear. This domain connects the N terminus with the L-shaped part of the C terminus, that is made up of the other two domains of the C terminus (aa487–aa755 and aa756–aa920). The first C-terminal domain seems to separate the two great parts of SbsB: the cell-binding domain with its SLH- and SCWP-binding regions and the main part of the C terminus responsible for the self-assembling process. Within the C terminus, 15 β -strands and 24 β -bridges can be found, whereas both the β -strands and the β -bridges are arranged as

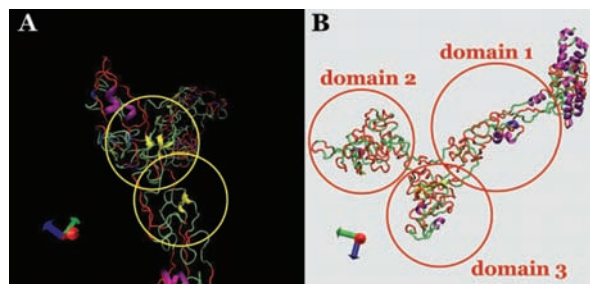


FIG. 9. (Color) Ribbon illustration of the C terminus. Productions runs covered 30 ns. (A) Antiparallel β -sheets are marked yellow. (B) C-terminal domains. Blue: β -bridge, green: turn, red: coil, yellow: β sheet, and purple: α -helix.

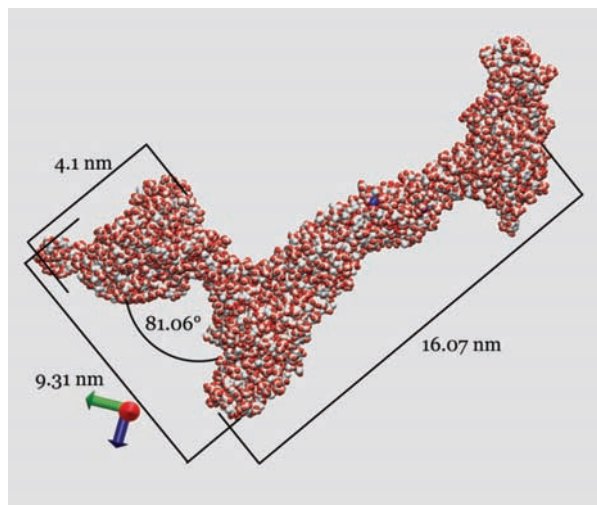


FIG. 10. (Color) Main distances and angles of SbsB.

antiparallel β -sheets. This secondary structure formation corresponds to the previous domain prediction that the C-terminal domains contain bacterial Ig-like group 2 and fibronectin type III domains, that are characterized by antiparallel β -sheets. The three C-terminal domains are linked together by various turns and coils that enable the domains to move relatively to each other. Domain 2 and domain 3 form a pocket, whereas domain 2 additionally forms an emargination. Regarding the structure of the molecular surface of the C terminus and the obvious flexibility of the C-terminal domains, this part of SbsB seems to play a major role in lattice formation by making the perfect fitting of two single proteins possible. Moreover, the C-terminal part of SbsB, especially domains 2 and 3, provide the smoother outer surface of the lattice and forms the regular pores. The roughness of the molecular surface may play a role in pore formation. The C-terminal domains contain much more hydrophilic residue side chains than the N-terminal part, primarily domain 2, that includes more hydrophilic than hydrophobic residues, whereas domain 1 contains mainly hydrophobic residues.

3. Distances and angles

Figure 10 shows the main distances of the protein SbsB and the angle between the two main C-terminal domains. Regarding the lattice constants $a=10.4$ nm, $b=7.9$ nm, and $\gamma=81^\circ$,⁹ the dimensions of the calculated tertiary structure of SbsB seems to be in a right range. The length of the protein is approximately 16.07 nm, and the width at the C-terminal domains is 9.31 nm. The width of the C-terminal domain 1, which is the smallest, is 1.97 nm. The angle between the two C-terminal domains 2 and 3 amounts to 81.06° , whereas the angle between domains 1 and 2 nearly reaches 90° .

G. Verification of the structure—MF method

The calculated tertiary structure of SbsB was systematically analyzed using the ABF method: parts of the protein were pulled along a chosen reaction coordinate ζ , and the protein was deformed to quantify the stability of the struc-

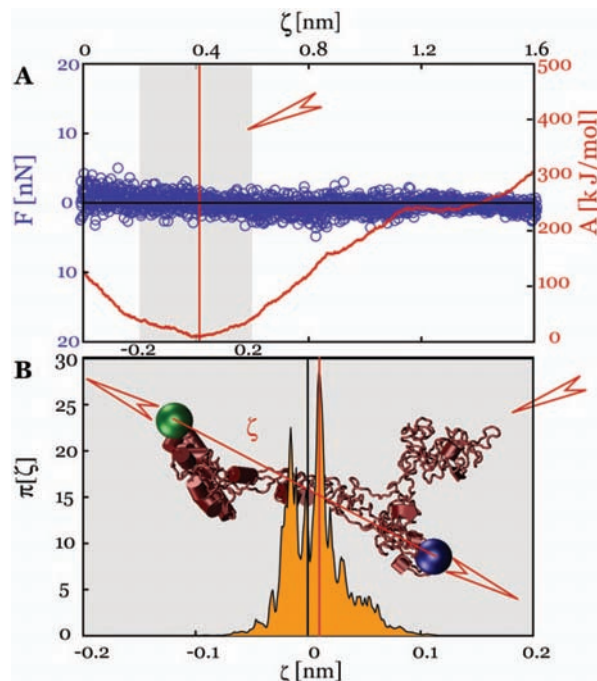


FIG. 11. (Color) (A) MF values indicated by open blue circles. The red full line gives the PMF. It has a minimum at $\zeta=0.4$ nm and a plateau at $\zeta=1.2$ nm. (B) $\pi(\zeta)$, line with yellow body, has been calculated within an interval of $|\zeta| \leq 0.02$, centered at the minimum of the free energy. The structure model of the protein is given as an insert, the reaction coordinate is indicated, and green and blue sphere indicate fixed regions.

ture. The free energy, in equilibrium it is equivalent to the potential of mean force, was calculated as function of the deformation. Parts of the protein were reversibly pulled in complementary directions, the MF was calculated, the results were joined. Hereof, the PMF and a local probability distribution $\pi[\zeta]$ were calculated, and the native form of SbsB was verified. Next, we shall present some examples of the analysis.

1. Structural analysis 1—N terminus

N terminus, Fig. 11: The N terminus was stretched along the x axis between the states $\zeta_1=0$ nm to $\zeta_2=1.6$ nm. The protein is elongated. There is a clear global free energy minimum at $\zeta=0.42$ nm, indicated by the red line. It is trivial that it corresponds to the global maximum of the local probability distribution. The average elongation $\bar{\zeta}$, however, is shifted left and is indicated by the black line. Likewise, the shape of the local energy and the corresponding local density probability distribution indicate a certain flexibility of the N terminus. Again, $\pi[\zeta]$ has nonzero values within an interval of $|\Delta\zeta| \leq 0.1$ nm.

2. Structural analysis 2—C-terminal domain 2

C-terminal domain 2, Fig. 12: The MF values are given by blue open circles, while the red line gives the PMF values. Two distinct shoulders are indicated by filled red arrows and correspond to metastable protein conformations. From the PMF, a local probability distribution $\pi[\zeta]$ has been cal-

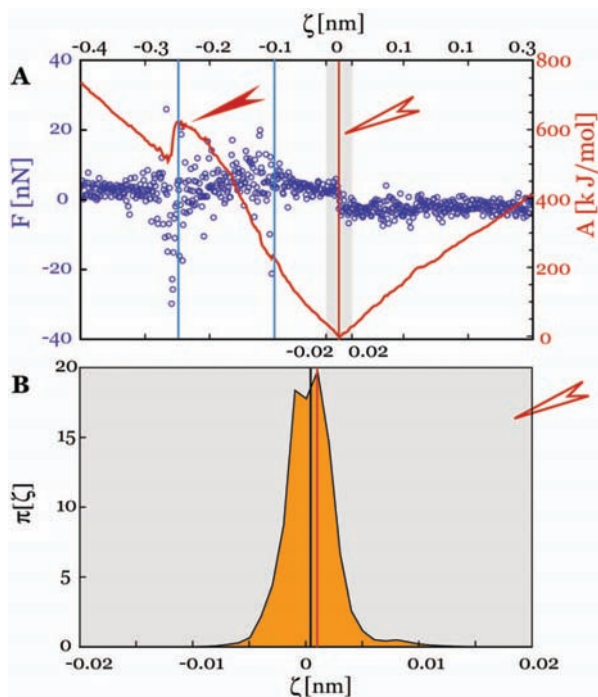


FIG. 12. (Color) (A) MF values are given by open blue circles. PMF given by the red full line jumps at $\zeta = -0.25$ nm to a lower value. Within the gray area, a local density probability distribution has been calculated. (B) $\pi(\zeta)$, line with yellow body, calculated from the PMF. The reaction coordinate was changed within an interval of $|\Delta\zeta| \leq 0.02$ nm.

culated over an interval of $|\Delta\zeta| \leq 0.1$ nm. The interval is centered at the global minimum of the PMF. The average elongation, indicated by the black line, is shifted left.

3. Structural analysis 3—C-terminal domains 2 and 3

Both C-terminal domains 2 and 3 were set in the $\pm y$ direction, Fig. 13. The blue circles indicate MF values, while the red line gives the PMF. The plot was centered at the global minimum of the PMF. In either direction, the free energy increases as a function of ζ . In the $+y$ direction, MF values are constant, and the PMF therefore increases nearly linear. For negative values of ζ , the MF starts scattering between $\zeta = -0.1$ nm and $\zeta = -0.25$ nm. Therein, the PMF develops a shoulder and jumps at $\zeta = -0.25$ to lower values, $\Delta A = 100$ kJ/mol. This jump indicates a metastable state of the protein. It is the most dense conformation of the two domains. This state may be useful while forming the S-layer lattice for providing the required flexibility for the binding of two single proteins. A local probability distribution, $\pi(\zeta)$, has been calculated over an interval of $|\Delta\zeta| \leq 0.02$ nm, that has been centered at the global minimum of the PMF. It is given by the graph with the yellow body. Trivial, the maximum of $\pi(\zeta)$ is located at zero. However, the average distance $\bar{\zeta}$ is shifted left and indicated by the black line.

H. The SbsB lattice

The protein SbsB assembles in an oblique lattice type that is characterized by the successively binding of single

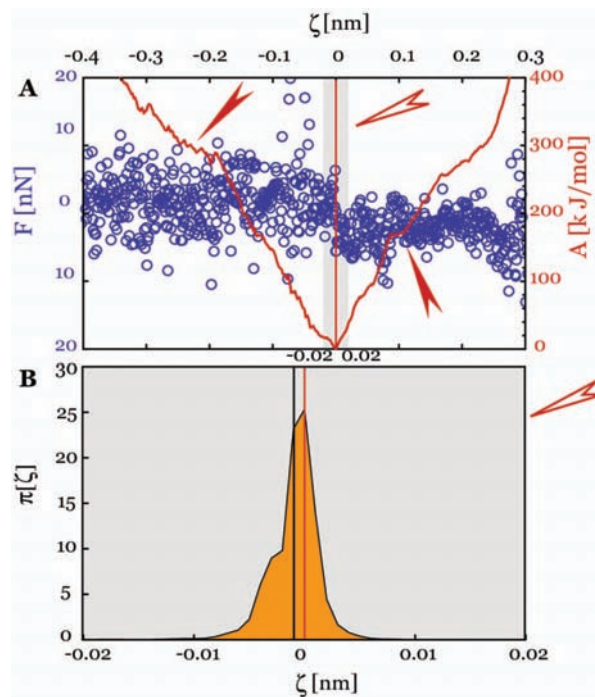


FIG. 13. (Color) (A) Calculated MF values are indicated by blue open circles. Integrated PMF profile is given by red full line. Full red arrow marks some plateau. The gray bar indicates an area, wherein a local density probability distribution $\pi(\zeta)$ has been calculated. (B) $\pi(\zeta)$ given by the yellow filled profile, calculated from the PMF within an elongation of $|\Delta\zeta| \leq 0.02$ nm.

SbsB proteins to each other, whereas there is no cluster building of prebonded proteins in contrast to square or hexagonal lattice types, where a cluster formation of three ($p3$), four ($p4$), or six ($p6$) proteins takes place primarily. Therefore, the graphical modeling of the SbsB lattice was done by joining single proteins with the program CINEMA4D. The N and the C termini of SbsB both contain pockets and inversions, which leads to the conclusion that these structural characteristics play a role in lattice formation. To find the optimal match of the monomers, a successive configuration search was performed graphically. Figure 14 illustrates the graphical model of the SbsB lattice. This model was created based on the optimal fitting of the pockets and inversions of the N and the C termini. Within this lattice model, the single monomers match perfectly to each other without overlapping. The inversion of the C-terminal domain 3 is complementary to the protuberance of domain 2. Additionally, the N-terminal binding pocket containing the SLH region and the SCWP-binding domain fit in the place, where domain 3 of one monomer binds domain 2 of the second monomer. In this way, a regular lattice occurs that contains periodic pores with a longer and a shorter side. The lattice constants of the SbsB lattice obtained via digital image reconstructions of rSbsB forms have a ratio of 1.3 ($a = 10.4$ nm, $b = 7.9$ nm). If using the measured distances of the SbsB model, the ratio of the graphical lattice model amounts to 1.29 ($a = 11.97$ nm, $b = 9.31$ nm) approximately, therefore, the dimensions of the modeled lattice seem to be in a right range. Figure 14 dis-

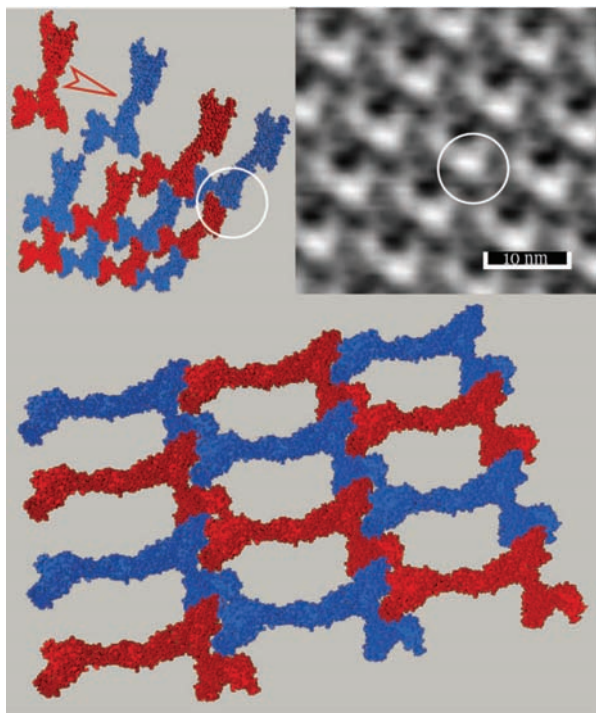


FIG. 14. (Color) Comparison of the graphical model of the SbsB lattice with a projection map of native SbsB. The white circles exemplarily mark higher regions in the lattice.

plays the analogy between the graphical model of the lattice and a projection map of native SbsB. The graphical model contains regular higher regions in the lattice due to the binding of the C-terminal domains 2 and 3. If compared with the projection map, the lattice topology is very similar concerning these higher regions. Moreover, the size and form of the pores look similar in both cases. Therefrom, both the perfect match of the single monomers and the topographic as well as the structural analogy between the model and the real lattice allow the assumption that the graphically modeled SbsB lattice is similar to the native lattice. We anticipate that the length difference of the base vectors between the native and the simulated structure can be fitted by assuming overlapping regions. Furthermore, the fitting of the inversions of the C-terminal domains in the lattice could verify the calculated tertiary structure of SbsB.

V. SUMMARY AND CONCLUSION

In this work, the tertiary structure of the S-layer protein SbsB of *Geobacillus stearothermophilus* PV72/p2 was predicted by using molecular dynamic simulations based on the amino acid sequence. In this way, the folding of the protein is determined only by thermodynamical and sterical parameters based on the amino acid composition. The obtained structure can be verified by computing the minimum of the free energy, that is, the potential of mean force (PMF), which is calculated based on the mean force that acts on a specific state of a system. This way of calculation is called the mean force method (MF method). Therefore, the amino acid sequence of SbsB was analyzed by performing a sequence

similarity search, a secondary structure prediction, and a domain prediction. The sequence of SbsB does not have any significant similarities to other proteins with clarified tertiary structure, whereas there are similarities to other S-layer proteins, especially concerning the N terminus (aa1–aa207), which leads to the suspicion that there are conserved domains within the S-layer protein family. Based on previous work on the structure-function relationship of SbsB, these conserved domains are SLH domains that are located at the N terminus and function as anchor for the protein in the cell surface. The C terminus of the protein is much more complex and could not be clarified by the sequence similarity search, whereas the domain prediction resulted in possible bacterial Ig-like group 2 and fibronectin type III domains, that are characterized by containing mainly antiparallel β -sheets, coils, and turns. This prediction goes in line with the secondary structure prediction that the N terminus mainly consists of α -helices as secondary structure elements and the C terminus of β -strands. Based on this information, the protein was split into eight structurally meaningful parts, whereas three domains for the N terminus (SLH domains) and five for the C-terminus, whereas two C-terminal domains could not be clarified by domain prediction. This splitting of SbsB is necessary to perform molecular dynamic simulations, because the length of SbsB avoids a conformation search including the whole protein. The protein would stay in the local energy minima, because it is too large to compute the whole phase space and, therefore, calculate the global free energy minimum. The C-terminal parts, that mainly consist of β -sheets, had to be premodeled by using a method called fold recognition. β -sheet formation is thermodynamically unfavorable because of the loss of configurational entropy. By creating 3D models of these parts, predetermined coordinates of the single atoms can be obtained that tend to form β -sheets. The original SbsB sequence was then matched to the modeled coordinates. With the eight parts of SbsB, molecular dynamic simulations were performed in water and in vacuum, whereas the folding process of the single parts could be computed. Subsequently, the parts were joined together and the tertiary structure of SbsB could be calculated. Using the ABF method implemented in the simulation software NAMD, the mean force values as well as the free energy of the protein were generated. These data provide the basis for calculating the PMF, which was done by integrating the mean force over a reaction coordinate ζ , that was determined based on the z coordinates of the atoms, that were used to manipulate the structure of the protein. In this way, the single domains of the protein were moved relatively to each other and, therefore, the free energies of the different conformations could be computed. The structural analysis of SbsB by the MF method lead to the result that the calculated tertiary structure of the protein represents a conformation lying in the global free energy minimum of the energy landscape, and is therefore the thermodynamically favorable state of SbsB in vacuum. There are various local energy minima and maxima, especially concerning the two main C-terminal domains 2 and 3. This indicates a certain flexibility within the protein that may play a role in lattice and, therefore, pore formation. Furthermore, the elongation of the N terminus

and the C-terminal domain 1 underlies a possible movement, which is evident regarding the function of the N terminus of attaching the cell surface. By using the MF method, the simulated native state of SbsB could be verified. The resulting tertiary structure of SbsB is L shaped, consists of three C-terminal domains, and the SLH region at the N terminus, which forms a binding pocket together with the SCWP-binding domain. The topography of the protein is very rough and there are a lot of inversions and pockets. Finally, the single SbsB monomers were joined together graphically to model an SbsB lattice. This model is based on the fitting of the inversions and pockets of the protein, whereas a good match was obtained by joining one C-terminal domain of one monomer with another C-terminal domain of a second monomer. The obtained model was compared with a projection map of native SbsB, and there is a great analogy especially concerning the topography of the lattice. In this way, the folded form of SbsB could be predicted and verified. This method may offer the possibility to specify the behavior of S-layer proteins, for example, by implementing various surfaces or conditions. Furthermore, the lattice itself could be simulated, whereas, in this case, a simplified model of the lattice should be taken to compute the characteristics of the lattice concerning pores, size, and functionalization. The results with molecular dynamic simulations and the MF method demonstrate that it is possible to determine the tertiary structure of a protein avoiding experimental problems in structure determination, which is the case concerning S-layer proteins.

Aside from the biochemical relevance of this work, the knowledge of the tertiary structure of S-layer proteins is also most important for the use of S-layers as patterning elements in life and nonlife sciences. For example, for the development of specific methods for the binding of biomolecules or nanoparticles, or for the precipitation of metal ions from solution, we need to know the location, orientation, and sterical accessibility of the involved binding sites in more detail. This is particularly true for S-layer fusion proteins where a specific functionality, such as the streptavidin binding domain, was fused to or inserted in the amino acid sequence. Up to our knowledge, this is the first published 3D-structure prediction of an S-layer protein. The results presented here will help us to address this question from a new perspective.

ACKNOWLEDGMENTS

Financial support from the Air Force Office of Scientific Research, USA (AFOSR, Project No. FA9550-06-1-0208) and Deutsche Forschungsgemeinschaft for the project "Adsorption aus wässrigen Lösungen", Az Fi 287/13-2, within the priority research program Molekular Modellierung und Simulation in der Verfahrenstechnik are gratefully acknowledged.

¹U. B. Sleytr, C. Huber, N. Ilk, D. Pum, B. Schuster, and E. M. Egelseer, *FEMS Microbiol. Lett.* **267**, 131 (2007).

²U. B. Sleytr, P. Messner, D. Pum, and M. Sára, *Angew. Chem., Int. Ed.* **38**, 1034 (1999).

- ³M. Sára, D. Pum, B. Schuster, and U. B. Sleytr, *J. Nanosci. Nanotechnol.* **10**, 1 (2005).
- ⁴U. B. Sleytr, M. Sára, P. Messner, and D. Pum, *J. Cell. Biochem.* **56**, 171 (1994).
- ⁵C. Huber, N. Ilk, D. Rünzler, E. M. Egelseer, S. Weigert, U. B. Sleytr, and M. Sára, *Mol. Microbiol.* **55**, 197 (2005).
- ⁶B. Kuen, A. Koch, E. Asenbauer, M. Sára, and W. Lubitz, *J. Bacteriol.* **179**, 1664 (1997).
- ⁷D. Rünzler, C. Huber, D. Moll, G. Köhler, and M. Sára, *J. Biol. Chem.* **279**, 5207 (2004).
- ⁸M. Sára, E. M. Egelseer, C. Dekitsch, and U. B. Sleytr, *J. Bacteriol.* **180**, 6780 (1998).
- ⁹D. Moll, Ph.D. thesis, University of Natural Resources and Applied Life Sciences, Vienna, 2002.
- ¹⁰D. T. Jones, *Protein Folds and Their Recognition from Sequence* (Oxford University Press, Oxford, 1996).
- ¹¹E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.* **156**, 472 (1989).
- ¹²A. Sergi, G. Ciccotti, M. Falconi, A. Desideri, and M. Ferrario, *J. Chem. Phys.* **116**, 6329 (2002).
- ¹³A. Ghoufi, P. Archiel, J.-P. Morel, N. Morel-Desrosiers, A. Boutin, and P. Malfeyt, *ChemPhysChem* **8**, 1648 (2007).
- ¹⁴W. Billes, B. Hegemark, M. Mecke, M. Wendland, and J. Fischer, *Langmuir* **19**, 10862 (2003).
- ¹⁵W. Billes, R. Tscheliessnig, L. Sobczak, M. Wendland, J. Fischer, and J. Kolafa, *Mol. Simul.* **33**, 655 (2007).
- ¹⁶R. Tscheliessnig, W. Billes, M. Wendland, J. Fischer, and J. Kolafa, *Mol. Simul.* **31**, 661 (2005).
- ¹⁷T. R. Forester and W. Smith, *J. Chem. Soc., Faraday Trans.* **93**, 3249 (1997).
- ¹⁸S. Consta, *Theor. Chem. Acc.* **116**, 373 (2006).
- ¹⁹G. Ciccotti, R. Kapral, and E. Vanden-Eijden, *ChemPhysChem* **6**, 1809 (2005).
- ²⁰D. Trzesniak, A.-P. E. Kunz, and W. F. van Gunsteren, *ChemPhysChem* **8**, 162 (2007).
- ²¹C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- ²²C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
- ²³J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, *Science* **296**, 1832 (2002).
- ²⁴D. A. Egolf, *Science* **296**, 296 (2002).
- ²⁵S. Park, F. Kahalili-Araghi, E. Tajkhorshid, and K. Schulten, *J. Chem. Phys.* **119**, 3559 (2003).
- ²⁶J. Henin and C. Chipot, *J. Chem. Phys.* **121**, 2904 (2004).
- ²⁷M. Fixman, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3050 (1974).
- ²⁸G. Ciccotti and M. Ferrario, *Mol. Simul.* **30**, 787 (2004).
- ²⁹E. Paci, G. Ciccotti, M. Ferrario, and R. Kapral, *Chem. Phys. Lett.* **176**, 581 (1991).
- ³⁰M. Sprik and G. Ciccotti, *J. Chem. Phys.* **109**, 7737 (1998).
- ³¹W. K. den Otter and W. J. Briels, *J. Chem. Phys.* **109**, 4139 (1998).
- ³²J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, *J. Comput. Chem.* **26**, 1782 (2005).
- ³³U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- ³⁴W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).
- ³⁵J. Verlet, *Phys. Rev.* **159**, 98 (1967).
- ³⁶N. Guex and M. C. Peitsch, *Electrophoresis* **18**, 2714 (1997).
- ³⁷L. A. Kelley, R. M. MacCallum, and M. J. Sternberg, *J. Mol. Biol.* **299**, 499 (2000).
- ³⁸C. Chipot and J. Hémin, *J. Chem. Phys.* **123**, 244906 (2005).
- ³⁹S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, *Nucleic Acids Res.* **25**, 3389 (1997).
- ⁴⁰Y. Luo, E. A. Frey, R. A. Pfuetzner, A. L. Creagh, D. G. Knoechel, C. A. Haynes, B. B. Finlay, and N. C. Strynadka, *Nature (London)* **405**, 1073 (2000).
- ⁴¹S. Yan, G. Gawlak, K. Makabe, V. Tereshko, A. Koide, and S. Koide, *J. Mol. Biol.* **368**, 230 (2007).
- ⁴²M. Pagni, V. Ioannidis, L. Cerutti, M. Zahn-Zabal, C. V. Jongeneel, and L. Falquet, *Nucleic Acids Res.* **32**, 332 (2004).
- ⁴³I. Letunic, R. R. Copley, B. Pils, S. Pinkert, J. Schultz, and P. Bork, *Nucleic Acids Res.* **34**, D257 (2006).
- ⁴⁴L. Y. Geer, M. Domrachev, D. J. Lipman, and S. H. Bryant, *Genome Res.* **12**, 1619 (2002).

- ⁴⁵A. Bateman, L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy, *Nucleic Acids Res.* **32**, D138 (2004).
- ⁴⁶J. L. Gattis, A. V. Washington, M. M. Chisholm, L. Quigley, A. Szyk, D. W. Mcvicar, and J. Lubkowski, *J. Biol. Chem.* **281**, 13396 (2006).
- ⁴⁷D. J. Leahy, I. Aukhil, and H. P. Erickson, *Proteins: Struct., Funct., Genet.* **19**, 48 (1994).
- ⁴⁸A.-S. Yang and B. Honig, *J. Mol. Biol.* **252**, 366 (1995).
- ⁴⁹C. M. Dobson, A. Sali, and M. Karplus, *Angew. Chem.* **110**, 908 (1999).

Surface layer protein characterization by small angle x-ray scattering and a fractal mean force concept: From protein structure to nanodisk assemblies

Christine Horejs,¹ Dietmar Pum,¹ Uwe B. Sleytr,¹ Herwig Peterlik,² Alois Jungbauer,³ and Rupert Tscheliessnig^{3,a)}

¹Department for Nanobiotechnology, University of Natural Resources and Applied Life Sciences, 1090 Vienna, Austria

²Faculty of Physics, University of Vienna, 1090 Vienna, Austria

³Institute for Applied Microbiology, University of Natural Resources and Applied Life Sciences, 1190 Vienna, Austria

(Received 2 June 2010; accepted 25 August 2010; published online 4 November 2010)

Surface layers (S-layers) are the most commonly observed cell surface structure of prokaryotic organisms. They are made up of proteins that spontaneously self-assemble into functional crystalline lattices in solution, on various solid surfaces, and interfaces. While classical experimental techniques failed to recover a complete structural model of an unmodified S-layer protein, small angle x-ray scattering (SAXS) provides an opportunity to study the structure of S-layer monomers in solution and of self-assembled two-dimensional sheets. For the protein under investigation we recently suggested an atomistic structural model by the use of molecular dynamics simulations. This structural model is now refined on the basis of SAXS data together with a fractal assembly approach. Here we show that a nondiluted critical system of proteins, which crystallize into monomolecular structures, might be analyzed by SAXS if protein-protein interactions are taken into account by relating a fractal local density distribution to a fractal local mean potential, which has to fulfill the Poisson equation. The present work demonstrates an important step into the elucidation of the structure of S-layers and offers a tool to analyze the structure of self-assembling systems in solution by means of SAXS and computer simulations. © 2010 American Institute of Physics.

[doi:10.1063/1.3489682]

I. INTRODUCTION

Surface layers (S-layers) are the most commonly observed cell surface structure of prokaryotic organisms. They are composed of planar assemblies of proteins, so-called S-layer proteins.^{1–3} The remarkable potential of this protein species lies in the capability to self-assemble into monomolecular crystalline lattices in suspension and on various solid substrates.^{4,5} Depending on the organism, different lattice symmetries could be observed: oblique ($p1$, $p2$), square ($p4$), and hexagonal ($p3$, $p6$).⁶ They exhibit pores of identical size and morphology and thus functional groups on the lattices are aligned in well-defined order and orientation. These characteristic features turn S-layers into a unique system for the controlled bottom-up assembly of functional supramolecular structures serving as building blocks and patterning elements in biomolecular construction kits. This opens up a broad spectrum of possible applications.^{7–13}

The basic building block of S-layers are proteins or glycoproteins with a molecular mass ranging from 40 to 200 kDa. There is only little knowledge about the structural details of these proteins due to the fact that they are far too large for investigations using NMR, and they are crystallizing into two-dimensional lattices avoiding the formation of

isotropic three-dimensional crystals required for x-ray crystallography.⁵ However, some information about the distribution of amino acids and parts of the structures of some s-layer protein species could be determined by introducing systematic point mutations, by producing diverse truncated forms of the proteins, or by the binding of special tags.^{14–18} A first atomistic model calculated by molecular dynamics simulations (MDS) was presented in our earlier work.¹⁹ Based on the amino acid sequence as well as experimental background we could compute a model of the tertiary structure of one S-layer protein, namely, the S-layer protein SbsB from *Geobacillus stearothermophilus*.^{20,21} As a continuative verification process of the calculated structure as well as a more profound understanding of structural details and potentials, we combine here the model from MDS and refine this model with the aid of small-angle x-ray scattering (SAXS) data of s-layer monomers and self-assemblies using the same s-layer protein as in the *in silico* experiments.

SAXS is a widely used method to obtain information on the structure of biological macromolecules in nearly physiological environments.^{22–28} Thereby a protein solution is exposed to x-rays, and the scattered intensity is recorded as a function of the scattering angle 2θ or the scattering vector Q . The x-ray pattern is determined by differences in the electron densities. Whereas from analytical methods only raw information on the size and shape of proteins such as spherical, cylindrical, or disklike is obtained, now numerical ap-

^{a)} Author to whom correspondence should be addressed. Electronic mail: rupert.tscheliessnig@boku.ac.at.

proaches are available, which allow to reconstruct the three-dimensional structure from one-dimensional scattering patterns.²⁵ A significant advantage of these methods is for example the enhanced insight into assembly and (un)folding processes of macromolecules.²⁵

Macromolecules in dilute solutions are usually analyzed using model functions, where no interaction of the molecules is required. However, the reconstruction of the structure of S-layer proteins faces the additional difficulty that the proteins crystallize. This affects the range of high Q -values, as the structure leads to Bragg reflections, as well as the range of small Q -values, because the proteins form large two-dimensional sheets. Thus, we separated the scattering intensity in the contributions by the proteins and the two-dimensional sheets, depending on the respective structure factor affecting the SAXS intensities from the form of the proteins. In a first step, the software programs GNOM²⁹ and DAMMIN²⁸ were applied to model the shape of the S-layer protein in solution from the experimental SAXS data at high Q -values, where the lower envelope (i.e., the signal without the Bragg reflections) was used. Driven by the reasonable size and similar shape to the structure obtained from MDS, we numerically calculated scattering profiles and refined the MD model on the basis of the SAXS data via Debye's formula for the form factor.³⁰ We extended Debye's model from the three-dimensional space to a general description in a space with a fractal dimension, as assembly of particles to clusters frequently leads to density distributions being best described by a fractal pair-pair correlation function.^{31–36} As the local density distribution and the mean potential of a system are correlated,³⁷ the mean potential is also fractal and it takes into account possible protein-protein interactions. It is also correlated with the tertiary structure of the protein itself. Therefore, the relation between mean potential and local densities was also formulated in a fractal space with the additional restriction that the local density distribution has to fulfill the Poisson equation. In a second step, the protein-protein interaction, which scatters due to the large dimension in the low Q -regime in reciprocal space, was described by objects with a disklike shape due to the two-dimensional nature of their assembly. To be consistent, the theoretical description in this Q -range was also formulated in fractal dimensions. By this, it was possible to describe the complete scattering curve and to derive the structure of the protein itself as well as its self-assembling behavior. The reconstruction process shown here is based on the combination of the results of the MDS with experimental SAXS measurements. Thereby the previously calculated model¹⁹ is refined by the SAXS data, but serves as a permanent corrective of the numerical reconstruction procedure.

II. THEORY

In general, the scattering intensity $I(Q)$ is the product of $P(Q)$, the form factor of particles, and $S(Q)$, the structure factor describing the arrangement of these particles. As S-layer proteins crystallize into monomolecular structures in solution, it is not possible to directly use conventional low resolution reconstruction procedures, which are available for

SAXS data from proteins in dilute solution.²⁵ The structure factor affects the signal in the high Q -regime in form of Bragg reflections and in the low Q -regime by the global arrangement of the proteins in form of S-layers being large in comparison to the size of the protein itself. Therefore, we separated the scattering data into two regimes, one at values $Q > 0.5 \text{ nm}^{-1}$ and one at values $Q \leq 0.5 \text{ nm}^{-1}$ for large objects in real space, which is indexed in the following by the symbol \gg . As either the protein is assembled from basic secondary structure elements as well as the proteins themselves self-assemble into large disklike layers, we extend the description of this system to fractal dimensions. The motivation is that particles or proteins that aggregate into clusters frequently adapt a density distribution that is best described by a fractal pair-pair correlation.^{31–36} Additionally, we take into account the relation between the local density distribution and the mean potential,^{37,38} where the local density distribution has to fulfill the Poisson equation. Therefore the mean potential does not only cause the structure of the monomers, but also takes into account protein-protein interactions. With these restrictions, we derive in a first step the equation to describe the scattering intensity of the protein itself in the high Q -range and in the second step for the self-assembly of the protein into disklike layers, accessible by SAXS data in the low Q -range. To be fully consistent, both Fourier transformations of local densities in the respective Q -range are formulated in the same functional fractal space.

The scattering intensity in SAXS arises from local density distributions, which are related to the mean potential of the system by $-\beta w(\zeta) \approx \log(\rho(\zeta))$,^{37,38} where $\beta = \frac{1}{k_B T}$ with k_B being the Boltzmann constant and T the temperature, ζ is a coordinate, ρ is the density, and w is the potential. Proteins are considered as critical due to their crystallization properties, where superstructures are assumed to appear on different spatial scales linked to each other. Due to the properties of clustering from small molecular units up to the arrangement of a large layer, a fractal mass distribution is assumed³⁶ and the scattering intensity $I(Q)$ is given by the Fourier transform of the autocorrelation function, which has a characteristic shape in this case

$$I(Q) = \mathcal{F}(\langle \rho(\zeta') \rho(\zeta) | \zeta' - \zeta | \rangle) [Q] = \mathcal{F}(\gamma_R \zeta^{3-D}) [Q], \quad (1)$$

$$= c D Q^{-D} R^{-D} \cos\left(\frac{D\pi}{2}\right) \Gamma(D-1). \quad (2)$$

D is the fractal dimension of the system, c a constant depending on experimental parameters such as beam intensity or electron density contrast, and γ_R is a proportionality constant. The fractal dimension is thus directly accessible in a double-logarithmic plot of intensities versus scattering vector Q .

A further requirement is that ionic solutions have to obey the Poisson equation. Solutions for systems with the assumption of a fractal charge distribution have already been developed by Lee and Hubbard³⁹ and Blender and Dietrich.⁴⁰

They linearized the density-mean potential relation and found with the help of a coordinate transformation that the mean potential takes the form of Bessel functions of the second kind. However their solution is nonlinear for the spatial coordinate ζ and thus an analytic Fourier transform is not possible.

A more general approach is to solve the Poisson equation for the mean potential by the use of the Green's function:⁴¹ $w(\zeta) = \int d^D \zeta' G_D(|\zeta - \zeta'|) \rho(\zeta')$ $= \int d^D \zeta' G_D(\zeta') \rho(|\zeta - \zeta'|)$ or as an iteration: $w(\zeta) \approx \int d^D \zeta' G_D(\zeta') \exp(-\beta w(|\zeta - \zeta'|))$.

Note that the fractal dimension in real space enters by the fractal Fourier transformation only, or in mathematical terms: $G_D(u) = \mathcal{F}_u^{-1}(\hat{G}(Q))$ and thus $\hat{G} = 1/(u^2 + Q^2)$.

$$\mathcal{F}(w(\zeta))[Q] \approx \mathcal{F}(\int d^D \zeta' G_D(\zeta') \langle \delta(\zeta') \delta(|\zeta' - \zeta|) \rangle)[Q], \quad (3)$$

$$\mathcal{F}(w(\zeta))[Q] \approx \mathcal{F}(\langle G_D(\zeta) \delta(\zeta) \rangle)[Q], \quad (4)$$

$$\approx 1/(2\pi)^D \int_0^\infty du^D \hat{G}(u) \mathcal{F}_\zeta(\delta(\zeta - u))[Q]. \quad (5)$$

$\hat{G}(u)$ is the Fourier transform of the Green's function and may be formally argued by polydispersity.⁴² The description of the scattering intensity of polydisperse systems is often weighted by a Schulz distribution function⁴³ to account for the size distribution of the particles. Here, the solution of $\hat{G}(u)$ is a Cauchy-Lorentz distribution, which was regarded as superior to the Schulz distribution⁴³ to describe the polydispersity of the system because it is directly related to the mean potential.

Rewriting Eq. (5) with the respective Fourier transforms explicitly leads to

$$\mathcal{F}(w(\zeta))[Q] \approx \frac{1}{(2\pi)^D} \int_0^\infty du u^{D-1} \frac{1}{\underbrace{u^2 + \kappa^2}_{f^D(Q)}} \frac{J_{D/2-1}(Qu)}{(Qu)^{D/2-1}} \quad (6)$$

$$\approx \frac{1}{(2\pi)^D} \kappa^{D/2-1} \frac{K_{D/2-1}(Q\kappa)}{(Q\kappa)^{D/2-1}}. \quad (7)$$

The measured x-ray intensities are then proportional to the form factor obtained from the square of the Fourier transform Eq. (6) by

$$I(Q) \propto P^D(Q) = |\mathcal{F}(Q)|^2. \quad (8)$$

In a three-dimensional space, the scattering intensity may be simplified to $I(Q) = P^3(Q)$ with the analytical form factor $P^3(Q) = \frac{1}{(2\pi)^3} \kappa^{1/2} \frac{K_{1/2}(Q\kappa)}{(Q\kappa)^{1/2}}$. The local densities obey the Poisson equation automatically here, which may be an advantage for a description of assembly processes driven by Coulomb forces. Equation (6) is an extension of the Debye formula³⁰ to a fractal space with arbitrary dimensions: Inserting $D=3$ in Eq. (6), setting the weight factor $f^D(Q)$ to 1 and replacing the integral term by a sum, Eq. (6) simplifies to the Debye expression³⁰ for the scattering intensity of scattering

centers with specific form factors $F^{i,j}$ and scattering densities $M^{i,j}$ at sites i and j , with C a normalization constant

$$I(Q) \approx P_{num}(Q) = \frac{1}{C} \sum_{i,j}^N \underbrace{M^i M^j F^i(Q) F^j(Q)}_{f^3(Q)} \frac{\sin(Q\zeta)}{Q\zeta} \quad (9)$$

where the index “num” indicates that this equation is basically used for the numerical reconstruction of structures from scattering curves.

The basic structural units, the proteins, self-assemble to form monomolecular layers. The scattering of this structure can be described in three dimensions with the form factor for a disk with a specific height H_\gg and a radius R_\gg ,⁴⁴ where the symbol \gg indicates that this dimension is now large in real space compared to the size of the protein itself. This form factor can be extended to a fractal description, which is motivated by the arrangement of the disks as clustered and not fully dense structures. Additionally, we take into account a specific orientation with respect to each other, which could arise from an interaction of the disks.

If we consider two fractal spaces that are orthogonal in respect to each other, then the volume element of a D -dimensional space⁴⁵ is given by $d\tau_\perp = \zeta_\perp^{D_\perp-1} \sin^{D_\perp-2} \theta \sin^{D_\perp-3}(\phi_1) \dots \sin(\phi_{D_\perp-3}) d\zeta_\perp d\theta \prod_i^{D_\perp-2} d\phi_i$. Its fractal complement is given by $d\tau_\parallel = \zeta_\parallel^{D_\parallel-1} \cos^{D_\parallel-2} \theta \sin^{D_\parallel-3}(\phi_1) \dots \sin(\phi_{D_\parallel-3}) d\zeta_\parallel d\theta \prod_i^{D_\parallel-2} d\phi_i$. Consequently the disklike structure may be reconstructed by a product of a δ -function $\delta(\zeta_\perp^2 - H_\gg^2)$ and a Heaviside Θ -function $\Theta(R_\gg - \zeta_\parallel)$. Therefore the local densities $\rho_\gg(\zeta)$ are given by $\rho_\gg(\zeta) = \delta(\zeta_\perp^2 - H_\gg^2) \Theta(R_\gg - \zeta_\parallel)$, with $\zeta_\perp = \zeta \sin \theta$ and $\zeta_\parallel = \zeta \cos \theta$. The D -dimensional Fourier transform is given by

$$\begin{aligned} S_\gg(Q) &= \mathcal{F}(\rho_\gg(\zeta))[Q] \\ &= \mathcal{F}_\perp(\delta(\zeta_\perp^2 - H_\gg^2))[Q] \mathcal{F}_\parallel(\Theta(R_\gg - \zeta_\parallel))[Q] \\ &= C_{\Gamma,D} \int_0^{\pi/2} \pi(\theta^*) \times \frac{J_{D_\perp/2-1}(QH_\gg \cos \theta^*)}{(QH_\gg \cos \theta^*)^{D_\perp/2-1}} \\ &\quad \times \frac{J_{D_\parallel/2}(QR_\gg \sin \theta^*)}{(QR_\gg \sin \theta^*)^{D_\parallel/2}} d(-\cos \theta^*) \end{aligned} \quad (10)$$

with a normalization constant $C_{\Gamma,D}$. Here we consider that the fractal disk dimensions D_\parallel and D_\perp are related by $D_\parallel = D_\perp - 1$ and applied two definitions of the Poisson integral representation of the Bessel function, $J_\nu(\zeta) = \pi^{-1/2} \Gamma^{-1}(\nu + 1/2) (\zeta/2)^\nu \int_{-\pi/2}^{\pi/2} \cos(\zeta \sin u) \cos^{2\nu} u du$ and $J_\nu(\zeta) = \pi^{-1/2} \Gamma^{-1}(\zeta/2)^\nu \int_0^\pi \sin(\zeta \sin u) \cos^{2\nu} u du$. This structure factor describes disks with radius R and height H , which are arranged with a certain angle distribution $\pi(\theta)$ with respect to each other.

The scattering intensity in the small Q -regime, i.e., in the regime of large objects in real space, is then

$$I(Q) = c_1 P^D(Q) \times S_>(Q) \times S_\gg(Q) + c_2. \quad (11)$$

c_2 is a small constant from a parasitic background arising from the background subtraction not allowing negative intensities to occur. The extension of the description of the scat-

tering intensity in Eq. (11) is valid due to the assumption that the mean potentials are independent and additive and thus superpose.

The whole scattering curve may finally be described by a form factor $P^D(Q)$ accounting for the structure of the protein monomers, a structure factor $S_>(Q)$ describing the arrangement of the monomers as self-assembled monomolecular sheets and a second structure factor $S_>>(Q)$, which takes into account the large-scale behavior of the two-dimensional sheets as nanodisks

$$P^D(Q) = \sum_i \frac{J_{D/2-1}(Qu_i)}{(Qu_i)^{D/2-1}}, \quad (12)$$

$$S_>(Q) = \left(\frac{1}{2\pi}\right)^D \int d^D u u^{D-1} \underbrace{\hat{G}_{u,\kappa} \rho_>(u)}_{K(u)} \frac{J_{D/2-1}(Qu)}{(Qu)^{D/2-1}} \quad (13)$$

$$S_>>(Q) = \left(\frac{1}{2\pi}\right)^D \int d^D R R^{D-1} \hat{G}_{R,\kappa} \mathcal{F}(\rho_>>(\zeta))[Q]. \quad (14)$$

Therein $P^D(Q)$ resembles the monomer structure, $S_>(Q)$ the self-assemblies and $S_>>(Q)$ represent the nano-discs.

In the case $\hat{G}_{R,\kappa}$ is a numerical delta function, and all disks are of equivalent size, $S_>>(Q)$ simplifies to $\mathcal{F}(\rho_>>(\zeta))[Q]$.

The fit parameters are the constants c_1 , c_2 , and κ . The fractal dimension D is taken from the decrease toward small Q -values in the double logarithmic plot. With these values fixed, the structure factor $S_>>$ is then additionally added for disks with a certain radius $R_>>$, a thickness $H_>>$, and a possible oriented distribution with respect to each other. Finally, the scattering intensity in the complete Q -range is used as experimental data for fitting the model.

III. MATERIALS AND METHODS

A. S-layer proteins

The s-layer protein SbsB from *Geobacillus stearothermophilus* PV72/p2 was isolated by default as previously described.^{46,47} After the cell wall preparation,⁴⁶ the proteins were extracted with 5M guanidine hydrochloride (pH 7.2) at room temperature for 30 min. Afterwards the cell wall fragments were removed by centrifugation (30 000 rpm, 30 min, 8 °C). The supernatant containing the proteins was dialyzed against 10 mM CaCl₂ for two hours at pH 5.9. A monomer solution was obtained by centrifuging for 30 min at 8 °C and 30 000 rpm. Afterward the monomer solution was concentrated by centrifugation in Amicon tubes (50 000 Da molecular weight cutoff) 35 min at 18 °C and 5000 rpm, to a final concentration of 3.3 mg/ml. The molecular weight of the s-layer protein is 98 kDa. The monomers are insoluble at these conditions.⁴⁸ For the self-assembly solution, we used the protein solution before centrifugation. At these conditions the s-layer protein SbsB self-assembles into two-dimensional sheets due to the higher monomer concentration, which was proved by electron microscopy studies.⁴⁷

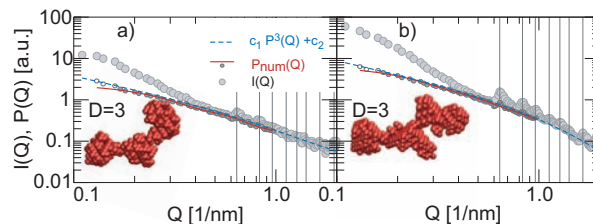


FIG. 1. Scattering data of the (a) monomeric solution and (b) the self-assemblies. Gray filled circles are experimental scattering intensities $I(Q)$ as a function of the scattering vector Q . The dashed blue line is the analytical fit from Eq. (11). Small circles are experimental intensities after division through a probable structure factor. Red lines give fits obtained by DAMMIN (Ref. 28). The corresponding structure models are given as insets. Vertical lines visualize Bragg reflections.

B. SAXS

SAXS was performed with Cu K α radiation from a rotating anode generator (Nanostar, BRUKER AXS, Karlsruhe, Germany) equipped with a pinhole camera and an area detector (VANTEC 2000, from BRUKER AXS). The SbsB monomer solution and the self-assemblies, respectively, were put into capillaries with 1 mm diameter and 10 μ m wall thickness (from Hilgenberg) and then sealed with polymeric caps, as the whole equipment operates in vacuum. The SAXS intensity patterns were taken at a sample to detector distance of 109 cm for 6 h. They were corrected for background scattering and then radially averaged to obtain the function $I(Q)$, where $Q = (4P/\lambda)\sin\theta$ is the scattering vector, 2θ is the angle between incident and diffracted beam, and $\lambda = 0.1542$ nm is the x-ray wavelength. The scattering intensities were normalized to each other in the Q -range between 2 and 2.5 nm⁻¹. In this region the scattering intensity is flat and dominated mainly by fluid scattering and an additional smaller contribution from the glass capillary. The scattering intensity from the solution without protein was then subtracted from the solution with protein. We cross-checked that this is consistent with the measured transmission of each sample.

The programs GNOM,²⁹ DAMMIN,²⁸ and DAMAVER⁴⁹ were used for a first evaluation of the scattering data. The software package GNOM is an indirect transform program for SAXS data processing. It calculates the distance or size distribution function from one-dimensional scattering curves for mono- and polydisperse systems, respectively. The software DAMMIN varies the number and position of subunits (usually spheres) of an object to match the distance distribution function using a simulated annealing minimization algorithm. This leads to a low resolution shape model for the object. In a Monte-Carlo algorithm, slightly different low resolution shape models are aligned and averaged by the software package DAMAVER⁴⁹ to result in the most probable shape.

IV. RESULTS AND DISCUSSION

The scattering intensity in the double logarithmic plot is shown in Fig. 1(a) for the monomer and Fig. 1(b) for the self-assemblies separately. Bragg reflections are traceable in both samples, which indicate that crystallization occurs even in the sample containing initially mainly monomers. Obvi-

ously, the proteins already start to self-assemble after a short period of time, a phenomenon, which has been observed before by means of light scattering.⁵⁰ If we compare scattering data at different times during the measurement, we observe the development of Bragg peaks, which suggest a dynamic crystallization process during the time of the measurement (see supplementary information⁵¹). The insoluble monomers at the given conditions start to self-assemble into two-dimensional lattices. In the first step to evaluate the data, the fractal dimension is set to $D=3$ in Eq. (6) and the lower envelope of the intensities in the high Q -range $Q > 0.5 \text{ nm}^{-1}$ is directly fitted with the form factor $P^3(Q)$. This corresponds to objects, built up densely in three-dimensional space, with local densities obeying the Poisson equation. Three fit parameters are required: Two of them are merely depending on the setup, a scaling factor c_1 corresponding to the total amount of the scattered volume (i.e., the thickness of the capillary and the intensity of the beam), and a small parasitic background c_2 . The main physical parameter is κ , which corresponds in the three-dimensional case to the radius of gyration and thus the size of the whole protein. The obtained fit values were $\kappa=6.3 \text{ nm}$ for the monomer (Fig. 1(a)) and $\kappa=6.8 \text{ nm}$ for the self-assemblies (Fig. 1(b)). This is in very good coincidence with the results from the MD simulations, which delivered a value of $\kappa=6.3 \text{ nm}$ ¹⁹ for the radius of gyration of the protein. Obviously, there is a strong increase in the intensities toward small Q -values (circles in Fig. 1), which is not described by the scattering curve of one single protein (solid lines in Fig. 1). This increased scattering intensity is therefore attributed to the structure factor from the self-assembly of proteins into crystalline layers, as any fit with conventional form factors would lead to a size about ten times larger than a single protein as determined by other methods such as transmission electron microscopy (TEM).⁴⁷ Thus, we concentrate first on the high Q -range, the shape and size of the protein, and in the second part on the self-assembled S-layers. Based on the calculated analytical form factors $P^3(Q)$ with the radius of gyration inserted from the MD model, we computed in a first rough approach a probable structure factor for the layered arrangement by $S(Q) \sim I(Q)/P^3(Q)$. Consequently we numerically reconstructed $S_{\text{num}}(Q)$ and divided the scattering intensities by this. These structure factor corrected data are shown in Fig. 1 as small circles up to a range of $Q = 1 \text{ nm}^{-1}$. For the fit of these data, the numerical software programs GNOM²⁹ and DAMMIN²⁸ were used to numerically reconstruct the three-dimensional structure. As starting configuration, prolate ellipsoids, slightly larger than the size of the MD model, with half axes $a=10.9 \text{ nm}$ and $b=8.6 \text{ nm}$, were chosen. 20 Monte-Carlo runs were performed for both the monomer and the self-assemblies and averaged with the use of DAMAVER.⁴⁹ The resulting bead models are given as inserts in Fig. 1. From this numerical reconstruction, the protein shape seems to differ between the monomer solution and the self-assemblies. For the proteins as part of self-assemblies a characteristic *bonelike* shape is obtained, which is very similar in shape and size to the MD model.¹⁹

This was the starting point to combine the results from

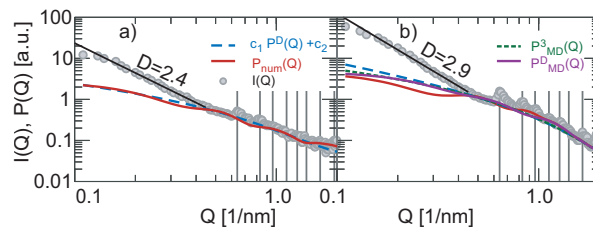


FIG. 2. SAXS data of the (a) monomeric solution and (b) the self-assemblies are given by gray filled circles and slopes for the fractal dimension are inserted as black lines: (a) $D=2.4$ and (b) $D=2.9$. The dashed blue line is the analytical fit P^D from Eq. (11) for (a) $P^{2.4}$ and (b) $P^{2.9}$, the dashed green line represents P^3_{MD} , and the full violet line $P^{2.9}_{\text{MD}}$. They are obtained by calculating the scattering intensity from the MD model. The full red line is the complete numerical reconstruction using a Monte-Carlo algorithm generalized to fractal dimensions. Vertical lines visualize Bragg reflections.

MD simulations with experimental SAXS measurements and a fractal theory describing local electron densities obeying the Poisson equation: The fractal dimension was estimated from the slope of the scattering intensities toward low Q -values. Then, the scattering intensities were described with Eq. (6) for fractal dimensions and a fit resulted in $\kappa=6.1 \text{ nm}$ and $D=2.4$ for the monomer in solution and $\kappa=4.7 \text{ nm}$ and $D=2.9$ for the self-assemblies, with the fit functions shown in Fig. 2 for the whole experimentally measured Q -range and in Fig. 3 for the range of high Q -values. Dashed lines represent fits with analytical functions and solid lines fits from the numerical reconstruction (which have the index *num* in all figures).

In fractal dimensions κ resembles more a characteristic value of the mass distribution and is not directly related to the radius of gyration anymore. In comparison to the three-dimensional form factor $P^3(Q)$, the fractal form factor $P^D(Q)$ (dashed blue line in Fig. 3) is able to follow some characteristic features of the SAXS curve, in particular the local minima better than the three-dimensional (3D) factor. We also compared both analytical form factors of the MD model (Fig. 3 green and magenta line), whereas $P^D(Q)$ follows the minima more precisely than $P^3(Q)$. In a final step, numerical fitting was extended to fractal dimensions with the help of Eq. (7), which is the analog of Debye's formula (Eq. (9)).

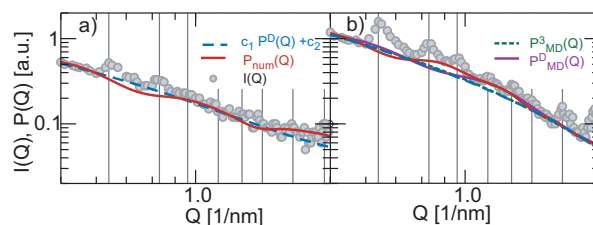


FIG. 3. SAXS data of the (a) monomeric solution and (b) the self-assemblies as in Fig. 2, but enlarged for the high Q -range to visualize how the different theories match the local minima. Slopes for the fractal dimension are inserted as black lines: (a) $D=2.4$ and (b) $D=2.9$. The dashed blue line is the analytical fit P^D from Eq. (11) for (a) $P^{2.4}$ and (b) $P^{2.9}$, the dashed green line represents P^3_{MD} , and the full violet line $P^{2.9}_{\text{MD}}$, either are based on the MD model. The full red line is the complete numerical reconstruction using a Monte-Carlo algorithm generalized to fractal dimensions. Vertical lines visualize Bragg reflections.

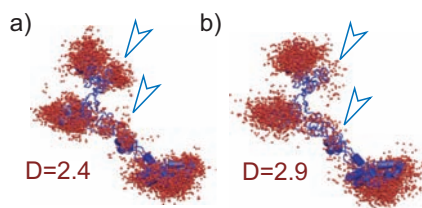


FIG. 4. (a) The numerical reconstruction of the SAXS scattering intensities in the monomeric sample indicates localized high electron densities (red beads) next to secondary structures such as β -sheets and α -helices. (b) In the samples containing self-assembled monomolecular sheets, high electron densities (red beads) are shifted compared to the monomer solution, visualized by arrows. The MD model is visible as blue cartoon.

The numerical reconstruction procedure of the 3D shape of the protein from one-dimensional scattering curves is then identical to available software programs such as SAXS3D⁵² or DAMMIN,²⁸ however, it uses basic functions to describe the fractal space instead of those for the 3D space. Therefore $f^D(Q)$ is set to one and local electron densities of the monomer are reconstructed by generating scattering patterns using a numerical algorithm. In order to restrict the problem, we closely connected the fitting procedure to the MD model, which serves as starting configuration for the electron densities. A sphere was created, which includes the entire atomistic protein model. Then, the numerical fitting algorithm changes positions of scatterers with the restriction that their position is within a size limit of 2 nm with respect to the MD model. The MD model is thus refined by the SAXS data, but serves as a permanent corrective of the numerical reconstruction procedure after each step. The final fitted scattering curves for the monomer and the self-assemblies are given by the red lines in Fig. 3 and the reconstructed electron contrasts in Fig. 4. For the monomers and the proteins as part of the lattice (Fig. 4) the red spheres represent the scatterers and thus the regions with the highest electron density. Whereas the refined structural model using the SAXS data is quite close to the original MD model and the highest electron contrast is still remaining in its vicinity, the protein as part of a self-assembly seems to be more extended and the higher electron contrast is shifted toward the termini of the protein. The differences are visualized by the arrows in Fig. 4. The different shape is interpreted in the way that the arrangement of the proteins in a two-dimensional (2D)-lattice leads to some overlapping of the proteins with neighboring proteins, which slightly shifts the protein shape in particular in these regions. In the case of the monomer, this is less pronounced, as the scattering intensity is an average of all configurations and the smaller degree of order leads to a smaller distortion of single proteins by neighboring ones.

For the analysis of the long-range order of the self-assembled structures, the experimental scattering intensity was divided by the scattering intensity of the reconstructed structure of the monomers and the self-assemblies, respectively, resulting in the structure factor $S_{\gg}^D(Q)$. This structure factor is fitted with Eq. (10) describing disks with a certain radius R , a height H , and an angle distribution $\pi(\theta)$ between them. The calculated structure factor of the disk together with the experimental values are shown in Fig. 5. A height of

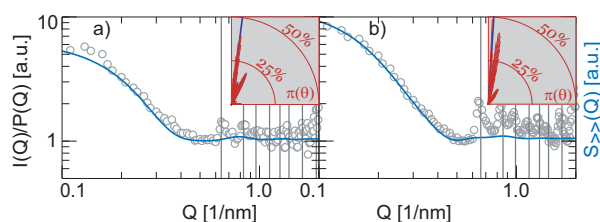


FIG. 5. The open circles give the scattering intensities divided by the form factor for the single proteins $I(Q)/P^D(Q)$, the blue line is the fit curve from Eq. (10) and describes the structure factor of disks with height $H=4.5$ nm and radii of (a) $R=20$ nm for the monomers and (b) $R=29$ nm for the self-assemblies. The insets give angle distributions of the disks with respect to each other from Eq. (10). Vertical lines visualize Bragg reflections.

$H=4.5$ nm and radii of $R=20$ nm for the monomers and $R=29$ nm for the self-assemblies were obtained. Therefore, even in the monomeric solution there do exist small oligomeric self-assemblies, the latter case. In the monomeric sample disklike S-layers are already formed, which are on the one hand smaller in radius and on the other hand more isotropically distributed in solution shown in the inset in Fig. 5. In the self-assembly sample, nanodisks self-assemble in a stacked order from single sheets with a height of 4.5 nm, which is in good agreement with results obtained by thin-section preparations of self-assembled sheets,⁵ and exhibit a fractal dimension close to three due to the formation of almost 3D structures.

The short-range order of the self-assembled S-layers is analyzed by the structure factor S_{\gg}^D using Eq. (13). Figure 6 gives the corresponding fitting function for $I(Q)/(P(Q)S_{\gg}^D(Q))$. We calculated the correlation function $\pi(u)$. For $\lim_{u \rightarrow \infty}$ it is an oscillating function, which indicates a critical system.

Whereas there is no long-range order of the sheets perpendicular to their axis, within the plane Bragg reflections were identified (Fig. 6) indicating an arrangement of the protein in a perfect monoclinic 2D-lattice with unit cell dimensions of $a=9.91 \pm 0.06$ nm, $b=7.66 \pm 0.05$ nm, and an angle between the axes of $81.1 \pm 0.1^\circ$. Bragg reflections were

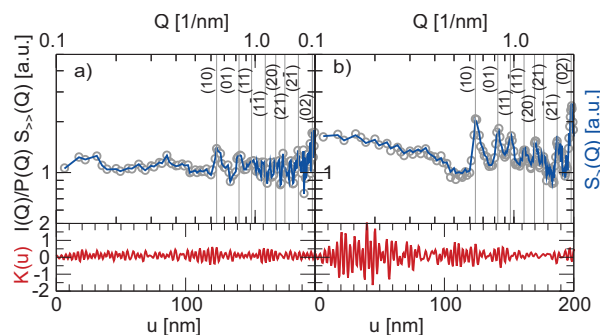


FIG. 6. Open circles are the scattering intensities divided by the structure factor at small Q -values, i.e., from the self-assembly into sheets, $I(Q)/(P(Q)S_{\gg}^D(Q))$. The vertical lines visualize Bragg reflections, being identified and attributed to a monoclinic 2D-lattice with unit cell dimensions of $a=9.91 \pm 0.06$ nm, $b=7.66 \pm 0.05$ nm, and an angle between the axes of $81.1 \pm 0.1^\circ$. The blue line gives fits for the structure factor S_{\gg}^D , while inserts give the correlation function $K(u)$ from Eq. (13). For $\lim_{u \rightarrow \infty}$ it is an oscillating function.

determined as described elsewhere,⁵³ where for a monoclinic lattice the third dimension in real space approaches infinity. The error results from the error in the determination of the sample position (i.e., the finite thickness of the capillary), the monochromaticity of the beam and the fit error. This is in perfect agreement with data obtained from transmission electron microscopy $a=10.4$ nm, $b=7.9$ nm, and $\alpha=81^\circ$,^{5,47} however, the SAXS data are much more precise in this case due to high resolution in reciprocal space.

V. CONCLUSIONS

Experimental SAXS data for the S-layer protein SbsB in solution were measured, where Bragg reflections in the high Q -range precisely determined the structural arrangement of the protein units in a 2D monoclinic lattice. In the low Q -range, scattering arises from the arrangement as 2D large objects with a disklike shape. The data were separated into these two regions and the structure was described as self-assembled units in a fractal space. Analytical solutions fulfilling the Poisson equation were developed and applied in the first step. In the second step, a reconstruction of the structure from SAXS data was performed completely numerically, where previous results from MD served both as starting configuration and as control to restrict the maximum allowed deviation of the reconstruction. Whereas the electron densities remained in the vicinity of the original MD model for the monomer, in the case of the assembled proteins electron densities were shifted, which is interpreted as overlapping zones of assembled proteins. The presence of Bragg reflections in both samples—though not equally strongly visible in the x-ray pattern—as well as different size magnitudes of the S-layers in the monomeric and the self-assemblies solution show that even in the monomeric sample small oligomeric self-assemblies are formed. We assumed a critical system due to the remarkable self-assembling behavior of the system: Even when monomeric samples are highly diluted, oligomeric small crystals are formed, which has to be taken into account into the fitting of the SAXS intensity data. By splitting the whole Q -range into two different regimes and by describing the S-layer system in a fractal space, which ensures that the analytical solutions fulfill the Poisson equation, we could reconstruct the electron densities of an S-layer protein, which confirms the previously predicted MD model and indicates some structural changes due to crystallization. However, even in the case of self-assembling into a monomolecular lattice, the shape of the protein remains remarkably stable. Electron densities are only slightly shifted toward the ends which might be explained by possible overlaps with neighboring proteins in the lattice.

This work shows that a nondiluted, critical system of proteins, which crystallize into monomolecular structures, might be analyzed with SAXS, if protein-protein interactions are taken into account by relating a fractal local density distribution to a fractal mean potential, which has to fulfill the Poisson equation. This mean potential governs the monomer structure and describes interactions in the system. The assumption of a fractal mass distribution was already successfully applied for analyzing SAXS data of particles that ag-

gregate to clusters.^{31–36} By adapting the fractal approach to proteins, both the shape of single proteins as well as their self-assembly behavior can be described.

ACKNOWLEDGMENTS

The authors thank Jacqueline Friedmann for the preparation of the protein samples. C.H. holds a DOC-fORTE fellowship of the Austrian Academy of Sciences. This project has been supported by the Air Force project FA9550-09-1-0342. This work was carried out in the Austrian Center of Biopharmaceutical Technology which is a Competence Center within the Kind-Knet program funded by the Austrian Federal Ministry of Economics and Labour (FWF) and the provinces Vienna and Tyrol.

¹U. B. Sleytr, *Nature* **257**, 400 (1975).

²U. B. Sleytr, P. Messner, D. Pum, and M. Sara, *Angew. Chem., Int. Ed.* **38**, 1034 (1999).

³M. Sara and U. B. Sleytr, *J. Bacteriol.* **182**, 859 (2000).

⁴E. S. Györfvay, E. Stein, D. Pum, and U. B. Sleytr, *J. Microsc.* **212**, 300 (2003).

⁵D. Pum, M. Sara, and U. B. Sleytr, *J. Bacteriol.* **171**, 5296 (1989).

⁶U. B. Sleytr, C. Huber, N. Ilk, D. Pum, B. Schuster, and E. M. Egelsser, *FEMS Microbiol. Lett.* **267**, 131 (2007).

⁷B. Kainz, K. Steiner, M. Möller, D. Pum, C. Schäffer, U. B. Sleytr, and J. L. Toca-Herrera, *Biomacromolecules* **11**, 207 (2010).

⁸S. R. Scheicher, B. Kainz, S. Köstler, M. Suppan, A. Bizzarri, D. Pum, U. B. Sleytr, and V. Ribitsch, *Biosens. Bioelectron.* **25**, 797 (2009).

⁹J. Tang, A. Ebner, B. Kraxberger, M. Leitner, A. Hykollari, C. Kephlinger, C. Grunwald, H. J. Gruber, R. Tampe, U. B. Sleytr, N. Ilk, and P. Hinterdorfer, *J. Struct. Biol.* **168**, 217 (2009).

¹⁰H. Tschiggerl, J. L. Casey, K. Parisi, M. Foley, and U. B. Sleytr, *Bioconjugate Chem.* **19**, 860 (2008).

¹¹N. Ilk, S. Küpcü, G. Moncayo, S. Klimt, R. C. Ecker, R. Hofer-Warbinek, E. M. Egelsser, U. B. Sleytr, and M. Sara, *Biochem. J.* **379**, 441 (2004).

¹²N. Ilk, C. Völlenkne, E. M. Egelsser, A. Breitwieser, U. B. Sleytr, and M. Sara, *Appl. Environ. Microbiol.* **68**, 3251 (2002).

¹³C. Schäffer, R. Novotny, S. Küpcü, S. Zayni, A. Scheberl, J. Friedmann, U. B. Sleytr, and P. Messner, *Small* **3**, 1549 (2007).

¹⁴T. Pavkov, M. Oberer, E. M. Egelsser, M. Sara, U. B. Sleytr, and W. Keller, *Acta Crystallogr. D* **59**, 1466 (2003).

¹⁵T. Pavkov, E. M. Egelsser, M. Tesarz, D. I. Svergun, U. B. Sleytr, and W. Keller, *Structure* **16**, 1226 (2008).

¹⁶S. Howorka, M. Sara, Y. Wang, B. Kuen, U. B. Sleytr, W. Lubitz, and H. Bayley, *J. Biol. Chem.* **275**, 37876 (2000).

¹⁷H. Kinns, H. Badelt-Lichtblau, E. M. Egelsser, U. B. Sleytr, and S. Howorka, *J. Mol. Biol.* **395**, 742 (2010).

¹⁸P. P. Fagan, D. Albesa-Jove, O. Qazi, D. I. Svergun, K. A. Brown, and N. F. Fairweather, *Mol. Microbiol.* **71**, 1308 (2009).

¹⁹C. Horejs, D. Pum, U. B. Sleytr, and R. Tscheliessnig, *J. Chem. Phys.* **128**, 065106 (2008).

²⁰B. Kuen, A. Koch, E. Asenbauer, M. Sara, and W. Lubitz, *J. Bacteriol.* **179**, 1664 (1997).

²¹D. Rünzler, C. Huber, D. Moll, G. Köhler, and M. Sara, *J. Biol. Chem.* **279**, 5207 (2003).

²²O. Glatter and O. Kratky, *Small-Angle X-Ray Scattering* (Academic, London, 1982).

²³J. Lipfert and S. Doniach, *Annu. Rev. Biophys. Biomol. Struct.* **36**, 307 (2007).

²⁴D. I. Svergun and M. H. J. Koch, *Rep. Prog. Phys.* **66**, 1735 (2003).

²⁵D. I. Svergun and M. H. J. Koch, *Curr. Opin. Struct. Biol.* **12**, 654 (2002).

²⁶C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, *Q. Rev. Biophys.* **40**, 191 (2007).

²⁷D. I. Svergun, M. V. Petoukhov, and M. H. J. Koch, *Biophys. J.* **80**, 2946 (2001).

²⁸D. I. Svergun, *Biophys. J.* **76**, 2879 (1999).

²⁹D. I. Svergun, A. V. Semenyuk, and L. A. Feigin, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **44**, 244 (1988).

- ³⁰P. Debye, *Ann. Phys.* **351**, 809 (1915).
- ³¹M. Daniel, S. Baskar, and M. M. Latha, *Phys. Scr.* **60**, 270 (1999).
- ³²S.-H. Chen and J. Teixeira, *Phys. Rev. Lett.* **57**, 2583 (1986).
- ³³A. C. Castellano, M. Barteri, A. Bianconi, E. Borghi, L. Cassiano, M. Castagnola, S. Della Longa, and A. La Monaca, *Biophys. J.* **64**, 520 (1993).
- ³⁴T. Freltoft, J. K. Kjems, and S. K. Sinha, *Phys. Rev. B* **33**, 269 (1986).
- ³⁵G. Beaucage, *J. Appl. Crystallogr.* **29**, 134 (1996).
- ³⁶M. Sahimi, *Rev. Mod. Phys.* **65**, 1393 (1993).
- ³⁷D. Chandler, *J. Chem. Phys.* **68**, 2959 (1978).
- ³⁸C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- ³⁹S. H. Lee and J. B. Hubbard, *J. Chem. Phys.* **98**, 1504 (1993).
- ⁴⁰R. Blender and W. Dietrich, *J. Phys. A* **19**, L785 (1986).
- ⁴¹S. Vembu, *Q. J. Math.* **12**, 165 (1961).
- ⁴²M. Kotlarchyk and S.-H. Chen, *J. Chem. Phys.* **79**, 2461 (1983).
- ⁴³M. J. Wagner, R. Krause, A. R. Rennie, and B. D'Aguzzo, *J. Chem. Phys.* **95**, 494 (1991).
- ⁴⁴A. Guinier and G. Fournet, *Small-Angle Scattering of X-Rays* (Wiley, New York, 1955).
- ⁴⁵L. E. Blumenson, *Am. Math. Monthly* **67**, 63 (1960).
- ⁴⁶U. B. Sleytr, M. Sara, Z. Küpcü, and P. Messner, *Arch. Microbiol.* **146**, 19 (1986).
- ⁴⁷D. Pum, M. Weinhandl, C. Hödl, and U. B. Sleytr, *J. Bacteriol.* **175**, 2762 (1993).
- ⁴⁸D. Moll, C. Huber, B. Schlegel, D. Pum, U. B. Sleytr, and M. Sara, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14646 (2002).
- ⁴⁹V. V. Volkov and D. I. Svergun, *J. Appl. Crystallogr.* **36**, 860 (2003).
- ⁵⁰R. Jaenicke, R. Welsch, M. Sara, and U. B. Sleytr, *Biol. Chem. Hoppe Seyler* **366**, 633 (1985).
- ⁵¹See supplementary material at <http://dx.doi.org/10.1063/1.3489682> for the development of the monomer sample during the measurement.
- ⁵²D. Walther, F. E. Cohen, and S. Doniach, *J. Appl. Crystallogr.* **33**, 350 (2000).
- ⁵³I. Solomonov, M. J. Weygand, K. Kjaer, H. Rapaport, and L. Leiserowitz, *Biophys. J.* **88**, 1809 (2005).

SUPPLEMENTARY MATERIAL

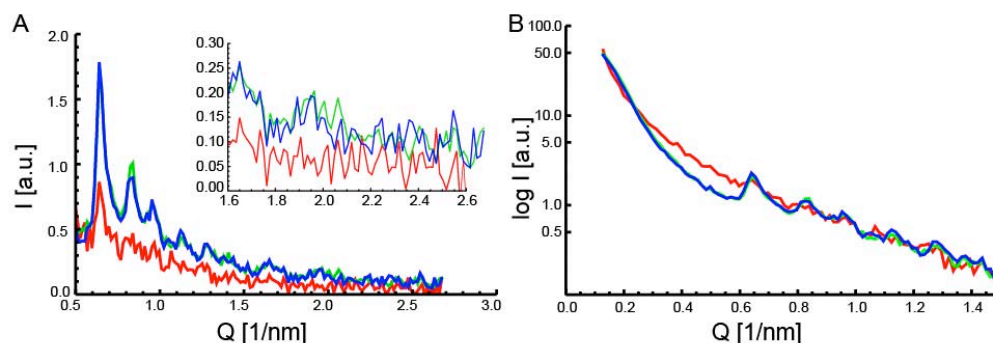


Fig. S1: Time evolution of the monomer sample during the measurement. The scattering intensity as a function of the scattering vector Q is shown. The system is critical and there is a clear development of Bragg peaks during the measurement. The insoluble monomers at the given conditions start to self-assemble into two-dimensional lattices. SAXS measurements were performed for one hour (red line, after 1 hour) and six hours (green line, after 34 hours and blue line, after 55 hours). A: Linear scale after background subtraction from parasitic pinhole scattering and further subtraction of a power-law background towards small q -values to better visualize the diffraction peaks. B: Logarithmic scale after background subtraction from parasitic pinhole scattering.

Atomistic Structure of Monomolecular Surface Layer Self-Assemblies: Toward Functionalized Nanostructures

Christine Horejs,[†] Harald Gollner,[†] Dietmar Pum,[†] Uwe B. Sleytr,[†] Herwig Peterlik,[‡] Alois Jungbauer,[§] and Rupert Tscheliessnig^{§,*}

[†]Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, [‡]Faculty of Physics, University of Vienna, Vienna, Austria, and [§]Austrian Centre of Industrial Biotechnology c/o Institute for Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

Surface layers (S-layers) represent an outstanding self-assembly system based on protein subunits, which form the crystalline outermost cell envelope of a great variety of prokaryotic cells.¹ These S-layer proteins self-assemble into monomolecular lattices with symmetries ranging from p1 to p6, exhibiting defined pores and a highly reproducible large-scale order.^{2–4} They have been successfully used as the basic unit for molecular construction kits because of their remarkable potential to self-assemble not only in their natural environment but also in solution, on various solid substrates and on lipids.^{3,5} A great number of S-layer-carrying organisms has been identified so far, whereas no atomistic structure of a single unmodified S-layer protein could be experimentally determined up to now because they are too large for NMR and they do not crystallize into isotropic three-dimensional crystals as required for X-ray crystallography. Recently, more information on structural details and self-assembly pathways could be gathered.^{6–15} However, the defined binding of molecules and nanoparticles on S-layers requires detailed structural information on the underlying S-layer proteins at an amino acid level. Here we present a combination of molecular modeling, small-angle X-ray scattering (SAXS), and transmission electron microscopy (TEM) to reveal the three-dimensional structure of one unit cell of the S-layer protein SbpA from *Bacillus sphaericus* CCM2177.¹⁶ SbpA consists of 1268 amino acids and self-assembles into a square lattice structure (p4). This S-layer protein has been extensively investigated in order to be used for nanobiotechnological applications at the nanoscale.^{17–24} In order to model the structure at an amino acid

ABSTRACT The concept of self-assembly is one of the most promising strategies for the creation of defined nanostructures and therefore became an essential part of nanotechnology for the controlled bottom-up design of nanoscale structures. Surface layers (S-layers), which represent the cell envelope of a great variety of prokaryotic cells, show outstanding self-assembly features *in vitro* and have been successfully used as the basic matrix for molecular construction kits. Here we present the three-dimensional structure of an S-layer lattice based on tetrameric unit cells, which will help to facilitate the directed binding of various molecules on the S-layer lattice, thereby creating functional nanoarrays for applications in nanobiotechnology. Our work demonstrates the successful combination of computer simulations, electron microscopy (TEM), and small-angle X-ray scattering (SAXS) as a tool for the investigation of the structure of self-assembling or aggregating proteins, which cannot be determined by X-ray crystallography. To the best of our knowledge, this is the first structural model at an amino acid level of an S-layer unit cell that exhibits p4 lattice symmetry.

KEYWORDS: self-assembly · SAXS · TEM · protein structure · computer simulations · nanoarrays

level of one SbpA unit cell consisting of four monomers, we combined three different approaches. In a first step, we defined possible domains of the protein based on information on the self-assembly behavior of truncated and modified recombinant forms of the protein. We already successfully applied this approach to model the tertiary structure of another S-layer protein.¹⁴ We defined seven individual domains of the protein, and on the basis of secondary structure predictions and structural homologies, we premodeled each domain implementing possible secondary structure elements. Consequently, we equilibrated each domain in a water sphere, joined them together, and minimized the whole structure while keeping every domain restraint using molecular dynamic simulations. In a second step, a three-dimensional density model has been calculated by performing tilting experiments with a transmission electron microscope. In a third step, small-angle

* Address correspondence to rupert.tscheliessnig@boku.ac.at.

Received for review December 23, 2010 and accepted February 14, 2011.

Published online March 04, 2011
10.1021/nn1035729

© 2011 American Chemical Society

X-ray scattering (SAXS) experiments were performed together with a theory of a fractal mean force potential, which can be adequately used to describe the behavior of S-layers in solution, as we showed in our earlier work.¹⁵ These studies resulted in a distribution of electron densities within an S-layer unit cell. We finally merged these data, the model of one monomer obtained by molecular dynamic simulations, and the density profile as well as the scattering distribution of one unit cell and modeled the three-dimensional structure of an SbpA unit cell at an amino acid level. The information of the location of individual amino acids on the inner and outer surface of the lattice as well as within the pores will facilitate the directed modification of this S-layer protein and the directed binding of molecules and nanoparticles on the lattice for the application of this self-assembly system in nanotechnology.

RESULTS AND DISCUSSION

The S-layer protein SbpA has aroused interest due to the fact that this protein self-assembles into lattices exhibiting a p4 lattice symmetry. The four monomers, which form one unit cell of the lattice, may perfectly serve as a matrix for the generation of functional nanoarrays.²⁵ The N-terminal part possesses three so-called S-layer homologous (SLH) domains, which is a common structural motive in this class of proteins that is mainly made up of α -helices.²⁶ The SLH domains are responsible for anchoring the proteins in the underlying cell wall. The successful self-assembly of SbpA demands the presence of bivalent cations,²⁶ which explains the structural homology of this protein to calcium binding proteins, which could be found by performing homology searches using various different bioinformatic tools. Interestingly, the lattice symmetry can be changed from p4 to p1 if 237 C-terminal amino acids are truncated. If additionally another 113 amino acid residues are removed, the proteins lose the ability to self-assemble.²⁷ Apparently a structural change of the monomers due to a shortening of the C-terminal domain leads to a loss of the intrinsic function. We also performed secondary structure predictions revealing S-layer proteins' common distribution of secondary structure elements, a mainly α -helical N-terminal region and mainly β -sheets in the central and C-terminal part. On the basis of all this information, we defined the following seven domains: three SLH domains (aa1–aa210, aa211–329, aa330–aa458), one central domain showing structural homology to calcium binding proteins (aa459–aa639), and three C-terminal domains showing homology to Ig-like proteins, which are mainly made up of β -sheets (aa640–aa888, aa889–aa1001, aa1002–aa1238). As a next step, we used the online algorithm PHYRE to model each domain based on fold recognition.²⁸ Domains one, two, and three (SLH domains) are mainly made up of

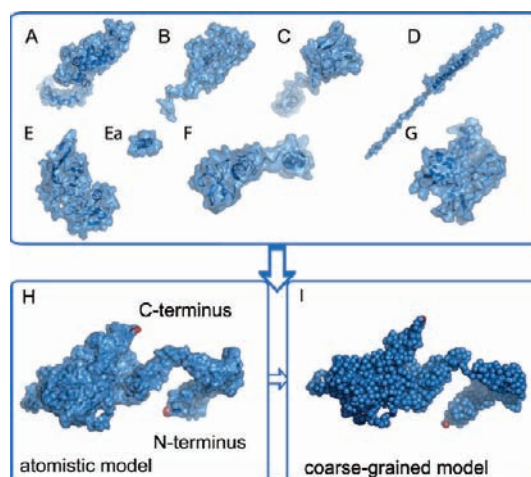


Figure 1. Results of the molecular dynamics simulations of the S-layer protein SbpA. The protein was first split into structurally meaningful domains, which were premodeled using fold recognition to obtain secondary structure elements. These domains were equilibrated in water. The results are given in A–G. (A–C) Three SLH (S-layer homologous) domains located at the N-terminal region of the protein, (D) central domain showing structural homology to Ca^{2+} -binding proteins, (E–G) domains showing structural homology to Ig-like domains. Without domains F and G, the protein loses its ability to self-assemble, and without domain G, the self-assembled products show p1 rather than p4 lattice symmetry.²⁷ (H) Intermediate structure of the whole protein after the single domains were joined and simulated in vacuum, where the secondary structures were kept rigid, ions were added, and external forces enabled. The termini are marked in red. (I) Coarse-grained protein structure based on the atomistic structure shown in H. Every amino acid is represented by a single bead. Consecutive beads are linked by a harmonic potential.

α -helices and coils as expected. Domains three, five, and seven exhibit parallel and antiparallel β -sheets, and domains four and six could not be modeled by fold recognition and show a rather elongated structure. Every premodeled domain was now equilibrated in a water sphere using molecular dynamics simulations. Figure 1A–G shows the results for every single domain. The three potential SLH domains (Figure 1A–C) are mainly made up of α -helices (especially the first SLH domain, Figure 1A) and random coils after equilibration in water. In clear contrast to our previous work,¹⁴ we could not equilibrate all of these domains in water. Even after a considerable large amount of production runs, two domains lack a sensible secondary structure (Figure 1D,E). The central domain (Figure 1D), which was premodeled using homologies to Ca^{2+} -binding proteins, is supposed to consist of mainly β -sheet structures. β -Sheets are rather complicated secondary structure elements, which may fold due to interactions between distant amino acid residues in the primary structure, which is why they hardly form in the framework of short molecular dynamics simulations. In contrast, those domains, which show homologies to Ig-like domains, show β -sheets, α -helices, and random coil

structures (Figure 1E–G) as expected by the secondary structure predictions.

All domains were now linked together, and an inverse steered molecular dynamics simulation was performed in vacuum. As the protein is charged, Na^+ and Cl^- ions were added to compensate the net charge. Moreover, the anticipated secondary structures, the α -helices, β -sheets, and coils, were kept rigid. In addition, all domains were linked to one another by an external force. This external force is physically motivated by the force that the absent water molecules would impose on the protein. We realize these potentials by harmonic springs, which are necessary to initiate the refolding process. However, the resulting structure, as shown in Figure 1H, represents an intermediate conformation. The resulting model was consequently used to reconstruct the three-dimensional unit cell by rational design based on the density distribution as obtained by electron microscopy studies.

The transmission electron microscope offers a great possibility to determine the density distribution and thereby a three-dimensional reconstruction of an S-layer unit cell by performing tilting studies with a negatively stained self-assembly product due to the symmetry of the S-layer lattice. This method is based on the projection theorem, which states that the two-dimensional Fourier transform of a plane projection of a three-dimensional density distribution is identical to the corresponding central section of the three-dimensional transform normal to the direction of view. Therefore, the three-dimensional transform can be built up section by section and consequently reconstructed by inverse Fourier transform.²⁹ The phases of the diffraction pattern are directly accessible from an image. Figure 2 shows the results of a tilting series with an SbpA self-assembly product, where the single sections used for the density reconstruction are shown. On the basis of these sections, a density model of the unit cell can be rebuilt, as shown in Figure 2B. This three-dimensional model was consequently used to fit the structural model of one SbpA monomer into the unit cell. Therefore, four monomers are arranged in a way to cover the entire unit cell, where overlaps are avoided. As this arrangement is based on an intermediate structural model, the reconstructed unit cell has to be equilibrated, which demands a coarse-graining of the whole structure (Figure 1I). Each amino acid is represented by a single bead. We introduce three different potentials. First, all consecutive beads are linked by a harmonic potential. As all seven domains but the outer five amino acids of each are kept rigid, the harmonic potential is of minor relevance. The interactions of the amino acids are controlled by two types of pair–pair potentials: an attractive screened Coulomb potential and an associative Gaussian potential. While the first potential type is a consequence of electrostatic forces,

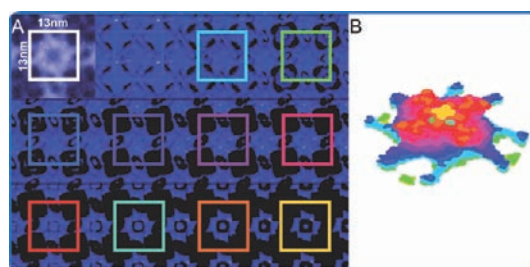


Figure 2. Reconstruction of the three-dimensional density distribution of one SbpA unit cell calculated by inverse Fourier transform of single sections of SbpA self-assembly products. A transmission electron microscope was used to obtain sections of the unit cell by performing tilting measurements. (A) Single sections of an SbpA unit cell. In the top left corner, an electron micrograph of a whole unit cell is shown. The density boundaries are systematically increased (starting in the top left corner). The sections provide the basis for the unit cell density distribution, which is shown by the colored squares. Every color reflects one single section of the unit cell in two dimensions. A superposition of all sections results in the three-dimensional unit cell as shown in B. (B) Three-dimensional density distribution as used for the reconstruction of the unit cell based on the coarse-grained model obtained from molecular dynamics simulations.

the second, in principle, should enable the formation of particular secondary structures. The coarse-grained model is consequently used to model a coarse-grained unit cell. The unit cell is periodic in plane with a period of 13.0 nm, while the periodicity out of plane is 6.0 nm. The amino acids are given appropriate masses, and the entire system is initially kept at 300 K and then cooled to 30 K in order to solidify the lattice.

This final structural model is used to reconstruct the entire small-angle X-ray scattering signal. The whole reconstruction procedure is based on a method presented in our earlier work, where we describe the S-layer system by a fractal mean potential.¹⁵

In Figure 3B, we give the background-corrected scattering contrasts for the monomeric and the self-assembled samples, where the scattering intensity is shown as a function of the scattering vector Q . Black open circles give the scattering contrast of a predominantly monomeric solution, while blue open circles represent the scattering contrast of self-assembled structures. Obviously the monomeric solution lacks any corrugation—all characteristic Bragg peaks are missing. We take this as a clear indication that, although secondary structure elements are formed, the tertiary structure of the protein remains non-native at this stage, suggesting that in their monomeric state the proteins adapt a different conformation than as part of a tetramer. This hypothesis that the proteins first condense into an amorphous cluster in an extended conformation before they restructure to form a crystal of folded tetramers has been recently proposed by Chung *et al.*¹² However, the slope of the scattering intensity is linear for small Q values and thus indicates a self-similar system. The fractal dimension of the

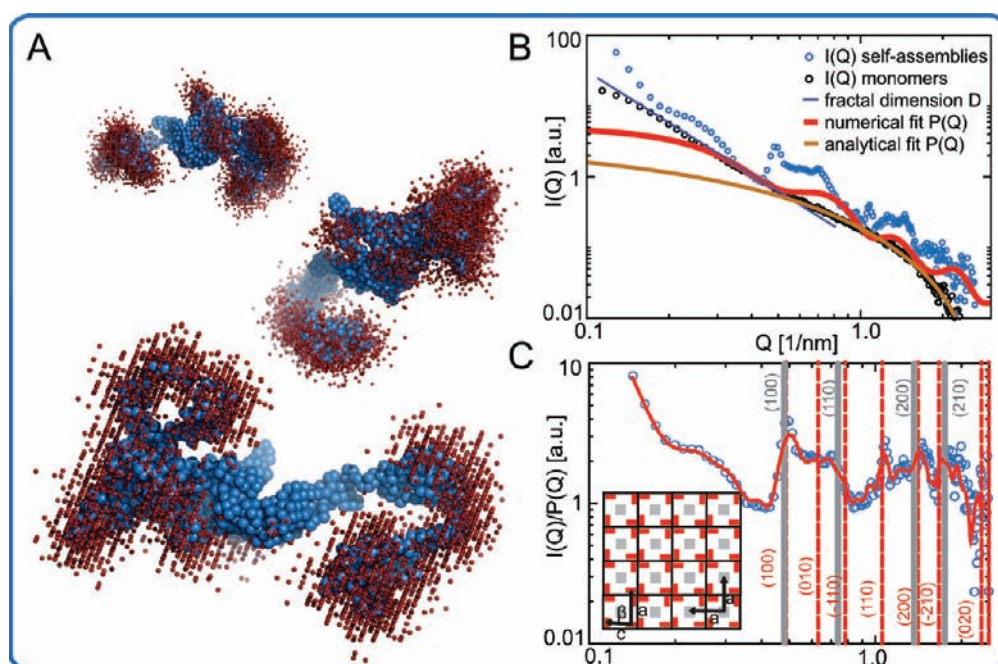


Figure 3. SAXS. (A) Distribution of scattering clusters (red beads) of one SbpA monomer (blue). The scattering sites represent high electron density contrast as determined by small-angle X-ray scattering (SAXS) and a Monte Carlo algorithm. (B) Intensities $I(Q)$ as a function of the scattering vector Q of the monomeric solution and the sample containing self-assemblies are given by black and blue open circles. The slope for the fractal dimension $D = 2.4$ is inserted as a dark blue line. The red line gives the numerical reconstruction using a Monte Carlo algorithm generalized to fractal dimensions and based on the coarse-grained monomer model. The yellow line gives the analytical fit of the monomer data. The reconstruction procedure is based on our previous work.¹⁵ (C) Open circles give scattering intensities divided by the form factor $I(Q)/P(Q)$. The red line is the fit curve from eq 1, which includes the contribution to the scattering signal of the S-layer self-assemblies in solution. On the basis of this fit, the height $H = 5.6$ nm, the radius $R = 7.2$ nm, and the radius of the self-assemblies $R = 72$ nm can be calculated. Vertical lines indicate Bragg reflections, where strong peaks are indicated by gray lines and weak peaks by red lines. Strong peaks indicate an arrangement of the tetramers in a cubic lattice with $a = 13.0 \pm 0.1$ nm, where weak peaks arise due to the arrangement of the proteins in the tetramer. These weak reflections indicate a 2D monoclinic sublattice with unit cell dimensions of $a = 13.0 \pm 0.1$ nm, $c = 10.5 \pm 0.1$ nm, and $\gamma = 80 \pm 1$ (a detailed view of the Bragg reflections is given in Figure S1 in the Supporting Information). A corresponding schematic drawing illustrates the arrangement of the lattice and sublattice.

monomers and self-assemblies is $D = 2.4$ (Figure 3B dark blue), which would be consistent with the formation of layers with a dimension between a two-dimensional plate and a three-dimensional solid. The entire signal with the exception of peaks originating from the structure factor is reconstructed from the baseline, that is, the minima of the scattering curve, by the analytical function $I(Q) \propto K_{D/2-1}(Q\kappa)/(Q\kappa)^{(D/2-1)}$, where K is a Bessel- K function, Q the scattering vector, D the fractal dimension, and κ a characteristic magnitude of the unfolded monomer (Figure 3B yellow line). The details of the reconstruction procedure of a SAXS signal produced by S-layers in solution have been given previously.¹⁵ The scattering contrast of the self-assembled S-layers is fitted numerically on the basis of the coarse-grained model (Figure 1I) and is shown as a red line in Figure 3B. The details of the numerical fitting procedure were given previously.¹⁴ On the basis of the coarse-grained model, it is possible to numerically fit the minima of the Bragg peaks (the red line in Figure 3B follows the pattern of the Bragg peaks in a defined way), which indicates that the proteins only exhibit their tertiary structure when assembled into the lattice structure. However, the fact that we can perfectly fit the

scattering contrast of the self-assembled layers using the coarse-grained monomeric model substantiates the presented calculated tertiary structure of the protein, at least on a coarse-grained level.

In a second step, we refine the simulated coarse-grained monomer model (Figure 1I) on the basis of the SAXS data. A reverse Monte Carlo algorithm was applied¹⁵ to reconstruct scattering sites, visible as red beads in Figure 3A. The set of identified clusters corresponds to secondary structure elements at these locations. The scattering intensity arises in particular from those regions with a high electron density contrast: it is visible in Figure 3A that this electron density contrast is concentrated at the terminal domains of the monomer, whereas those parts of the structure, where overlapping takes place to result in the final tetramer, are exhibiting less contrast. This decrease of electron density contrast as a consequence of the formation of bonds supports the presented structural model of the S-layer tetramer.

For the reconstruction of the whole scattering signal of S-layer self-assemblies, we divide the signal by the form factor and thus obtain contributions of the self-assembled monomolecular sheets. The reconstruction

as shown in Figure 3C is calculated using eq 1 (we give the detailed deduction of eq 1 in the Methods section). If the self-assembled monomolecular layers in solution adapt a particular orientation with respect to the scattering vector, the contribution to the scattering contrast is supposed to be dominant over all other orientations. This is in line with the results we observed in our previous work,¹⁵ where another S-layer protein was analyzed in solution by means of SAXS. In that case, we described the self-assembled layers as nanodisks, which were oriented parallel to each other and thereby formed some kind of nematic liquid. Here we refine this theory, as we anticipate that parallel oriented nanodisks contribute most to the SAXS signal.

$$\mathcal{F}(\rho(\zeta))[Q]/\hat{\rho}^{(1)}(Q) = - \int_0^1 d(\cos \theta) f(\cos \theta) \hat{\delta}_{\parallel}(QR \cos \theta) \hat{\delta}_{\perp}(H^2 Q \cos \theta) \hat{\delta}_{\parallel}(QR_{\gg} \cos \theta) \quad (1)$$

Therein H represents the height of the layer, R is the in-plane radius of the protein, and R_{\gg} a multiple of R (i.e., the layer size).

On the basis of eq 1, we can reconstruct the entire structure factor and deduce a possible characteristic height $H = 5.6$ nm, radius $R = 7.2$ nm, and nanodisk size $R_{\gg} = 72$ nm. This size, obtained from the model reconstruction, which takes into account the whole

scattering curve, is completely consistent with the size obtained from the identified Bragg reflections (Figure 3C vertical gray lines). They indicate an arrangement of the tetramers in a cubic lattice with $a = 13.0 \pm 0.1$ nm. An additional substructure is visible by a number of weaker reflections (Figure 3C vertical red lines). They are attributed to the slightly oblique arrangement of the monomers along the main axes of the cubic lattice and are thus described by a 2D monoclinic sublattice with unit cell dimensions of $a = 13.0 \pm 0.1$ nm, $c = 10.5 \pm 0.1$ nm, and $\gamma = 80 \pm 1^\circ$ (a detailed view of the Bragg reflections is given in Figure S1 in the Supporting Information).

The resulting structural model of one SbpA unit cell is shown in Figure 4. In the tetramer, the monomers are interlocked into one another, where both termini are accessible at either surface. The N-termini are located at the inner surface of the tetramer, which is anchored on the cell surface *via* SLH domains (Figure 4B), whereas the C-termini are accessible at the outer surface of the S-layer unit cell (Figure 4C). Figure 4D,E shows the scattering clusters reconstructed from the SAXS curves as red beads of one monomer in the unit cell. Apparently, the electron density contrast is higher in those domains that are not part of direct overlaps in the lattice. These tetramers build up the S-layer lattice (Figure 4A), leading to the formation of defined pores between the single unit cells. These pores exhibit a

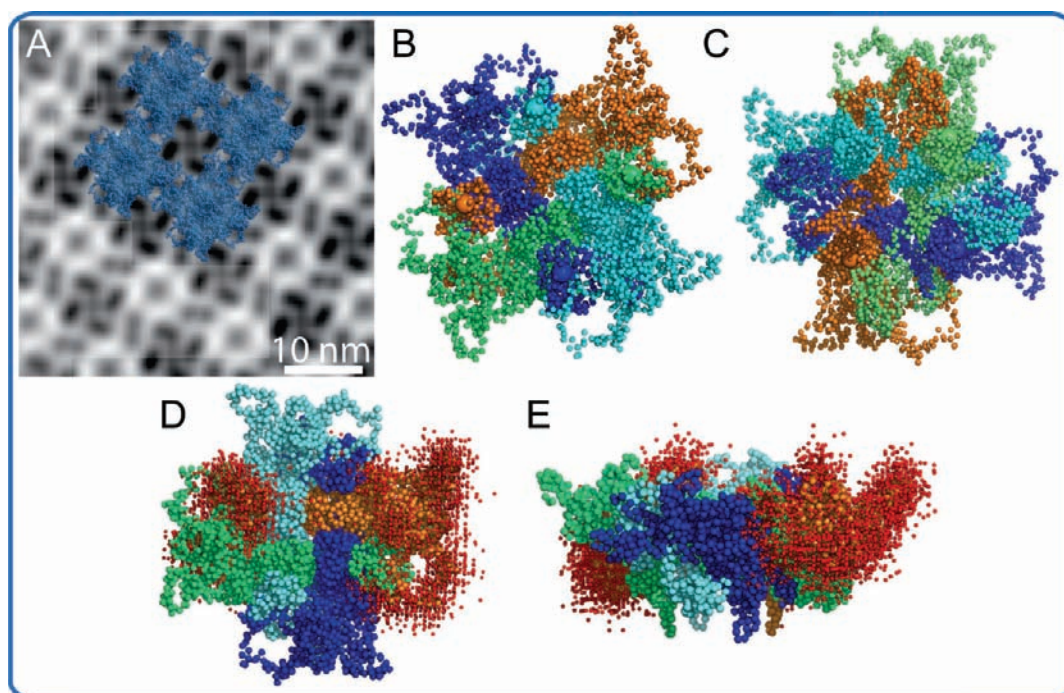


Figure 4. Structure of the resulting SbpA unit cell. (A) Comparison of the calculated structure with an electron micrograph showing an SbpA lattice. (B–E) Every monomer in the tetramer is illustrated in a different color. The proteins are interlocked into each other. (B) Inner surface of the tetramer, which anchors the protein on the cell surface. The N-termini are represented by magnified beads and are accessible on the surface. (C) Outer surface of the tetramer, which is exposed to the surroundings of the cell. The C-termini are also accessible and marked as magnified beads. (D,E) Red beads represent scattering clusters of one monomer of the tetramer as shown in Figure 3A. The overlapping or interacting parts do not show high electron density contrast.

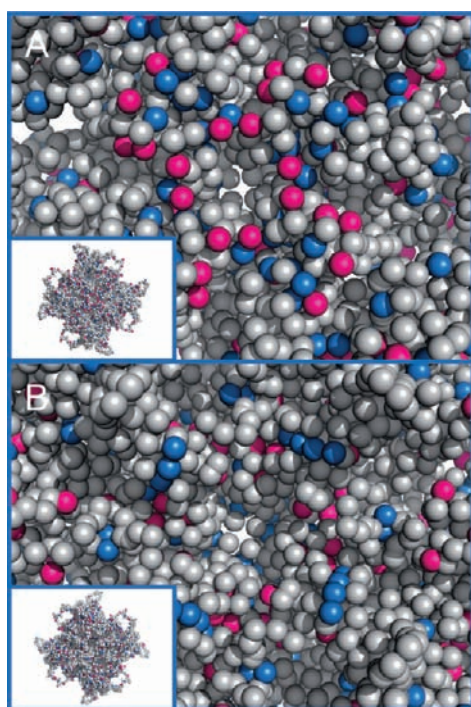


Figure 5. Enlarged view of the central region of one tetramer: (A) inner surface, which is mainly positively charged in the central region; (B) outer surface exhibiting mainly negatively charged amino acids. Pink beads, positively charged amino acids; blue beads, negatively charged amino acids.

very regular cubic arrangement (with a pore-to-pore distance of 13.0 nm) and lead to the strong Bragg peaks in the scattering curve in the high Q regime (see Figure 3C and Figure S1 in the Supporting Information). In the center of the unit cells, there is a clear anisotropic charge distribution at the outer and inner surface, where at the inner surface the center is surrounded by positively charged residues and at the outer surface by negatively charged residues, as shown in detail in Figure 5A,B. The accessibility of the respective termini and the different charge distribution of the outer and the inner central region of the unit cell have already been demonstrated experimentally.^{16,21,25,30} Various S-layer fusion proteins have been experimentally investigated, where truncated forms of the protein were used.^{22,26,31,32} Figure 6 shows the correspondent truncated forms using the structural model of the unit cell. The deletion of 237 C-terminal amino acids leads to the change of the lattice symmetry from p4 to p1.²⁷ Figure 6A,B shows the unit cell, where the deleted amino acids are colored blue. If the C-terminus is shortened by another consecutive 113 amino acids, the proteins are not capable of self-assembling anymore. These residues are shown in Figure 6C,D. Figure 6E,F shows the corresponding monomers, where it becomes clear that the main part of the C-terminal domain, which can be deleted without any loss of functionality, is located outside of the lattice plane and is not part of any distinct overlaps. However, the deletion of this

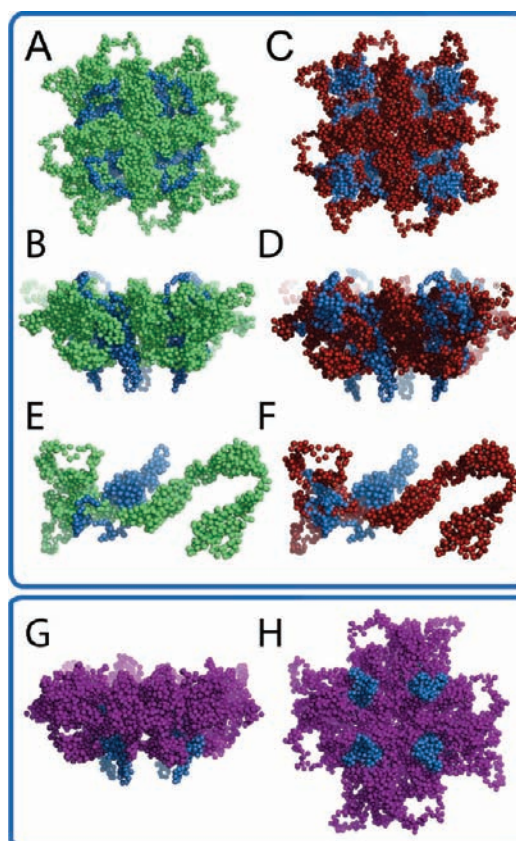


Figure 6. Analysis of the resulting structural model of the SbpA unit cell. The tetramer is compared to experimentally investigated truncated recombinant forms of the protein. (A,B) Blue beads represent the last 237 C-terminal amino acids. The deletion of the blue part leads to a change of the lattice symmetry from p4 to p1.²⁷ (E) Corresponding monomer. (C,D) Further deletion of 113 amino acids (blue part) leads to a loss of functionality. The proteins are not able to self-assemble anymore. (F) Corresponding monomer. The tetrameric structure shows clearly that the C-terminal part that is not located in the plane of the lattice can be deleted without any loss of function. If additional parts in the plane of the unit cell are deleted, the monomers are not able to form the tetramer. (G,H) Unit cell made up of a recombinant form of the monomer that lacks the last 170 amino acids (rSbpA 1068). This form of the protein has been extensively studied for applications as fusion partner for various molecules. It was reported that this recombinant form is better accessible than other truncated forms.^{17,20–22,31} The cartoons clearly show that the C-terminal part, which is located out of the plane of the unit cell, is deleted, which smoothes the surface and makes the tetramer better accessible.

separate domain leads to a clear change in the monomer structure, which might very well explain the change in lattice symmetry, where a p4 symmetry demands a more complicated and flexible monomeric structure than a p1 symmetry. Apparently, the further truncation of parts of the protein, which are located inside the plane, leads to a loss of functionality. Interestingly, the electron density contrast, as determined by SAXS (Figure 4D,E), gives very low scattering sites at exactly these locations, which leads to the hypothesis that this part is involved in essential interactions in the tetramer. The most investigated and successfully used

S-layer fusion protein contains a C-terminally truncated form, which maintained the residues 1–1068, where it was reported that this recombinant protein provides a highly accessible fusion partner.^{17,20–22,31} Figure 6G,H shows this recombinant protein as part of a unit cell, where the blue part represents the truncated domain. The better accessibility of the fused molecule might be explained by the smoothening of the outer surface through deletion of the C-terminal part, which is located outside of the lattice plane and therefore sticking out of the unit cell.

CONCLUSIONS

In summary, we have demonstrated the successful combination of *in silico* and experimental methods to calculate the three-dimensional structure of an S-layer lattice, which will pave the way for the directed bottom-up design of nanoscale structures based on S-layer proteins. On the basis of an atomistic model calculated using molecular dynamics simulations and molecular modeling, the unit cell was modeled by combining the monomer structure and three-dimensional density distribution data as obtained from tilting studies using electron microscopy. The resulting unit cell was equilibrated again by molecular dynamics simulations and subsequently used to fit small-angle X-ray scattering intensity data of self-assemblies and monomers in solution. Interestingly, the calculated scattering clusters, which represent high electron density contrasts, are located at domains of the monomer that are not part of significant interactions in the tetrameric model, which shows that electron density contrasts are diminished—at least at a resolution achieved by SAXS—if interactions take place. This consistency between the SAXS results and the simu-

lated unit cell structure regarding overlapping parts strongly substantiates the structural model of the unit cell. Furthermore, the SAXS results of the monomeric sample indicate that the proteins are not fully structured in their monomeric state. The hypothesis of the importance of conformational transformations guiding the S-layer self-assembly has been also recently proposed by Chung *et al.*¹² and legitimizes our approach to base the calculation of the tetrameric unit cell on an intermediate monomeric structure. The resulting structural model shows an anisotropy regarding the charge distribution in the center of the tetramer, which has been seen experimentally before. We explained the structure of a tetramer based on truncated recombinant forms of the S-layer protein that have been investigated experimentally and used for nanobiotechnological applications to better understand the behavior of these forms when assembled into S-layer lattices. The presented structure of an S-layer unit cell can serve as a basis for the specific and directed binding of various molecules. So far, the usage of S-layers as building blocks was somehow based on a trial and error approach due to the lack of structural details. The exact location of N- and C-terminal domains in the unit cell, the charge distribution on the inner and outer central region of the S-layer lattice, as well as the altered surface architecture due to the truncation of monomers will help to better understand the behavior of nanostructures based on S-layers and to specifically alter certain parts of the S-layer lattice for the production of nanostructures with different behavior and various architectures. Together with already available experimental data of S-layer proteins, this approach opens the path to determining locations, type, and distribution of amino acids in the S-layer lattice.

METHODS

Protein Preparation. The S-layer protein SbpA from *Bacillus sphaericus* CCM2177 was isolated by default as previously described.³³ After the cell wall preparation, the proteins were extracted with 5 M guanidine hydrochloride (pH 7.2) at room temperature for 30 min. Afterward, the cell wall fragments were removed by centrifugation (30 000 rpm, 60 min, 8°C). The supernatant containing the proteins was dialyzed against 10 mM CaCl₂ for 2 h at pH 5.9. The dialyzed sample, which contains self-assemblies of SbpA, was used for the SAXS measurements. For the TEM measurements, one drop of the protein sample was adsorbed on a grid and negatively stained as previously described.³⁴ A monomer solution was obtained by centrifuging for 30 min at 8 °C and 30 000 rpm.

Molecular Modeling. The S-layer protein SbpA was modeled using different bioinformatics tools and molecular dynamics simulations. Sequence homology searches were performed using BLAST³⁵ at www.ncbi.nlm.nih.gov/BLAST, where homologies could be found to other S-layer proteins, especially regarding the S-layer homology domains (SLH domains at the N-terminus). Additionally, we found homologies to RTX toxins

and related Ca²⁺-binding proteins. The S-layer protein SbpA needs bivalent cations to self-assemble in solution into the typical two-dimensional sheets, which might explain structural homologies to Ca²⁺-binding proteins. We also found homologies to Ig-like proteins, which characteristically contain β -sheet structures and to fibronectin type III domains. A comparison of the SbpA sequence with these structural homologues resulted in 26% identity to the S-layer protein SbsB, which has been modeled in our previous work,¹⁴ 34% identity to an Ig-like protein, 25% to RTX toxins and the related Ca²⁺-binding protein, 32% to a fibronectin type III domain, and 42% identity to the SLH domain of the S-layer protein of *Bacillus cereus*. Secondary structure predictions were made using PSIPRED³⁶ available at www.psipred.net/psiform.html, DOMPRED³⁷ at <http://bioinfadmin.cs.ucl.ac.uk/dompred>, and GOR4³⁸ at <http://pbil.ibcp.fr/htm/index.php>. All three algorithms predict mainly α -helices for the N-terminal region and β -sheets for the rest of the protein, which is a common structural motive for S-layer proteins. Finally, we performed domain predictions using CDART³⁹ available at www.ncbi.nlm.nih.gov/Structure/lexington/lexington.cgi and Pfam⁴⁰ at <http://pfam.sanger.ac.uk>.

Both algorithms predicted three SLH domains at the N-terminus of the protein. Pfam additionally predicted two Ig-like domains and CDART two fibronectin type III domains. On the basis of all these predictions and the information of the self-assembly behavior of truncated forms of the protein, we defined seven structurally meaningful domains: three SLH domains, one Ca^{2+} -binding domain, three Ig-like or fibronectin type III domains. The molecular dynamic simulations were performed using LAMMPS,⁴¹ which is distributed by the Sandia National Laboratories and free to download at <http://lammps.sandia.gov>. All simulations were computed at the SUN cluster Phoenix at phoenix.zserv.tuwien.ac.at. Visualizations were done using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrodinger, LLC.

Transmission Electron Microscopy. Tilt series were performed using a Philips transmission electron microscope CM12 operated at 80 kV. The tilt range was $\pm 60^\circ$. For the three-dimensional image reconstruction, we used a software package from Caldiris: CRISP 2.1.a (to edit tilted electron micrographs and perform FFT (fast Fourier transformation)), TriMerge 1.6.a (to reconstruct the three-dimensional density based on Fourier transforms of electron micrographs and to perform the inverse Fourier transform²⁹), and TriView 1.3 (to visualize three-dimensional density data). All in all, we performed six tilt series in 5 and 10° steps at a magnification of 37 000.

Small-Angle X-ray Scattering. Small-angle X-ray scattering (SAXS) was performed with Cu K α radiation from a rotating anode generator (Nanostar, BRUKER AXS, Karlsruhe, Germany) equipped with a pinhole camera and an area detector (VANTEC 2000 from BRUKER AXS). The sample containing self-assembled SbpA proteins was put into capillaries with 1 mm diameter and 10 μm wall thickness (from Hilgenberg) and then sealed with polymeric caps, as the whole equipment operates in vacuum. The SAXS intensity patterns were taken at a sample to detector distance of 109 cm for 6 h. They were corrected for background scattering and then radially averaged to obtain the function $I(q)$, where $q = (4\pi/\lambda)\sin \theta$ is the scattering vector, 2θ the angle between incident and diffracted beam, and $\lambda = 0.1542$ nm the X-ray wavelength. The scattering intensities were normalized to each other in the q range between 2 and 2.5 nm^{-1} . In this region, the scattering intensity is flat and dominated mainly by fluid scattering and an additional smaller contribution from the glass capillary. The scattering intensity from the solution without protein was then subtracted from the solution with protein. We cross-checked that this is consistent with the measured transmission of each sample.

For the analysis of the scattering data, $I(Q)$, we adapt our theory that we have recently formulated to access protein self-assemblies.¹⁵ We follow the Green's approach and give the local density distribution of all proteins in solution by

$$\mathcal{F}(\rho(\zeta))[Q] = I(Q) = \mathcal{F}\left(\int_R^\infty \left(\int_R^\infty G(u' - u'')\rho^{(1)}(u'')du''\right) (1 - \beta w(\zeta - u'))du'\right) [Q] \quad (2)$$

For the Green's function, we assume $G(u' - u'') = \delta(u' - u'')$, where this gives a possible solution of the wave equation. Therein the local density distribution of a single protein is given by $\rho^{(1)}(\zeta)$, where protein–protein interactions are approximated by $(1 - \beta w(\zeta - u'))$. This approximation is exact for hard-core potentials. In reciprocal space, the complex convolution given in eq 2 simplifies to $I(Q) = \bar{G}(Q)w(Q)^{(1)}(Q)$. Consequently, we calculate the structure factor of the entire system based on the form factor of the protein $\rho^{(1)}(Q)$, which represents the intramean potential of one monomer.¹⁵

We anticipate that the proteins self-assemble in a nanodisk-like manner.¹⁵ Furthermore, we assume that the nanodisks interact only by hard-core potentials. Then $w(\zeta) = \delta(\zeta^2 - H^2)\Theta(\zeta - R) = \int \pi(R_i)\delta(\zeta^2 - H^2)\delta(\zeta - R_i)dR_i$, with H being the disk height and R the disk radius. In contrast to our previous work,¹⁵ we introduce a radial distribution function $\pi(R_i)$. If there is no disk radius favored over the other, this probability function may be set to 1. However, we now assume that large disks, that is, disks of a radius R_∞ contribute most. Thus we set $\pi(R_i) = \delta(\zeta - R_\infty)$ and

give the protein structure factor by

$$\mathcal{F}(\rho(\zeta))[Q]/\bar{\rho}^{(1)}(Q) = - \int_0^1 d(\cos \theta) f(\cos \theta) \hat{\delta}_\parallel(QR \cos \theta) \hat{\delta}_\perp(HQ \sin \theta) \hat{\delta}_\parallel(QR_\infty \cos \theta) \quad (3)$$

Therein $\hat{\delta}_{\perp\parallel}(\dots) = J_{D\perp\parallel/2-1}(\dots) (\dots)^{D\perp\parallel/2-1}$ indicates the fractal Fourier transform of the respective $\delta_{\perp\parallel}$ functions.

We emphasize that $f(\cos \theta)$ may be argued as a summation of all possible orientations of the nanodisk with respect to the scattering vector. We now introduce a partition function in terms of the system's mean potential $w(\zeta, \zeta')$, with the system coordinates ζ and an arbitrary linear reaction coordinate ζ' .^{42–44} The partition function may be given by $\Pi(\zeta') = \langle \exp(-\beta w(\zeta, \zeta')) \rangle_\zeta$. An equivalent formulation is

$$\Delta \Pi = -\beta^{-1} \left\langle \int_{-\infty}^0 d\zeta' \partial_{\zeta'} \left(w(\zeta, \zeta') \exp(-\beta w(\zeta, \zeta')) \right) \right\rangle_\zeta \quad (4)$$

with $\beta = 1/k_B T$. In the literature, this approach is described as λ -integration^{42,45} or mean force method.⁴⁶ To change the integration boundaries, we introduce a coordinate transformation $\zeta' = \exp(-\lambda)$ and rewrite eq 4 by

$$\Delta \Pi = -\beta^{-1} \left\langle \int_0^1 d\lambda \partial_\lambda \exp(-\beta(w(\zeta, \lambda) - \beta^{-1} \ln|\lambda|)) \right\rangle_\zeta \quad (5)$$

The Jacobi determinant is essential for the calculation of the mean force,⁴³ and here it is $\exp(\lambda)$. If we compare eq 2 and eq 5, it is straightforward that $f(\cos \theta)$, which equals $-\beta \partial_{\cos \theta} w(\cos \theta)$, is rather a mean force than a distribution of the orientation of the nanodisks, given by $\cos \theta$. If we aim for the particular distribution function, we have to calculate the Boltzmann weighted mean potential: $\Delta \Pi(\cos \theta) = \exp(-\beta w(\cos \theta))$.⁴⁷

We now simplify eq 5 taking into account $\tan \theta = H/Q$:

$$\mathcal{F}(\rho(\zeta))[Q]/\bar{\rho}^{(1)}(Q) = \int_0^1 d(\cos \theta) (\partial_{\cos \theta} w(\cos \theta) - 1) \hat{\delta}(QR \cos \theta) \hat{\delta}_\perp(H^2 \cos \theta) \hat{\delta}_\parallel(QR_\infty \cos \theta) \quad (6)$$

Finally do note that the term $R_\infty \cos \theta$ may as well be interpreted as particular distances ζ that the proteins comprise when forming self-assemblies.

Acknowledgment. The authors thank Jacqueline Friedmann for the preparation of the SAXS samples. C.H. holds a DOC-FORTE fellowship of the Austrian Academy of Sciences. This project has been supported by the Air Force Project FA9550-09-1-0342.

Supporting Information Available: Detailed view of Figure 3 and the mean potential as determined by SAXS. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- Sleytr, U. B. Regular Arrays of Macromolecules on Bacterial Cell Walls: Structure, Chemistry, Assembly, and Function. *Int. Rev. Cytol.* **1978**, *53*, 1–64.
- Sleytr, U. B. Heterologous Reattachment of Regular Arrays of Glycoproteins on Bacterial Surfaces. *Nature* **1975**, *257*, 400–402.
- Sleytr, U. B.; Messner, P.; Pum, D.; Sara, M. Crystalline Bacterial Cell Surface Layers (S-Layers): From Supramolecular Cell Structure to Biomimetics and Nanotechnology. *Angew. Chem., Int. Ed.* **1999**, *38*, 1034–1054.
- Sara, M.; Sleytr, U. B. S-Layer Proteins. *J. Bacteriol.* **2000**, *182*, 859–868.
- Egelseer, E. M.; Ilk, N.; Pum, D.; Messner, P.; Schäffer, C.; Schuster, B.; Sleytr, U. B. Nanobiotechnological Applications of S-Layers. In *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*; Flicking, M. C., Ed.; Wiley and Sons: Weinheim, Germany, 2009; pp 4424–4448.
- Pavkov, T.; Oberer, M.; Egelseer, E. M.; Sara, M.; Sleytr, U. B.; Keller, W. Crystallization and Preliminary Structure Determination of the C-Terminal Truncated Domain of the

- S-Layer Protein SbsC. *Acta Crystallogr. Sect. D: Biol. Crystallogr. D* **2003**, 59, 1466–1468.
7. Pavkov, T.; Egelseer, E. M.; Tesarz, M.; Svergun, D. I.; Sleytr, U. B.; Keller, W. The Structure and Binding Behavior of the Bacterial Cell Surface Layer Protein SbsC. *Structure* **2008**, 16, 1226–1237.
 8. Howorka, S.; Sara, M.; Wang, Y.; Kuen, B.; Sleytr, U. B.; Lubitz, W.; Bayley, H. Surface-Accessible Residues in the Monomeric and Assembled Forms of a Bacterial Surface Layer Protein. *J. Biol. Chem.* **2000**, 275, 37876–37886.
 9. Kinns, H.; Badelt-Lichtblau, H.; Egelseer, E. M.; Sleytr, U. B.; Howorka, S. Identifying Assembly-Inhibiting and Assembly-Tolerant Sites in the SbsB S-Layer Protein from *Geobacillus stearothermophilus*. *J. Mol. Biol.* **2010**, 395, 742–753.
 10. Fagan, R. P.; Albesa-Jove, D.; Qazi, O.; Svergun, D. I.; Brown, K. A.; Fairweather, N. F. Structural Insights into the Molecular Organization of the S-Layer from *Clostridium difficile*. *Mol. Microbiol.* **2009**, 71, 1308–1322.
 11. Norville, J. E.; Kelly, D. F.; Knight, T. F., Jr.; Belcher, A. M.; Walz, T. 7 Å Projection Map of the S-Layer Protein SbpA Obtained with Trehalose-Embedded Monolayer Crystals. *J. Struct. Biol.* **2007**, 160, 313–323.
 12. Chung, S.; Shin, S.-H.; Bertozzi, C.; De Yoreo, J. Self-Catalyzed Growth of S Layers via an Amorphous-to-Crystalline Transition Limited by Folding Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 107, 16536–16541.
 13. Whitelam, S. Control of Pathways and Yields of Protein Crystallization through the Interplay of Nonspecific and Specific Attractions. *Phys. Rev. Lett.* **2010**, 105, 088102–1–088102-4.
 14. Horejs, C.; Pum, D.; Sleytr, U. B.; Tscheliessnig, R. Structure Prediction of an S-Layer Protein by the Mean Force Method. *J. Chem. Phys.* **2008**, 128, 065106–1–065106-10.
 15. Horejs, C.; Pum, D.; Sleytr, U. B.; Peterlik, H.; Jungbauer, A.; Tscheliessnig, R. Surface Layer Protein Characterization by Small Angle X-ray Scattering and a Fractal Mean Force Concept: From Protein Structure to Nanodisk Assemblies. *J. Chem. Phys.* **2010**, 133, 175102–1–175102-8.
 16. Ilk, N.; Völlenkne, C.; Egelseer, E. M.; Breitwieser, A.; Sleytr, U. B.; Sara, M. Molecular Characterization of the S-Layer Gene, *sbpA*, of *Bacillus sphaericus* CCM 2177 and Production of a Functional S-Layer Fusion Protein with the Ability To Recrystallize in a Defined Orientation while Presenting the Fused Allergen. *Appl. Environ. Microbiol.* **2002**, 68, 3251–3260.
 17. Badelt-Lichtblau, H.; Kainz, B.; Völlenkne, C.; Egelseer, E. M.; Sleytr, U. B.; Pum, D.; Ilk, N. Genetic Engineering of the S-Layer Protein SbpA of *Lysinibacillus sphaericus* CCM 2177 for the Generation of Functionalized Nanoarrays. *Bioconjugate Chem.* **2009**, 20, 895–903.
 18. Delcea, M.; Krastev, R.; Gutberlet, T.; Pum, D.; Sleytr, U. B.; Toca-Herrera, J. L. Thermal Stability, Mechanical Properties and Water Content of Bacterial Protein Layers Recrystallized on Polyelectrolyte Multilayers. *Soft Matter* **2008**, 4, 1414–1421.
 19. Moreno-Flores, S.; Kasry, A.; Butt, H.-J.; Vavilala, C.; Schmitt, M.; Pum, D.; Sleytr, U. B.; Toca-Herrera, J. L. From Native to Non-Native Two-Dimensional Protein Lattices through Underlying Hydrophilic/Hydrophobic Nanoprotrusions. *Angew. Chem., Int. Ed.* **2008**, 120, 4785–4788.
 20. Ilk, N.; Küpcü, S.; Moncayo, G.; Klimt, S.; Ecker, R. C.; Hofer-Warbinek, R.; Egelseer, E. M.; Sleytr, U. B.; Sara, M. A Functional Chimaeric S-Layer-Enhanced Green Fluorescent Protein To Follow the Uptake of S-Layer-Coated Liposomes into Eukaryotic Cells. *Biochem. J.* **2004**, 379, 441–448.
 21. Huber, C.; Egelseer, E. M.; Ilk, N.; Sleytr, U. B.; Sara, M. S-Layer-Streptavidin Fusion Proteins and S-Layer-Specific Heteropolysaccharides as Part of a Biomolecular Construction Kit for Application in Nanobiotechnology. *Microelectron. Eng.* **2006**, 83, 1589–1593.
 22. Völlenkne, C.; Weigert, S.; Ilk, N.; Egelseer, E. M.; Weber, V.; Loth, F.; Falkenhagen, D.; Sleytr, U. B.; Sara, M. Construction of a Functional S-Layer Fusion Protein Comprising an Immunoglobulin G-Binding Domain for Development of Specific Adsorbents for Extracorporeal Blood Purification. *Appl. Environ. Microbiol.* **2004**, 70, 1514–1521.
 23. Schuster, B.; Gufler, P. C.; Pum, D.; Sleytr, U. B. Interplay of Phospholipase A₂ with S-Layer-Supported Lipid Monolayers. *Langmuir* **2003**, 19, 3393–3397.
 24. Tang, J.; Ebner, A.; Badelt-Lichtblau, H.; Völlenkne, C.; Rankl, C.; Kraxberger, B.; Leitner, M.; Wildling, L.; Gruber, H. J.; Sleytr, U. B.; *et al.* Recognition Imaging and Highly Ordered Molecular Templating of Bacterial S-Layer Nanoarrays Containing Affinity-Tags. *Nano Lett.* **2008**, 8, 4312–4319.
 25. Györfv, E.; Schroedter, A.; Talapin, D. V.; Weller, H.; Pum, D.; Sleytr, U. B. Formation of Nanoparticle Arrays on S-Layer Protein Lattices. *J. Nanosci. Nanotechnol.* **2004**, 4, 115–120.
 26. Ilk, N.; Kosma, P.; Puchberger, M.; Egelseer, E. M.; Mayer, H. F.; Sleytr, U. B.; Sara, M. Structural and Functional Analyses of the Secondary Cell Wall Polymer of *Bacillus sphaericus* CCM 2177 That Serves as an S-Layer-Specific Anchor. *J. Bacteriol.* **1999**, 181, 7643–7646.
 27. Huber, C.; Ilk, N.; Rünzler, D.; Egelseer, E. M.; Weigert, S.; Sleytr, U. B.; Sara, M. The Three S-Layer-like Homology Motifs of the S-Layer Protein SbpA of *Bacillus sphaericus* CCM2177 Are Not Sufficient for Binding to the Pyruvylated Secondary Cell Wall Polymer. *Mol. Microbiol.* **2005**, 55, 197–205.
 28. Kelley, L. A.; Sternberg, M. J. E. Protein Structure Prediction on the Web: A Case Study Using the Phyre Server. *Nat. Protoc.* **2009**, 4, 363–371.
 29. Amos, L. A.; Henderson, R.; Unwin, P. N. Three-Dimensional Structure Determination by Electron Microscopy of Two-Dimensional Crystals. *Prog. Biophys. Mol. Biol.* **1982**, 39, 183–231.
 30. Tang, J.; Badelt-Lichtblau, H.; Ebner, A.; Preiner, J.; Kraxberger, B.; Gruber, H. J.; Sleytr, U. B.; Ilk, N.; Hinterdorfer, P. Fabrication of Highly Ordered Gold Nanoparticle Arrays Templated by Crystalline Lattices of Bacterial S-Layer Protein. *ChemPhysChem* **2008**, 9, 2317–2320.
 31. Pleschberger, M.; Neubauer, A.; Egelseer, E. M.; Weigert, S.; Lindner, B.; Sleytr, U. B.; Muyldermans, S.; Sara, M. Generation of a Functional Monomolecular Protein Lattice Consisting of an S-Layer Fusion Protein Comprising the Variable Domain of a Camel Heavy Chain Antibody. *Bioconjugate Chem.* **2003**, 14, 440–448.
 32. Huber, C.; Liu, J.; Egelseer, E. M.; Moll, D.; Knoll, W.; Sleytr, U. B.; Sara, M. Heterotetramers Formed by an S-Layer-Streptavidin Fusion Protein and Core-Streptavidin as a Nanoarrayed Template for Biochip Development. *Small* **2006**, 2, 142–150.
 33. Sleytr, U. B.; Sara, M.; Küpcü, Z.; Messner, P. Structural and Chemical Characterization of S-Layers of Selected Strains of *Bacillus stearothermophilus* and *Desulfotomaculum nigrificans*. *Arch. Microbiol.* **1986**, 146, 19–24.
 34. Messner, P.; Hollaus, F.; Sleytr, U. B. Paracrystalline Cell Wall Surface Layers of Different *Bacillus stearothermophilus* Strains. *Int. J. Syst. Bacteriol.* **1984**, 34, 202–210.
 35. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, 25, 3389–3402.
 36. McGuffin, L. J.; Bryson, K.; Jones, D. T. The PSIPRED Protein Structure Prediction Server. *Bioinformatics* **2000**, 16, 404–405.
 37. Marsden, R. L.; McGuffin, L. J.; Jones, D. T. Rapid Protein Domain Assignment from Amino Acid Sequence Using Predicted Secondary Structure. *Protein Sci.* **2002**, 11, 2814–2824.
 38. Garnier, J.; Gibrat, J.-F.; Robson, B. GOR Method for Predicting Protein Secondary Structure from Amino Acid Sequence. *Methods Enzymol.* **1996**, 266, 540–553.
 39. Geer, L. Y.; Domrachev, M.; Lipman, D. J.; Bryant, S. H. CDART: Protein Homology by Domain Architecture. *Genome Res.* **2002**, 12, 1619–1623.
 40. Finn, R. D.; Mistry, J.; Tate, J.; Coghill, P.; Heger, A.; Pollington, J. E.; Gavin, O. L.; Gunasekaran, P.; Ceric, G.;

- Forslund, K.; *et al.* The Pfam Protein Families Database. *Nucleic Acids Res.* **2010**, *38*, 211–222.
41. Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
42. Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987.
43. Otter, W. K.; Briels, W. J. The Calculation of Free-Energy Differences by Constrained Molecular-Dynamics Simulations. *J. Chem. Phys.* **1998**, *109*, 4139–4146.
44. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
45. McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Mill Valley, CA, 1976.
46. Tscheliesnig, R.; Geyrhofer, L.; Wendland, M.; Fischer, J. Adsorption from Oversaturated Aqueous Solution: Mean Force Molecular Simulations. *AIChE J.* **2008**, *54*, 2479–2486.
47. Glatter, O.; Kratky, O. *Small Angle X-Ray Scattering*; Academic Press: London, 1982.

Atomistic Structure of Monomolecular Surface Layer Self-Assemblies: Toward Functionalized Nanostructures - Supporting Information

Christine Horejs,[†] Harald Gollner,[†] Dietmar Pum,[†] Uwe B. Sleytr,[†] Herwig
Peterlik,[‡] Alois Jungbauer,[¶] and Rupert Tscheliessnig*,[¶]

*Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna,
Austria, Faculty of Physics, University of Vienna, Austria, and Austrian Centre of Industrial
Biotechnology c/o Institute for Applied Microbiology, University of Natural Resources and Life
Sciences, Vienna, Austria*

E-mail: rupert.tscheliessnig@boku.ac.at

*To whom correspondence should be addressed

[†]Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

[‡]Faculty of Physics, University of Vienna, Austria

[¶]Austrian Centre of Industrial Biotechnology c/o Institute for Applied Microbiology, University of Natural Resources and Life Sciences, Vienna, Austria

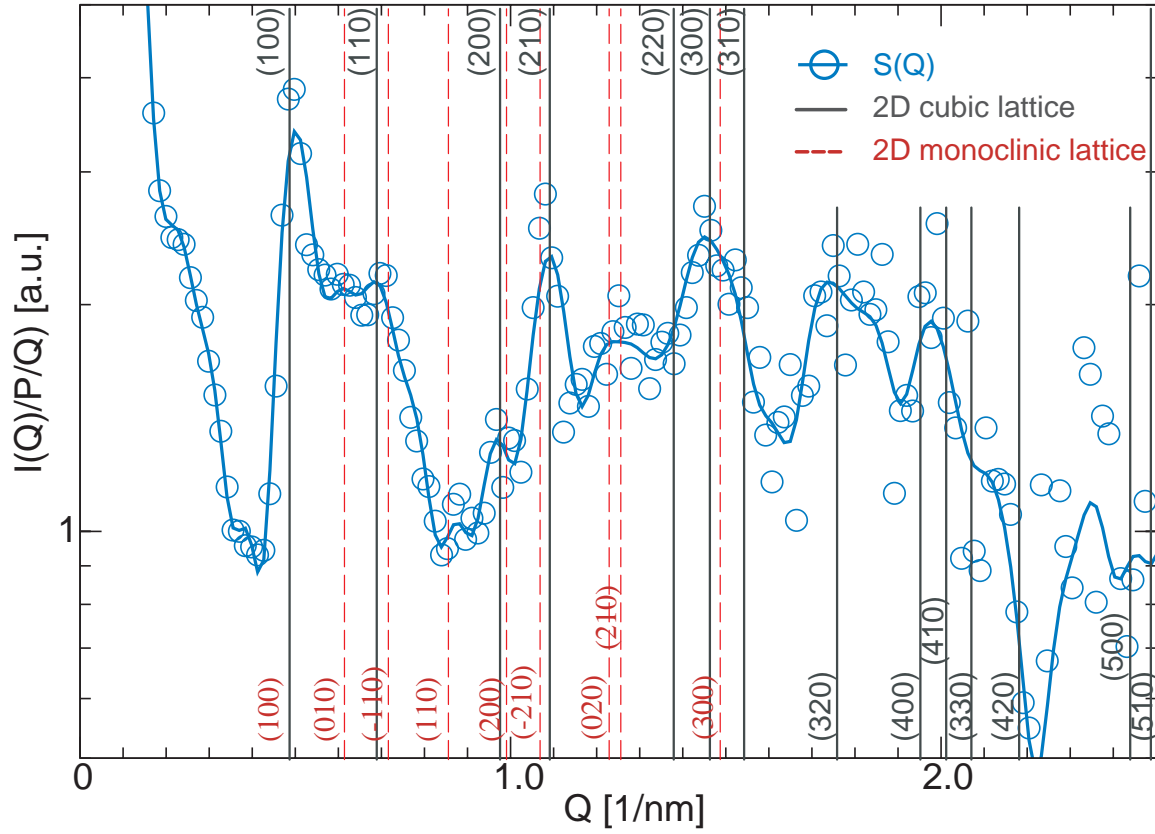


Figure S1: Detailed view of Fig.3. Open circles give scattering intensities divided by the form factor $I(Q)/P(Q)$. The red line is the fit curve from Eq.(1), which equals the structure factor as a function of the scattering vector Q . The corresponding Bragg peaks are indicated. The large number of additional small reflections is probably due to distortions close to the boundaries of the domains. The size of these domains is in a raw estimation about 200 nm, obtained from the 100 reflection and Scherrers formula for 2D planes.

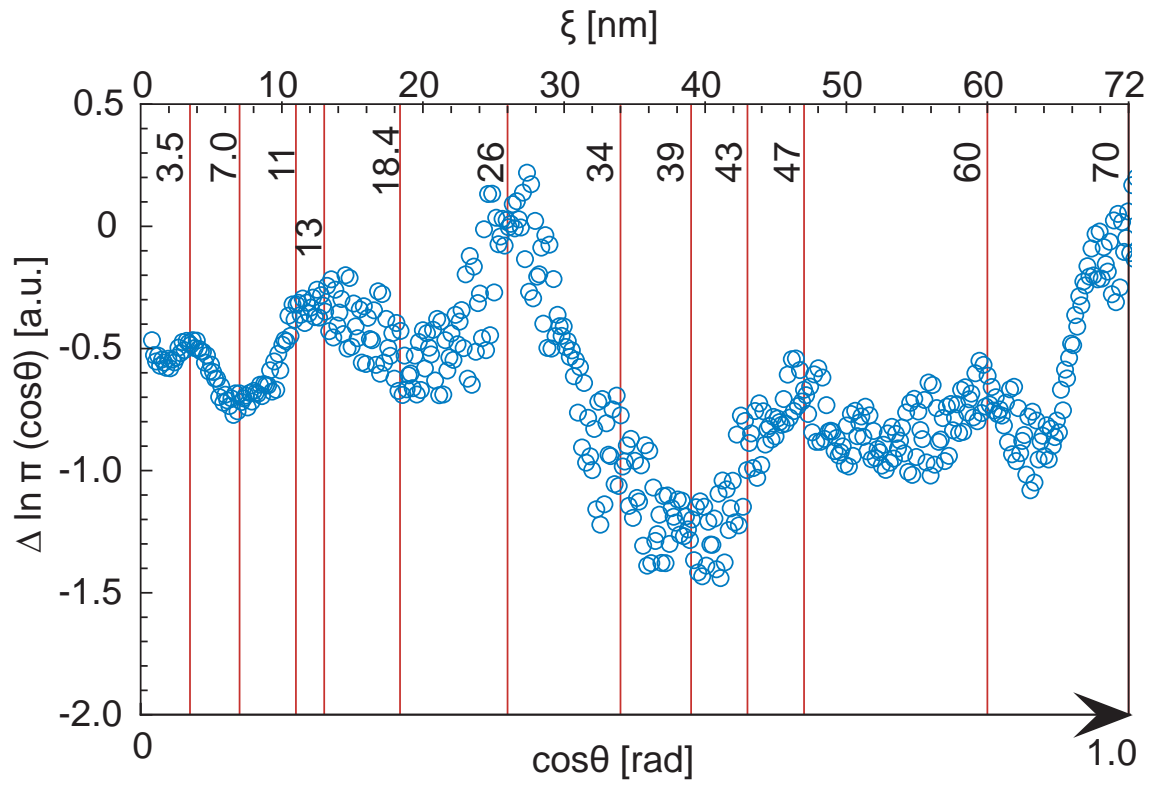


Figure S2: The mean potential is given as function of the reaction coordinate ζ . Here the reaction coordinate is interpreted as protein-protein distance [nm] and as relative orientation of the nanodisk with respect to the scattering vector. Maxima correlate to high local densities, i.e. 13 nm (unit cell dimension), while minima correspond to lower distributions of matter and indicate pores or domain boundaries. Some of these characteristic features are indicated by vertical lines.

Monte Carlo study of the molecular mechanisms of surface-layer protein self-assembly

Christine Horejs,¹ Mithun K. Mitra,² Dietmar Pum,¹ Uwe B. Sleytr,¹ and Murugappan Muthukumar^{2,a)}

¹*Department for Nanobiotechnology, University of Natural Resources and Applied Life Sciences, 1190 Vienna, Austria*

²*Department for Polymer Science and Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA*

(Received 17 November 2010; accepted 23 February 2011; published online 22 March 2011)

The molecular mechanisms guiding the self-assembly of proteins into functional or pathogenic large-scale structures can be only understood by studying the correlation between the structural details of the monomer and the eventual mesoscopic morphologies. Among the myriad structural details of protein monomers and their manifestations in the self-assembled morphologies, we seek to identify the most crucial set of structural features necessary for the spontaneous selection of desired morphologies. Using a combination of the structural information and a Monte Carlo method with a coarse-grained model, we have studied the functional protein self-assembly into S(surface)-layers, which constitute the crystallized outer most cell envelope of a great variety of bacterial cells. We discover that only few and mainly hydrophobic amino acids, located on the surface of the monomer, are responsible for the formation of a highly ordered anisotropic protein lattice. The coarse-grained model presented here reproduces accurately many experimentally observed features including the pore formation, chemical description of the pore structure, location of specific amino acid residues at the protein–protein interfaces, and surface accessibility of specific amino acid residues. In addition to elucidating the molecular mechanisms and explaining experimental findings in the S-layer assembly, the present work offers a tool, which is chemical enough to capture details of primary sequences and coarse-grained enough to explore morphological structures with thousands of protein monomers, to promulgate design rules for spontaneous formation of specific protein assemblies. © 2011 American Institute of Physics. [doi:10.1063/1.3565457]

I. INTRODUCTION

Self-assembly is one of nature's strategies to organize matter on a large scale and thereby create order from disorder.^{1–3} The process is ubiquitous for a great variety of biological molecules, such as lipids,^{4,5} DNA,⁶ polymers,^{7,8} self-assembled monolayers,⁹ and viruses,¹⁰ and understanding the driving forces behind this process is one of the central challenges in biological physics.¹¹ Proteins, however, tend to stay soluble in solution or aggregate into various structures rather than self-assemble into defined patterns. This is because of their complex structure exhibiting different conformations and a close-knit relationship between structure and function. Aggregation into three-dimensional composites thus generally leads to a loss of functionality.^{12,13} However, S(surface)-layer proteins represent a remarkable exception to this general trend.¹⁴ Another exception is the formation of essentially one-dimensional fibers from amyloidogenic proteins.¹⁵ The S-layer proteins crystallize into monomolecular arrays on the cell surface of a great variety of bacterial and all archaeal cells thereby providing the outermost cell envelope (S-layer). The crystallization of this kind of proteins then facilitates their function rather than forming a

nonfunctional state as is the case for most of the ill-aggregated complex proteins. The self-assembly process is also remarkably robust with the protein monomers forming patterns with defined lattice symmetries not only in their natural environment but also in solution and on a variety of surfaces and interfaces, which makes them an interesting object for the investigation of the design of biomolecular self-assembly processes that do not lead to aggregation.¹⁶

There has been a wide range of studies over the past few decades in order to investigate the genetics, structure, function, and nanotechnological applications of S-layer proteins. While this has provided considerable insight into the structure–function relationship of several S-layer proteins, a fundamental understanding of the molecular mechanisms guiding the self-assembly process has been elusive. Recent studies on the crystallization of an S-layer protein, which self-assembles into structures exhibiting lattices with a p4 symmetry, have given an insight into the kinetics and pathways of the S-layer self-assembly on surfaces. This work was based on high-resolution atomic force microscopy in combination with computer simulations.^{17,18} However, only recently, a combination of simulation and experimental techniques has enabled the determination of an atomistic structural model of one particular S-layer protein,^{19,20} SbsB from *Geobacillus stearothermophilus* pV71/p2, which assembles into two-dimensional sheets in solution exhibiting a lattice with one

^{a)} Author to whom correspondence should be addressed. Electronic mail: muthu@polysci.umass.edu.

monomer per unit cell (p1 symmetry).²¹ This structural model provides now the opportunity to investigate the specific interactions between protein monomers that lead to self-assembly as opposed to aggregation in solution.

Coarse-grained modeling of proteins has become an important tool to address various questions regarding their structure and function. Depending on the lengthscale of interest, different strategies of coarse-graining have been implemented.^{22–24} Further, depending on the questions that are of interest, these models can then be simulated using either Monte Carlo²⁵ or Langevin dynamics^{26–28} to investigate the properties of the system. In a recent study,²⁹ we have investigated the kinetics of amyloid fibrillization by combining a coarse-grained model of folded polypeptides and the lattice Monte Carlo method. The results obtained by this modeling methodology are in remarkable agreement with all phenomenological results. Propelled by this advance, we present here an analogous coarse-grained model for the assembly of S-layer proteins.

In the present study, we use a coarse-graining strategy in which every amino acid is represented by a single coarse-grained bead placed at the center of mass of the α -carbon and the side chain. We use off-lattice Monte Carlo simulations of this coarse-grained model to determine the interaction energies for a pair of monomers in terms of their relative orientation and separation distance. Using these energies as an input, we adopt the method of Ref. 29 to elucidate the molecular mechanisms leading to the lattice formation and to follow the kinetics of morphological assembly from hundreds of protein molecules. Our simulations afford an amino acid level understanding of the interactions between monomers and enable us to identify the essential residues required for self-assembly. At a broader level, the simulations also offer valuable insight into the role of specific interactions, such as hydrophobic or electrostatic interactions in the overall self-assembly process which might be applicable in a wider context.

The simulation results are in excellent agreement with known experimental results, while additionally offering the opportunity for predicting the behavior of possible recombinant proteins and different features of the self-assembly process in different environmental conditions. Our results demonstrate the important role that simple techniques, such as Monte Carlo simulations can play in understanding the complex phenomenon of protein–protein interactions. To the best of our knowledge this is the first theoretical study of the self-assembly processes leading to S-layers in solution based on atomistic structural details of one S-layer protein. This example might help to better understand the aspects of protein–protein interactions that are critically required for the formation of well-defined functional morphologies in solution.

II. MODEL AND SIMULATION METHOD

The modeling of the formation of a crystalline two-dimensional lattice from hundreds of protein monomers is carried out in three stages. In the first stage, a coarse-grained model of a monomer is generated from the atomistic details of all amino acid residues of the monomer. In this united-

atom coarse-grained model for the monomer, the polymer sequence and charge decoration on the amino acid residues, and their structural correlation with the tertiary structure are maintained. In the second stage, two such monomers are allowed to undergo rotation and translation, relative to each other, and the pairwise interaction energy was computed as a function of rotational and translational degrees freedom using a continuum Monte Carlo method. Only a couple of configurations of dimers dominate the energy landscape among numerous possibilities allowed by rotations and proximity. In the third stage, these two configurations and their energies are taken as an input in a lattice Monte Carlo method to follow the kinetics and morphology of spontaneously assembling structures from hundreds of the coarse-grained protein monomers. This three-stage multiscale simulation protocol allows the modeling of self-assembly from a collection of large numbers of protein monomers and at the same time maintaining the details related to the sequence and charge decoration of the protein monomer. The above mentioned three stages are described below.

A. Coarse-grained model

The coarse-grained model used in this work is based on the atomistic structure that we recently determined using *ab initio* molecular dynamic simulations¹⁹ and small angle x-ray scattering.²⁰ The protein monomer is made up of 920 amino acid residues. Figure 1(a) shows the surface structure of the protein. The S-layer protein is L-shaped, and contains an N-terminal alpha-helical part that is responsible for anchoring the protein to the cell surface [Fig. 1(a) red part], a C-terminal part that contains mainly beta-sheet structures [Fig. 1(a) blue part], and an unfolded part that links the two domains of the protein. The L-shaped form as well as the architecture of the surface accessible amino acids are supposed to be mainly responsible for the way these proteins self-assemble.

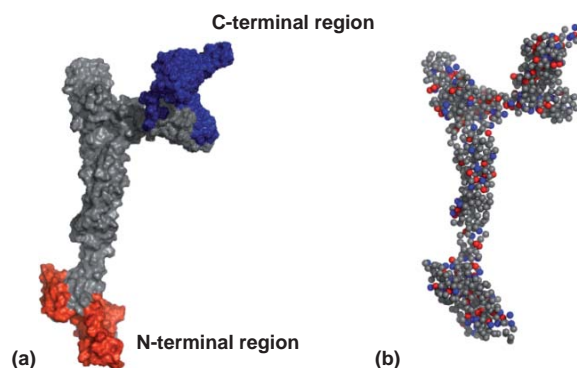


FIG. 1. S-layer protein SbsB from *Geobacillus stearothermophilus* pV72/p2. (a) Surface of an atomistic model calculated by molecular dynamics simulations (Ref. 19). The N-terminal region, which anchors the protein on the cell surface, is colored in red and the C-terminal region in blue. (b) Coarse-grained model used for the calculation of the self-assembly process. Each amino acid residue is represented by one bead of 0.65 nm diameter, which is located in the center of mass of the C_α and the side chain. Negatively charged beads are colored in red and positively charged beads are colored in blue. The protein is treated as a rigid body.

In order to model this protein, we represent every amino acid residue by a single coarse-grained bead placed at the center of mass of the α -carbon and the side chain, as shown in Fig. 1(b). This scheme thus preserves some information of the specific location of the side chain of the particular amino acid, and produces a faithful structural replica of the original protein. The folded protein monomer is treated as a rigid object, motivated by experimental observations that showed that the tertiary structure of the protein remains stable and is the same for monomers in solution as in self-assembled sheets.²⁰ We also account for the charges of the amino acid side chains, thus enabling us to study the effects of electrostatic interactions in the self-assembly process.

B. Dimer simulations

In order to determine the interprotein interaction energies, we performed off-lattice Monte Carlo simulations with two protein monomers. The two monomers were randomly placed inside a simulation box with hard walls, taking care only to ensure that the monomers did not overlap with each other. The interaction potential between the two monomers was initially taken as a sum of two contributions, a modified Miyazawa–Jernigan (MJ) contact potential, U_{contact} (Ref. 30) and an electrostatic energy term $U_{\text{electrostatic}}$ that accounts for the interactions between charged amino acids.

The MJ contact potential was developed to accurately model the interaction between amino acids based on a knowledge-based statistical analysis of known proteins in the protein database. These potentials have been shown to accurately capture the behavior of interacting amino acids in different situations.^{30,31} In the original form, every pair $\{ij\}$ of amino acids is assigned a contact energy denoted by M_{ij} (in units of $k_B T$), which is defined as the energy difference accompanying the formation of contacts between i and j types of amino acids from those amino acids exposed to the solvent. Thus, the contact energy is given by M_{ij} if the two amino acids are closer than some threshold distance, and is zero if they are larger than this threshold. However, physically, since a complex amino acid residue is represented by a single coarse-grained bead, there is no sharp boundary region for the bead, and to model this effect, we softened this contact potential by a sigmoidal function (S_{ij}) over an interaction range defined by a lower bound (r_{LB}) and an upper bound (r_{UB}). The contact potential can then be written as the product of two terms,

$$U_{\text{contact}} = \sum_{i \in A} \sum_{j \in B} M_{ij} S_{ij}, \quad (1)$$

where the sum runs over all possible amino acid pairs between the two monomers A and B . The sigmoidal function S_{ij} is defined such that the potential varies smoothly between M_{ij} at the lower bound and zero at the upper bound,

$$S_{ij} = \left(\frac{1}{1 + e^{r_{ij}}} - \frac{1}{1 + e^{r_{\text{UB}}}} \right) / \left(\frac{1}{1 + e^{r_{\text{LB}}}} - \frac{1}{1 + e^{r_{\text{UB}}}} \right). \quad (2)$$

In our simulations, the lower cutoff was chosen to be $r_{\text{LB}} = 0.65$ nm. This was chosen as an optimum value since smaller values resulted in unphysical highly interpenetrating conformations, whereas larger values did not yield any stable ground state conformations. This lower cutoff can also be interpreted as the radius of the coarse-grained bead. For the upper cutoff, a value of $r_{\text{UB}} = 0.9$ nm was chosen to ensure conformity with previous work on coarse-grained models of proteins.²⁸ This softening of the potential is not expected to significantly alter the results,³⁰ but was nevertheless introduced to take into account the coarse-grained nature of the amino acid residues in our model. The MJ values for the contact potential were chosen after a comparison with alternative formulations of the contact energies by Thomas and Dill.³² Their classification of amino acids in 3, 5, 10, and 20 groups was tested. It was found that for our protein, only the MJ values for the potential led to stable dimer conformations in the simulations.

For the electrostatic interaction, we used the Debye-Hückel potential (in units of $k_B T$):

$$U_{\text{electrostatic}} = \sum_{i \in A} \sum_{j \in B} \frac{q_i q_j \ell_B \exp(-r_{ij}/\kappa)}{r_{ij}}, \quad (3)$$

where $\kappa = 1$ nm is the Debye length and $\ell_B = 1$ nm is the Bjerrum length, corresponding to the dielectric constant of 56 for the ambient solution. The choice of this value for the dielectric constant is simply to take ℓ_B and κ as 1 nm. A slight variation in the value of ℓ_B is not expected to change the physical conclusions described below. The charges on the individual amino acid residues are denoted by q_i (± 1 or 0), and are taken to be their standard charges at physiological pH, relative to the individual pK_a values of the residues.³³ For the purpose of our simulations, the histidine residues in the monomer sequence were taken to be positively charged.

Using these two potential terms, we then simulated two protein monomers to obtain the minimum energy conformations. The two monomers were placed in a simulation box of size $8L$, where $L \simeq 227$ nm is the typical size of the protein, defined as the length of the space diagonal for the minimal rectangular cuboid that can contain the protein monomer. It was ensured that the monomers did not overlap with each other in the initial configuration. We then performed a standard Metropolis Monte Carlo simulation.³⁴ A trial move for the simulation consists of a combination of rotation and translation moves. A translation move involves shifting all the amino acid residues in the monomer by a randomly chosen displacement, with a maximum value of the displacement being given by $0.1L$, to ensure that the two monomers did not undergo large translations in a single move. A rotation move consists of randomly choosing a set of Euler angles (ψ, θ, ϕ) , where $\psi, \phi \in [-\pi, \pi]$ and $\theta \in [-\pi/2, \pi/2]$. The whole monomer is then rotated in accordance with the rotation matrix defined by these Euler angles $R(\psi, \theta, \phi)$. The trial move is considered valid if no amino acid residues overlap with any other in the resultant conformation, and if both the monomers lie within the simulation box. A valid move is then accepted or rejected according to the standard Metropolis rules.

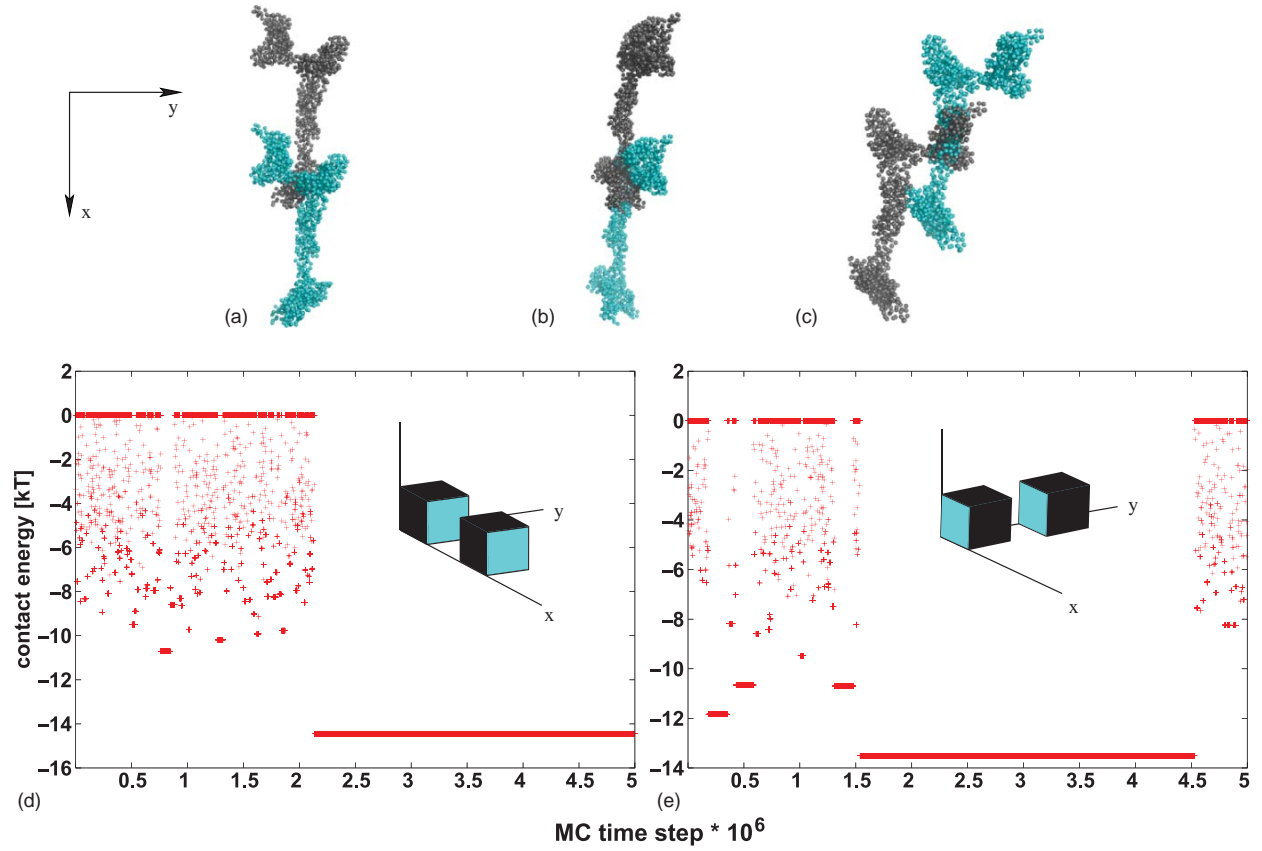


FIG. 2. Monte Carlo simulation of the interaction of two monomers resulting in two stable dimer conformations. (a, b) Hook conformation: one monomer is hooked to the other one by its N-terminus along the x -direction. (c) Parallel conformation: The two monomers associate along the y -direction and form a charged hydrophilic pore. (d, e) Sample contact energy profiles. (d) The stable energy minimum at $-14.5k_B T$ corresponds to the hook conformation. (e) The stable energy minimum at $-13.8k_B T$ corresponds to the parallel conformation. In these simulations, the amino acid residues were assigned their standard charges at physiological pH with histidine taken to be positively charged. Insets: Dimer conformations are represented by unit cubes (as a mnemonic representation) to be used in the lattice Monte Carlo simulation of the large-scale assembly process. These are provided to explain the transition from the off-lattice Monte Carlo scheme for two monomers to the lattice Monte Carlo scheme for simulating hundreds of monomers. Blue colored faces represent the interaction directions.

As described in the next section, we found that there were two minimal energy conformations for the dimers (see Sec. III A and Fig. 2). In addition, we also observed a third highly interpenetrating conformation, in which the two monomers lie almost on top of each other, but are slightly displaced, ensuring that the self-avoidance constraint of the individual amino acid beads is satisfied. This is clearly an unphysical conformation, which arises due to the coarse-grained nature of the model where we replace spatially extended amino acid side-chains by a finite bead, which leaves spurious empty spaces where another monomer can penetrate. Interestingly, it was observed that in all these minimum energy conformations, the two monomers lie in the same plane with the same orientation. It thus appeared that rotation moves were not important in determining the ground state conformations for our particular protein. As a result, in all subsequent runs, a trial move was taken to consist of a simple random translation, which considerably decreases the running time of our simulations, since scanning the general phase space of all rotational angles is unnecessarily expensive in computational time.

In order to exclude the unphysical interpenetrating conformation described above, we introduce an additional cutoff

term, U_{cutoff} , which specifies the minimum approach distance between the centers of mass of the two monomers. The form of this term is considerably simplified by virtue of the fact that the minimum energy conformations were found to lie in a plane, and hence the effect of rotations is not important, allowing us to use a single value of the cutoff parameter to exclude highly overlapping conformations. This effect is then modeled by

$$U_{\text{cutoff}} = \begin{cases} 0 & r_{CM}^A - r_{CM}^B > \eta L \\ \infty & r_{CM}^A - r_{CM}^B < \eta L \end{cases}, \quad (4)$$

where, $r_{CM}^{(A,B)}$ denotes the center of masses of the two monomers and η is the cutoff parameter. Multiple trial runs were performed using different values of η , and we chose the minimum value which did not result in the unphysical third state (described above) as one of the ground state conformations. This threshold is given by $\eta = 0.15$ in our simulations.

Our final potential energy is then given by a sum of these three terms,

$$U_{\text{total}} = U_{\text{contact}} + U_{\text{electrostatic}} + U_{\text{cutoff}}, \quad (5)$$

where the monomers are assumed to undergo only translation moves (as rotations are found to be irrelevant for the energy minima for a pair of monomers studied here). The ground state conformations of the dimer system were then obtained using this potential via the standard Metropolis Monte Carlo algorithm. Indeed, the energies of the stable states of the dimer are the same as the results without the cutoff and fully allowed rotations of the monomers.

C. Lattice Monte Carlo

In order to gain an insight into the kinetics of the growth process and the morphology of the final assembled structures, we also performed a lattice Monte Carlo using the results of the dimer simulation. The detailed off-lattice dimer simulations revealed that there are two ground state conformations that are dominant, and in these conformations, the two monomers lie in the same plane and with the same orientation. Then, given that rotations were not significant for the interaction among monomers for this protein, we use a mnemonic representation, whereby the protein monomers are represented as cubes, randomly distributed in a cubic lattice which constitutes the simulation box. This is only a mapping for computational ease, and maintains the off-lattice pairwise interaction energies of the states with energy minima. The interaction energies are extracted from the dimer simulations as the ground state energies, and in the lattice simulations these are assigned according to the direction of interaction of the two monomers. This is explicitly shown in the insets of Figs. 2(d) and 2(e), where the blue faces represent the interaction directions corresponding to the two ground states (as discussed below in Sec. III). A trial move then consists of selecting a monomer at random and displacing it to a random position on the cubic lattice. The trial move is considered valid if the attempted new position is unoccupied. If in the new position, the monomer is a neighbor to another monomer, whether in isolation or as a part of a cluster, it is assigned an interaction energy depending on which two faces are in contact. The values of these different interaction energies are obtained as an input from the dimer simulations, as mentioned above. The trial move is then accepted or rejected according to the standard Metropolis rules. The moves are only the nearest neighbor motions, where the self-assembled structures are not allowed to make concerted moves. For the lattice Monte Carlo, we have taken the simulation box to be a cube of linear size 50 in units of the size of each monomer.

We have performed two sets of lattice Monte Carlo simulations for a given protein concentration. In the first, a seed is placed at the center of the simulation box, and the monomers are allowed to adhere only to the seed and its continuously growing growth surfaces with the pairwise interaction energies as obtained above. Here, there is only one growing cluster and there is a continuous depletion of the monomers from the solution without any competition from any other growing cluster (as this is deliberately suppressed in the first set of simulations). We have monitored the growth rate as a function of protein concentration. In the second set, there are no seeds, and the monomers adhere to each other to form multiple clus-

ters which then coarsen among themselves competitively. We have monitored the time evolution of cluster size distributions in the second set of simulations.

III. RESULTS AND DISCUSSION

A. Interaction of two monomers

The S-layer protein under investigation typically assembles into two-dimensional sheets in aqueous solutions, with the monomers arranged in a p1 lattice symmetry.²¹ To analyze possible conformations between protein monomers that finally lead to the two-dimensional sheets, we first simulated the interaction between two monomers. In addition to establishing the primary stable conformations of a pair of interacting monomers, we have monitored the relative weights of hydrophobic and electrostatic interactions by deliberate manipulation of charges on the amino acid residues. Furthermore, we have also investigated the role of different segments of the primary sequence of the monomer in determining the most stable conformations of the dimer by, for example, cutting out a segment in the N-terminus of the monomer.

The off-lattice Monte Carlo simulations show the presence of two dominant ground state conformations (Fig. 2). In the first conformation [Figs. 2(a) and 2(b)] the proteins are hooked into each other and in the second conformation [Fig. 2(c)] the monomers are arranged parallel to each other while establishing a pore. As marked in Fig. 2, we assign the x and y directions to correspond to the direction of hooking and the direction of parallel arrangement, respectively. Two typical contact energy plots are given in Figs. 2(d) and 2(e). The energy minimum in Fig. 2(d) corresponds to the hooked conformation, and the one in Fig. 2(e) corresponds to the parallel state. As explained in Sec. II, the insets in Figs. 2(d) and 2(e) denote, respectively, the most favorable approaches along the x -axis and y -axis corresponding to the hooklike and parallel conformations of the dimer.

In addition to the two ground states, the energy plots also show the existence of many metastable states with higher contact energy [Figs. 2(d) and 2(e)]. Some of these metastable conformations are shown in detail in Fig. 3. Most conformations resemble the ground state conformations but are slightly shifted toward x - or y -direction, suggesting that they may play a role in further stabilizing the spontaneously formed lattice toward its final minimum energy conformation. We know from transmission electron microscopy studies in solution as well as from atomic force microscopy investigations on the S-layer formation on surfaces that these proteins do not tend to form three-dimensional aggregates under appropriate conditions as is the case in our simulations. The self-assembled sheets rather form patches with defined edges on solid or lipid surfaces^{35,36} or various different structures like sheets, cylinders, or tubes in solution.³⁷ To investigate the role of those metastable states in the process of the lattice formation, future studies using replica exchange and other simulation methods might be interesting to get information about the underlying free energy landscape. However, for the present work, these metastable states are not included in the large-scale simulations as the main focus of our work lies

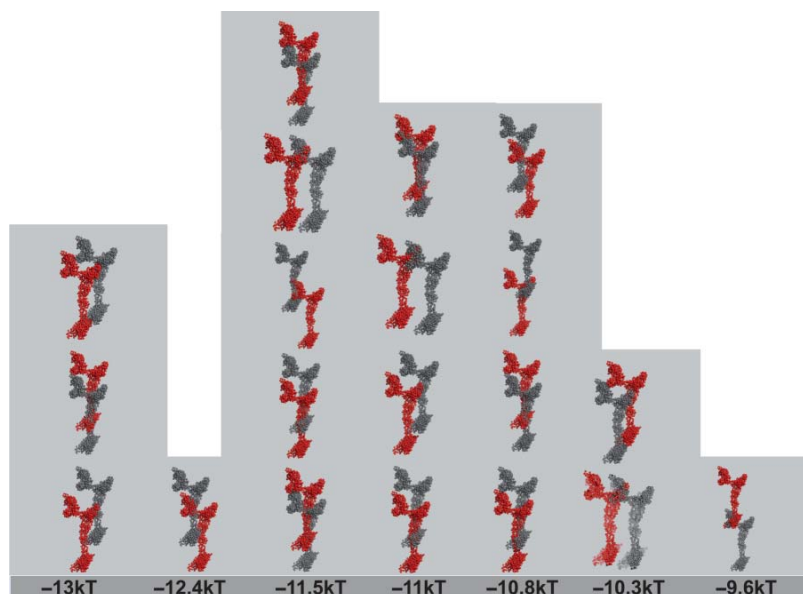


FIG. 3. Metastable states as obtained from the dimer simulations with different contact energy values. Most conformations are similar to the hook or to the parallel ground state, but slightly shifted in x - or y -direction.

on the structural details of the lattice in its minimum energy state.

It has been shown experimentally that the N-terminal region of the protein, comprising up to 208 amino acid residues, can be truncated without hindering the formation of the p1 lattice.³⁸ In order to find whether our coarse-grained model has the capacity to reproduce this experimental fact, we carried out the dimer simulations of the coarse-grained model without this N-terminal region, and indeed the ground state conformations were the same as that found for simulations of the entire protein. As a result, removal of the N-terminus with 208 amino acid residues is not detrimental to the assembly of a layerlike structure.

Additionally, the form of our interaction potential allows us to investigate separately the role of hydrophobic and electrostatic interactions in the self-assembly process. To this end, we performed simulations in which all amino acid side chains were kept uncharged. Interestingly, this did not affect the two ground state conformations obtained in the presence of electrostatic interactions. This result suggests that hydrophobic interactions are the dominant mechanism responsible for the self-assembly of the proteins.^{39–41} Analysis of the ground state conformations also suggests that it is the amino acid residues on the surface of the monomer that are responsible for the dimer formation. Further support for this conjecture is provided by simulations performed with models with a second level of coarse-graining, where only hydrophobic amino acid side chains located on the surface of the monomer were taken into account, which resulted in the same two conformations for the ground state as the full protein. This stripped coarse-grained model contains only 137 amino acids. This small subset of necessary amino acid residues then explains the robustness of the S-layer self-assembly

process with respect to genetic engineering and fusion proteins.

Experimentally, it is also known that S-layer proteins self-assemble in aqueous solutions independent of environmental conditions, such as pH , temperature, and salt concentration, which is necessary in order for them to carry out their function *in vivo*. Although charged groups are present on the surface of the monomers, our results show that the hydrophobic interactions are much stronger than the electrostatic interactions and only a few hydrophobic surface sites are involved in the required protein–protein interactions. This dominance of the hydrophobic interactions then presents an explanation for the robustness of the assembly process across the pH range, since this affects only the electrostatic interactions, which do not play a crucial role in the layer formation.

B. Large scale assembly

In order to investigate the growth kinetics of the large-scale assembly of S-layer sheets, we performed lattice Monte Carlo simulations using the energies of the two dominant ground state interactions obtained from the dimer simulation. The lattice simulations were performed in two ways. In order to investigate the growth rate, we performed seeded growth studies at different monomer concentrations. In these simulations, one protein monomer, the seed, was held fixed in the center of the simulation box, and the monomers are allowed to interact with the ground state energies only if they are part of the cluster which also comprises the seed monomer. The simulations were started from a random initial configuration [Fig. 4(a)], and the final configuration is a single

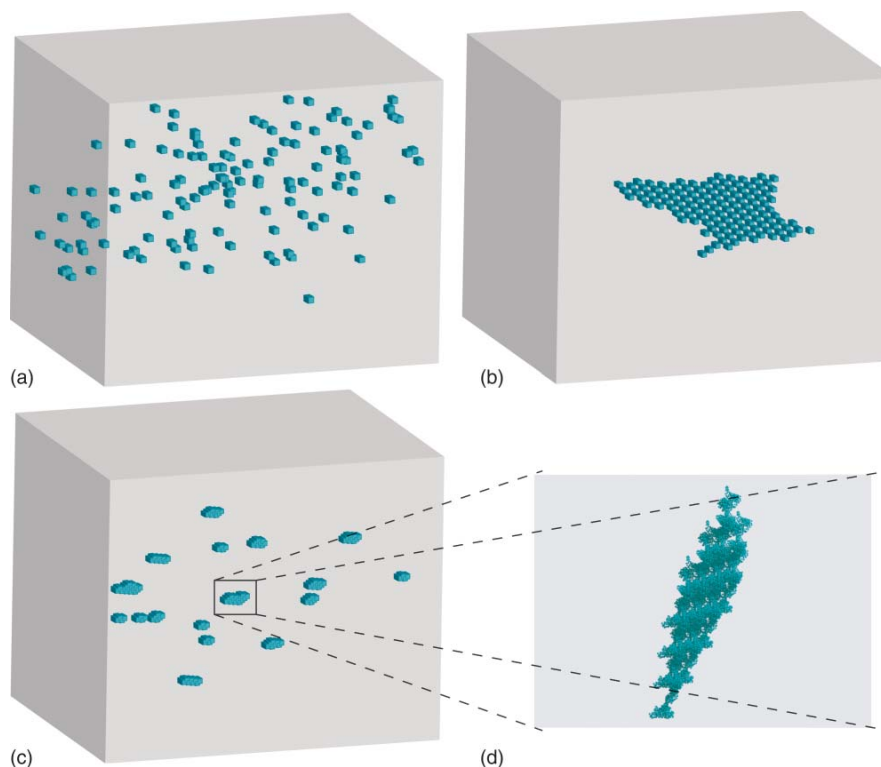


FIG. 4. Representation of the lattice Monte Carlo simulations of the large-scale self-assembly process. Energy values were taken from the interaction of two monomers. Each protein is represented by a unit cube, which is only a mnemonic representation. (a) Initial configuration: cubes are randomly distributed in the simulation box. (b) Single-seeded growth study. One monomer was kept fixed as the seed. Proteins interact with the seed resulting in a self-assembled two-dimensional sheet. (c) Competitive growth study. No seed was introduced to the system. Multiple sheets start to grow during the initial period. (d) Magnified view of the corresponding S-layer sheet using the coarse-grained model.

assembled sheet [Fig. 4(b)]. The number of monomers in the growing cluster increases with time before it reaches the saturation limit defined by the number of monomers in the system. This growth curve is shown for three different concentrations of monomers in Fig. 5(a). The growth curves are characterized by the absence of any lag time, which implies spontaneous assembly into layers. The growth curves show the typical features of a brief sigmoidal part for very short times, a linear dependence for a substantial time duration, and a saturation regime at very long times. The slope in the linear regime is taken as the growth rate. For the monomer concentrations $c = 0.001$, 0.0015 , and 0.002 (corresponding to 125, 187, and 250 protein monomers, respectively), the growth rates are found to be 0.473×10^{-2} , 0.765×10^{-2} , and 1.18×10^{-2} , respectively (in units of number of monomers in the cluster per unit time). Thus, we have found that the growth rate increases linearly with the concentration of the monomer.

We have also studied the competitive growth of different S-layer sheets in the general case when no seed is introduced into the system. Sheets can now nucleate at random throughout the system, and multiple sheets start to grow during the initial period. It must be noted that we do not observe any nucleation barrier for the present situation, unlike the situation with the amyloid fibrillization.²⁹ As free monomers be-

come depleted from the system, the different sheets compete with each other, and the growth of larger sheets takes place at the expense of dissolution of the smaller sheets, through the familiar Ostwald ripening or coarsening process. A sample configuration of such a system is shown in Fig. 4(c). We have also studied the distribution of cluster sizes in this case, and the distribution goes from a sharply peaked distribution about unity (corresponding to the monomers) for short times to a broader distribution peaked about a higher cluster size at later times [Fig. 5(b)]. As an example, at the Monte Carlo time of 10^6 , the cluster size distribution is peaked at 6 with a very long tail extending to the number of monomers inside the lattice to be even more than 20. A decade of time later, the cluster size distribution is still peaked at 6, but now there are more clusters with higher number of monomers in them. This is analogous to our previous study on amyloid growth.²⁹ In the present study, we do not attempt to establish the time-exponent for the coarsening kinetics,⁴² by relegating this to a future report.

C. Morphology of the assembled sheet

Figure 6 shows the resulting lattice structure. The architecture of the lattice exhibits identical pores with a diameter of 3.25 nm, which is in good agreement with

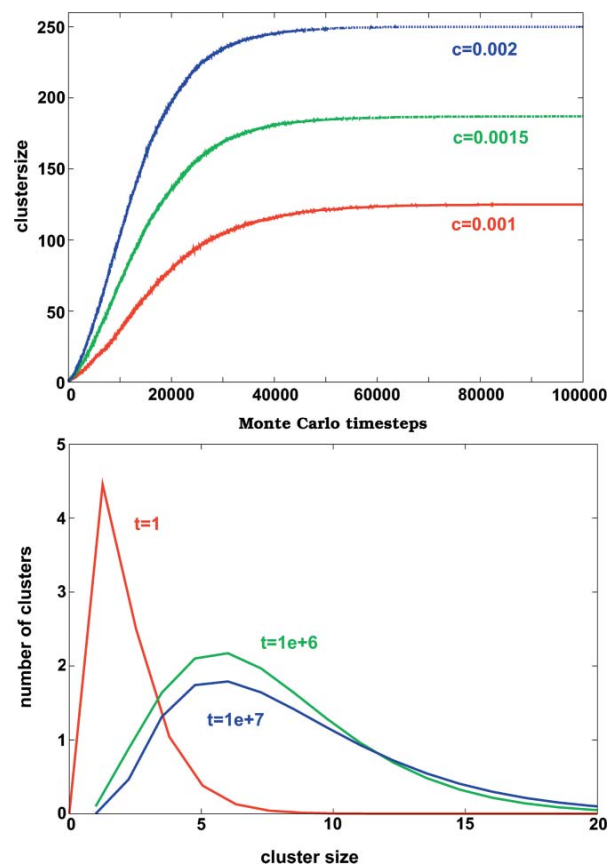


FIG. 5. Large-scale self-assembly. (a) Growth rate of the seeded cluster for different monomer concentrations. (b) Number distribution of the cluster size in the competitive growth case. The plots were smoothed using a Bezier curve. The initial distribution at $t = 1$ is scaled down by a factor of 10.

experimentally determined values.²¹ The pores consist of mainly hydrophilic residues (magnified view in Fig. 6: blue beads), among which six are positively charged and three are negatively charged. Altogether 24 residues contribute to the pore. These residues are R100, V104, R187, G188, D189, Q192, T332, S333, S334, and D335 of the first monomer (mainly the N-terminal part), and S476, K478, A479, S480, F481, V483, F485, D487, K490, R491, T492, F493, K746, and L747 of the second monomer (mainly the C-terminal part). These pores are responsible for the transport of ions and other molecules into and out of the underlying cell, thereby functioning as selective ion gates, both with respect to charge as well as size, as has been experimentally determined.⁴³ The resulting lattice is 4.5 nm thick, which is consistent with experimentally measured values.²¹ The lattice parameters as determined by small angle x-ray scattering and TEM experiments are $a = 9.9$ nm, $b = 7.6$ nm, and $\gamma = 81^\circ$.²¹ Our simulations yield parameter values of $a = 11.1$ nm, $b = 6.4$ nm, and $\gamma = 88^\circ$. This difference might arise due to the local conformational adaptivity of the monomers, which was not included in this study. The parameters are in reasonable agreement given the simple nature of our model.

The final morphology of the p1 lattice clearly shows that the N-terminal region [Fig. 6(b) in red] and a large part of the

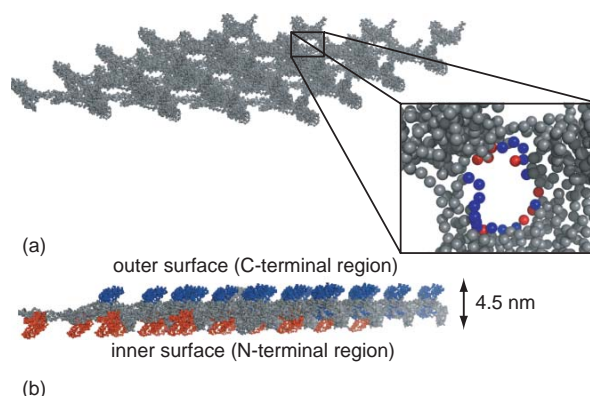


FIG. 6. Morphology of the self-assembled S-layer lattice as obtained by lattice Monte Carlo simulations. (a) Lattice made up of 16 protein monomers, represented by the coarse-grained model. The highly ordered structure is arranged as a p1 lattice and exhibits pores of identical size and morphology. The magnified view shows details of the pore. Hydrophilic residues are colored in blue while hydrophobic residues are colored in red. The pore is mainly hydrophilic with a net positive charge. Altogether 24 residues contribute to the pore. (b) Cross section of the lattice. The lattice has a width of 4.5 nm. The outer surface is composed of the C-terminal region (blue), the inner surface (*in vivo* anchored to the cell membrane) is composed of the N-terminal region (red). This architecture leads to the anisotropy of the lattice due to the different charge composition of the two terminal regions.

C-terminal region [Fig. 6(b) in blue] extend out of the lattice plane and are not directly involved in the interactions responsible for the formation of the lattices. S-layer fusion proteins have been produced by fusing the partner molecule to the N-terminus of the protein.^{44,45} These recombinant proteins did not lose their ability to self-assemble. The location of the N-termini in the lattice explains this behavior, because possible fusion partners do not disturb the necessary interactions directly.

The fact that both the N-terminal and the C-terminal regions are separated from the lattice surface leads to an insight about intersheet interactions. Based on the pK values³³ of the amino acid residues and the pH of the solution in which the monomers assemble, different levels of stacking of layers can be expected. Although the individual protein monomer has essentially a zero net charge under physiological pH conditions, the C-terminal region (300 aa) has a net negative charge (58 negatively charged and 26 positively charged residues) with mainly hydrophilic residues, whereas the N-terminal region (210 aa) has a very small net positive charge (28 positively charged and 23 negatively charged) with mainly hydrophobic amino acids [Fig. 1(a)].^{35,46–49} This anisotropy of charge distributions gives an insight about the formation of bilayers and even multilayers. The interactions between two layers in solution can change depending on the pK values of the specific residues, and hence can affect the assembly along the direction perpendicular to the sheets. Below a pH of 3.5, both surfaces of the S-layer lattice are positively charged. This implies that the individual S-layer lattices repel each other to yield individual monolayers as the final assembled structure. Conversely, at a $pH \gtrsim 13$, both surfaces of the lattice are negatively charged, and again we expect monolayers. Near physiological pH , all polar residues are fully charged, and hence

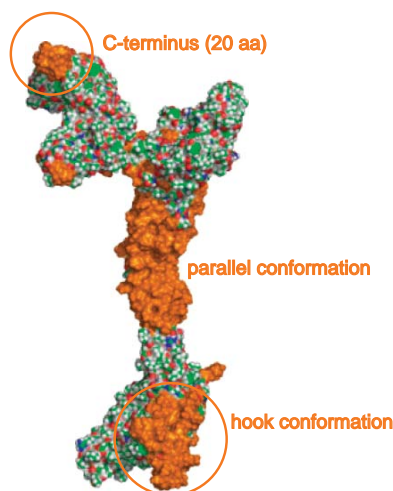


FIG. 7. Structural model of the S-layer protein SbsB. Amino acid residues involved in interactions in the pI lattice are colored in orange. Twenty residues at the C-terminus are part of interactions as well as 35 residues near the N-terminus, which are involved in the formation of the hook conformation. The mainly unfolded domain in the middle part of the protein is part of the interactions required for the formation of the parallel dimer conformation. All orange colored residues are essential for establishing a stable lattice structure.

the C-terminal face of the assembled lattice has a net negative charge, while the N-terminal face has a net positive charge. In this scenario, oppositely charged faces of individual lattices attract each other, and the final structure is a multilayer of individual S-layer lattices. For pH ranges close to the pI of one protein ($pH \sim 5$), the inner surface of the lattice is nearly neutral, while the outer surface is net negatively charged. In this situation, the two neutral faces can come together to form bilayers. We therefore conjecture that the S-layer sheets will assemble as monolayers in the very basic and very acidic regimes, double layers close to the pI of the protein (5.2) and multilayers close to the physiological pH (7.0).

In order to quantitatively estimate how many amino acid residues are essential for the dimer formation, we define a sphere of radius four times the bead size around every amino acid residue. If any amino acid residue from the second monomer lies within this “overlap sphere,” we defined the two residues to be part of overlaps. The overlap criterion as characterized by the radius of this sphere was determined as the one which gave physically realistic answers for the overlapping amino acid residues. It was found that altogether 241 amino acid residues of one monomer are involved in overlaps in the lattice, which corresponds to only 26% of the total number of residues. Figure 7 shows all required amino acid residues in detail (colored in orange). Analysis of the overlapping amino acid residues required for self-assembly shows that the very end of the C-terminus is required for the stability of the lattice structure (Fig. 7), which is in agreement with experimental observations.⁴⁴ This C-terminal region required for self-assembly is composed of 20 amino acid residues, with 15 of these being hydrophobic, three positively charged, and two negatively charged. This region plays a crucial role in the formation of the parallel dimer conformation, and cannot

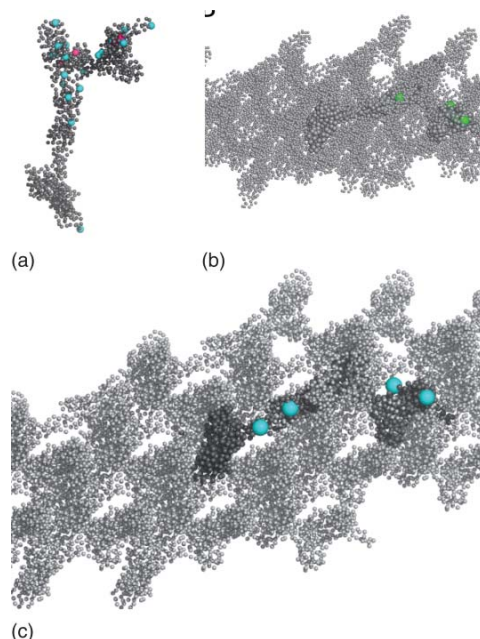


FIG. 8. Comparison of the calculated monomer and lattice structure with experimental results addressing the accessibility of individual residues (Ref. 51). (a) Twenty three residues have been experimentally determined as being accessible in the monomer structure. In the calculated structure 19 out of these 23 amino acids are located on the surface, which are colored in blue. Four residues are not accessible in the simulated monomeric model, which are represented by pink beads. (b) Residues that were experimentally determined to be part of intermolecular cross-links in the S-layer lattice are also seen in our model as represented by green beads. (c) Out of the five residues that were experimentally determined to be very accessible at the outer surface of the lattice, four are given by our simulations as shown in blue color.

be truncated without loss of the ability to self-assemble correctly. The predominance of hydrophobic amino acid residues (75%) in this interacting part supports the hypothesis of hydrophobic interactions being regarded as the driving force for the self-assembly of S-layers. Thirty five amino acid residues near the N-terminal region (outside the 208 residues which can be deleted without affecting the self-assembly) and a part in the middle of the protein (Fig. 7 orange amino acids) are also essential parts of the interactions within the lattice by providing hydrophobic side chains for the formation of the two dimer ground states.

We also compared our lattice architecture with the results obtained by recent experimental studies addressing the surface accessibility of amino acid side chains of this S-layer lattice,^{50–53} which is shown in Fig. 8. In this study, single-cysteine mutants of the protein were generated and then analyzed using the substituted cysteine accessibility method. Using this method, the experiments identified 23 residues that were accessible in the monomer, which implies that they are located on the surface of the monomer. In our coarse-grained model, 19 out of these 23 amino acid residues were found to be located on the surface [Fig. 8(a)]. The small discrepancy can arise either due to minor errors in the original structural model, or because of the nature of the coarse-graining strategy. Since the amino acid side chains are extended objects, some of these side chains might be accessible in the

experiments, even though the center of masses of these amino acids, corresponding to the position of our coarse grained beads, can lie just underneath the surface. Nevertheless, the success rate is appreciable, given the coarse-grained nature of our model. Further, these 23 amino acid residues were subject to a chemical cross-linking screen to identify their position in the assembled lattice. It was found that five out of these 23 amino acids were very accessible on the outer surface of the assembled S-layer lattice. In our simulations, we found that four out of these five residues are located on the outer surface [Fig. 8(c) blue beads], with the single residue that did not match being a part of the original four (out of 23) amino acid residues that did not match the surface accessibility data of the monomer. Out of the remaining 18 residues that are not on the outer surface of the assembled lattice, it was postulated experimentally that four are located at the interface between the interacting monomers in the lattice (intermolecular cross-links). All four of these were found to lie at the interface in our model also [Fig. 8(b) green beads]. These results demonstrate that the correct location of amino acid side chains was captured in the simple coarse-grained model of the protein, and thus offers a simple strategy to identify surface-accessible residues in the S-layer assemblies in place of extensive mutagenesis and chemical cross-linking studies.

IV. CONCLUSIONS

Our simulations of the self-assembly process of S-layer proteins using a coarse-grained model and Monte Carlo algorithms reveal the molecular structure of an S-layer lattice and the nonbonded interactions driving the self-assembly of these proteins. Our results reproduce all of the essential features observed experimentally: the anisotropy of the highly ordered lattice structure, the presence of hydrophilic pores bearing a net positive charge, the significant difference of the C- and N-terminal regions regarding the stability and interactions in the lattice, and the location of individual residues on the surface or at the interface of the two-dimensional sheets.

The underlying molecular mechanisms leading to the self-assembly of S-layer proteins are guided by the interaction of only few hydrophobic amino acid residues located on the surface of the proteins. The importance of hydrophobic interactions for the self-assembly of molecules is a well known phenomenon for lipids, small peptides, and other simple inorganic molecules. Our simulations show that the same basic principles also hold true for S-layer proteins, despite being larger in size and much more complicated in their tertiary structure, and can explain their self-assembly into functional structures with a defined morphology. S-layer proteins of different organisms differ enormously in their amino acid sequence and do not show any significant sequence homology to other protein classes. Nevertheless they all self-assemble into a variety of different lattice structures, and it would be interesting to study if the assembly in these proteins is also guided by the same hydrophobic interactions, and hence adhere to these common design rules.

In addition, our work also has potential applications in nanobiotechnology. Our analysis enables us to identify the essential amino acid residues required for the lattice formation.

The rest of the sequence may be altered by introducing functionalized groups for different applications. Thus, this provides a tool in designing recombinant proteins for structures with desired features on the large scale by a specific variation of the primary sequence.

ACKNOWLEDGMENTS

Acknowledgment is made to AFOSR (FA9550-10-1-0159), AFOSR (FA9550-09-1-0342), NSF (DMR-0706454), and the Materials Research Science and Engineering Center (MRSEC) at the University of Massachusetts for support. C.H. holds a DOC-fFORTE fellowship of the Austrian Academy of Sciences.

- ¹G. M. Whitesides and B. Grzybowski, *Science* **295**, 2418 (2002).
- ²G. M. Whitesides and M. Boncheva, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4769 (2002).
- ³J. Israelachvili, *Intermolecular and Surface Forces*, 2nd ed. (Academic, London, 1991).
- ⁴R. Lipowsky, *Nature (London)* **359**, 475 (1991).
- ⁵E. Sackmann, *Structure and Dynamics of Membranes: Generic and Specific Interactions* (Elsevier, Amsterdam, 1995).
- ⁶S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, *Nature (London)* **459**, 414 (2009).
- ⁷M. Muthukumar, C. K. Ober, and E. L. Thomas, *Science* **277**, 1225 (1997).
- ⁸J. T. Chen, E. L. Thomas, C. K. Ober, and G.-P. Mao, *Science* **273**, 343 (1996).
- ⁹D. K. Schwartz, *Annu. Rev. Phys. Chem.* **52**, 107 (2001).
- ¹⁰D. G. Angelescu and P. Linse, *Soft Matter* **4**, 1981 (2008).
- ¹¹M. McCullagh, T. Prytkova, S. Tonzani, N. D. Winter, and G. C. Schatz, *J. Phys. Chem. B* **112**, 10388 (2008).
- ¹²C. M. Dobson, *Nature (London)* **426**, 884 (2003).
- ¹³C. M. Dobson, *Nature (London)* **418**, 729 (2002).
- ¹⁴U. B. Sleytr, *Nature (London)* **257**, 400 (1975).
- ¹⁵J. C. Rochet and P. T. Lansbury, *Curr. Opin. Struct. Biol.* **10**, 60 (2000).
- ¹⁶U. B. Sleytr, P. Messner, D. Pum, and M. Sára, *Angew. Chem., Int. Ed.* **38**, 1034 (1999).
- ¹⁷S. Chung, S.-H. Shin, C. R. Bertozzi, and J. J. De Yoreo, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16536 (2010).
- ¹⁸S. Whitelam, *Phys. Rev. Lett.* **105**, 088102 (2010).
- ¹⁹C. Horejs, D. Pum, U. B. Sleytr, and R. Tscheliessnig, *J. Chem. Phys.* **128**, 065106 (2008).
- ²⁰C. Horejs, D. Pum, U. B. Sleytr, H. Peterlik, A. Jungbauer, and R. Tscheliessnig, *J. Chem. Phys.* **133**, 175102 (2010).
- ²¹B. Kuen, A. Koch, E. Asenbauer, M. Sára, and W. Lubitz, *J. Bacteriol.* **179**, 1664 (1997).
- ²²V. Tozzini, *Curr. Opin. Struct. Biol.* **15**, 144 (2005).
- ²³J. K. Cheung and T. M. Truskett, *Biophysical J.* **89**, 2372 (2005).
- ²⁴A. P. Heath, L. E. Kaviraki, and C. Clementi, *Proteins* **68**, 646 (2007).
- ²⁵*Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, Berlin, 1986).
- ²⁶M. P. Allen and D. J. Tildesley, *Computer Simulation Of Liquids* (Clarendon, Oxford, UK, 1987).
- ²⁷*Computer Simulations of Surfaces and Interfaces*, edited by B. Dünweg, D. P. Landau, and A. I. Milchev (Springer, Berlin, 2002).
- ²⁸A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten, *J. Phys. Chem. B* **110**, 3674 (2006).
- ²⁹J. Zhang and M. Muthukumar, *J. Chem. Phys.* **130**, 035102 (2009).
- ³⁰S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.* **256**, 623 (1996).
- ³¹J. Skolnick, L. Jaroszewski, A. Kolinski, and A. Godzik, *Protein Sci.* **6**, 676 (1997).
- ³²P. D. Thomas and K. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11628 (1996).
- ³³T. E. Creighton, *Proteins: Structures and Molecular Properties* (W. H. Freeman, San Francisco, 1992).
- ³⁴N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ³⁵D. Pum, M. Weinhandl, C. Hödl, and U. B. Sleytr, *J. Bacteriol.* **175**, 2762 (1993).

- ³⁶J. L. Toca-Herrera, R. Krastev, V. Bosip, S. Küpcü, D. Pum, A. Fery, M. Sára, and U. B. Sleytr, *Small* **1**, 339 (2005).
- ³⁷P. Messner, D. Pum, and U. B. Sleytr, *J. Ultrastruct. Mol. Res.* **97**, 73 (1986).
- ³⁸D. Rünzler, C. Huber, D. Moll, G. Köhler, and M. Sára, *J. Biol. Chem.* **279**, 5207 (2004).
- ³⁹C. Tanford, *Science* **200**, 1012 (1978).
- ⁴⁰D. Chandler, *Nature (London)* **437**, 640 (2005).
- ⁴¹E. E. Meyer, K. J. Rosenberg, and J. Israelachvili, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15739 (2006).
- ⁴²M. K. Mitra and M. Muthukumar, *J. Chem. Phys.* **134**, 044901 (2011).
- ⁴³S. Sotiropoulou, S. S. Mark, E. R. Angert, and C. A. Batt, *J. Phys. Chem. C* **111**, 13232 (2007).
- ⁴⁴D. Moll, C. Huber, B. Schlegel, D. Pum, U. B. Sleytr, and M. Sára, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14646 (2002).
- ⁴⁵H. Tschiggerl, J. L. Casey, K. Parisi, M. Foley, and U. B. Sleytr, *Bioconjugate Chem.* **19**, 860 (2008).
- ⁴⁶M. Sára and U. B. Sleytr, *J. Bacteriol.* **169**, 2804 (1987).
- ⁴⁷D. Pum, M. Sára, and U. B. Sleytr, *J. Bacteriol.* **171**, 5296 (1989).
- ⁴⁸M. Sára, D. Pum, and U. B. Sleytr, *J. Bacteriol.* **174**, 3487 (1992).
- ⁴⁹C. Mader, S. Küpcü, M. Sára, and U. B. Sleytr, *Biochim. Biophys. Acta* **1418**, 106 (1999).
- ⁵⁰S. Howorka, M. Sára, Y. Wang, B. Kuen, U. B. Sleytr, W. Lubitz, and H. Bayley, *J. Biol. Chem.* **275**, 37876 (2000).
- ⁵¹H. Kinns and S. Howorka, *J. Mol. Biol.* **377**, 589 (2008).
- ⁵²T. O. Yeates and J. E. Padilla, *Curr. Opin. Struct. Biol.* **12**, 464 (2002).
- ⁵³R. F. Service, *Science* **298**, 2322 (2002).

SINGLE-MOLECULE FORCE SPECTROSCOPY REVEALS THE INDIVIDUAL MECHANICAL UNFOLDING PATHWAYS OF A SURFACE LAYER PROTEIN

Christine Horejs¹, Robin Ristl¹, Rupert Tscheliessnig², Uwe B. Sleytr¹, Dietmar Pum¹

From Department for Nanobiotechnology¹, and the Austrian Centre of Industrial Biotechnology c/o Institute for Biotechnology²,

University of Natural Resources and Life Sciences, 1190 Vienna, Austria

Running head: Single-molecule force spectroscopy of S-layer proteins

Address correspondence to: Dietmar Pum, Muthgasse 11, 1190 Vienna, Austria

Phone: +431476542205, Fax: +4314789112; E-mail: dietmar.pum@boku.ac.at

Surface layers (S-layers) represent an almost universal feature of archaeal cell envelopes and are probably the most abundant bacterial cell proteins. S-layers are monomolecular crystalline structures of single protein or glycoprotein monomers, which completely cover the cell surface during all stages of the cell growth cycle thereby performing their intrinsic function under a constant intra- and intermolecular mechanical stress. In gram positive bacteria the individual S-layer proteins are anchored by a specific binding mechanism to polysaccharides (secondary cell wall polymers) that are linked to the underlying peptidoglycan layer. In this work AFM-based single-molecule force spectroscopy and a polyprotein approach are used to study the individual mechanical unfolding pathways of an S-layer protein. We uncover complex unfolding pathways involving the consecutive unfolding of structural intermediates, where a mechanical stability of 87 pN is revealed. Interestingly, a change of the unfolding pathway is observed when individual S-layer proteins interact with secondary cell wall polymers, which is a direct signature of a conformational change induced by the ligand. Moreover the mechanical stability increases up to 110 pN. This work demonstrates that single-molecule force spectroscopy offers a powerful tool to detect subtle changes in the structure of an individual protein upon binding of a ligand and constitutes the first study of surface layer proteins at the single-molecule level.

Crystalline bacterial cell surface layers, referred to as S-layers (1), have now been identified in hundreds of different species from nearly every taxonomical group of walled bacteria and represent an almost universal feature in archaea (2). S-layers, which are formed by a highly specific and robust self-assembly process of S-

layer proteins, constitute the simplest type of protein membrane developed during evolution. S-layer proteins self-assemble into monomolecular lattices with different symmetries ranging from p1 to p6 thereby exhibiting pores of identical size and morphology. As these crystalline layers represent the outermost barrier between the living cell and its environment, they are supposed to play a crucial role regarding stabilization and protection of the cell (3), which implies a certain resistance against physiologically relevant forces (4-6). The remarkable feature of S-layer proteins to self-assemble also *in vitro* has arisen the interest of nanotechnologists to use them as a matrix for molecular construction kits following a bottom-up approach (7). Although considerable knowledge has been accumulated on the genetics, the self-assembly behavior, the structure-function relationship, and the biochemical as well as biophysical properties of S-layer proteins (7-18), an overall tertiary structure of one native unmodified protein remains elusive. Classical experimental structure determination methods such as NMR and X-ray crystallography pose problems due to the size and two-dimensional crystallization behavior of S-layer proteins, as in solution they form crystalline monomolecular layers rather than isotropic three-dimensional crystals. The dissolved proteins immediately interact in order to form small oligomers, which provide the nucleation seed for the formation of large layers (19). Recently, a structural model of an S-layer protein, namely the protein SbsB of *Geobacillus stearothermophilus* pV72/p2 (20) could be determined by means of molecular dynamics simulations and small angle X-ray scattering (21,22). This S-layer protein is made up of 920 amino acids, has a molecular weight of 98 kDa and forms two-dimensional layers exhibiting p1 lattice symmetry. SbsB possesses three N-

terminal S-layer homologous (SLH) domains (23), which are of α -helical character, and fibronectin type III as well as Ig-like domains, which are composed of β -sheets. The SLH domains are located at the N-terminus of the protein within the first 208 amino acids (24). This structural architecture is very common for proteins located on the surfaces of cells, which have been successfully investigated using AFM-based single-molecule force spectroscopy (25-27). The SLH domains enable the binding of non-classical secondary cell wall polymers (SCWPs) in order to anchor the individual proteins to the underlying cell wall (24,28-30). The SCWP of *Geobacillus stearothermophilus* pV72/p2 was shown to consist of substoichiometrically substituted tetrasaccharide repeats of the following structure: $[(\beta\text{-D-GlcpNAc})_{-0.3}\text{-}4)\text{-}\beta\text{-D-ManpNAc}\text{-}(1\text{-}3)\text{-}\beta\text{-D-GlcpN}(\text{Ac})_{-0.3}\text{-}[(\text{S})\text{-Pyr-}4,6\text{-}\alpha\text{-D-ManpNAc}\text{-}(1\text{-}6)\text{-}\alpha\text{-D-GlcpNAc}\text{-}(1\text{-})_n]$ (29). The specific binding between S-layer proteins and SCWPs has been intensively studied using molecular biological and biochemical methods, transmission electron microscopy, and surface plasmon resonance spectroscopy (28,31). However, the consequences of ligand binding regarding the conformation of individual S-layer proteins as well as the exact binding sites and mechanisms remained in the dark. The morphogenesis of an S-layer protein is a very complex process in which the amount of the protein component, its translocation through the cell wall, and its specific incorporation into the existing S-layer lattice have to be coordinated with the growth rate of the organism and the synthesis of other cell wall components. Kinetic studies have shown that to keep the surface of a bacterium with a generation time of 20 min completely covered with a closed protein lattice, approximately 500 protein subunits per second must be synthesized and incorporated into the lattice (32). In order to constantly provide such a high number of monomers, self-assembly-disabled and highly mobile proteins are permanently stored in the peptidoglycan layer of the cell wall. The fact that after removing the cytoplasmic membrane, these premature S-layer proteins are capable to form an additional closed S-layer on the inner surface of the peptidoglycan layer, proves that the binding of SCWP somehow induces the large-scale self-assembly

(33). The exact secretory pathway of individual S-layer proteins is not well understood yet, but the SCWPs, which account for 50% of the rigid cell wall, are supposed to be key players in guiding the proteins on their way to the cell surface and subsequently enabling a defined orientation of the lattice.

AFM-based single-molecule force spectroscopy offers an opportunity to study the mechanical unfolding pathways of individual proteins thereby elucidating the mechanical stability and possible transition states (25,34-41). S-layer proteins perform their intrinsic function under constant mechanical stress due to the fact that they completely cover whole cells during all stages of the cell growth and cell division (3,42). In order to better understand the structure-function relationship and the role SCWPs play during the transport of S-layer proteins to the cell surface, it is of great importance to study the mechanical unfolding of individual proteins under the influence of a pulling force. Single-molecule force spectroscopy was already successfully used to investigate the influence of a ligand on the mechanical unfolding of various proteins. It was reported that due to ligand binding the mechanical stability of proteins often increases even when the binding event takes place at a position far from those regions that are critical for mechanical stability. However, the mechanical unfolding pathway remains the same (43-49). On the contrary, mechano-sensitive proteins show a conformational adaption to mechanical stress. Mechanical force can induce a conformation that is able to bind certain substrates, where it has been shown that this structure-function relationship plays an important role *in vivo*. (41,50-54).

Here we report for the first time an apparent structural change of a protein upon ligand binding, where the ligand binds prior to any mechanically induced structural change or stretching of the protein. We conducted mechanical unfolding experiments at a constant velocity using polypeptides, where the S-layer protein is embedded within five I27 modules of titin (35,40,55,56). The mechanical unfolding pathway of the I27 module of the human muscle protein titin has been intensively investigated and serves as a mechanical fingerprint in our experiments. This approach permits the direct detection of unfolding events related to the mechanical unfolding of SbsB and serves as an

internal standard for our set-up. A schematic cartoon of the polyprotein approach is illustrated in Fig. 1A, where the calculated structural model of SbsB is shown (21,22). In order to address the question, of how and if the protein is influenced due to ligand binding, the identical experiments were performed under the presence of secondary cell wall polymers.

Experimental Procedures

Construction, expression and purification of the (I27)₅-SbsB polyproteins – The polyprotein (I27)₂-SbsB-(I27)₃ was constructed by ligating the coding sequence of the S-layer protein SbsB from *Geobacillus stearothermophilus* pV72/p2 into a modified pET3-vector. This vector contains five I27 domains, an N-terminal His-tag (HIS)₆ for the purification of the expressed polyproteins and two C-terminal cysteine residues, which enable the binding of the polyproteins to a gold surface. PCRs and cloning were performed according to standard procedures (57) and *Escherichia coli* DH5 α was used as cloning strain. The template DNA for PCR was chromosomal DNA of *Geobacillus stearothermophilus* pV72/p2. The gene encoding SbsB was amplified using the following primers: 5'-ATCGgcgcgAGCTTCACAGATGTTGCG-3' (forward primer) and 5'-ATCGgcgcgTTTTGTCACAGTCACATTGAC-3' (reverse primer). The PCR product and the modified pET3 vector were digested with BssHII (Invitrogen) and then ligated using T4 DNA ligase (Invitrogen) yielding the vector pSbsB-I27. In the plasmid the coding sequence of the S-layer protein SbsB is flanked by two and three coding sequences of I27 domains, respectively (Fig. 1A). Finally, the plasmid was transformed into *Escherichia coli* BL21 cells by electroporation. Protein expression was performed as described previously (58). The polyproteins were purified by Ni-His affinity chromatography using 1-mL HisTrapTM HP prepacked columns (GE Healthcare) and anion exchange chromatography using Uno Q-1 columns (Bio-Rad), both under non-denaturing conditions. Finally the proteins were dialyzed against MilliQ water containing 1mM DTT yielding a protein concentration of 0.13 mg/mL. The purity of the protein was verified by SDS-PAGE.

Purification and modification of SCWP – The SCWP of *Geobacillus stearothermophilus* pV72/p2 was isolated and pyridyl disulfide-activated as described previously (28,59).

Single-molecule force spectroscopy – Force-extension measurements were conducted using the JPK NanoWizard I atomic force microscope. To study the mechanical unfolding of the polyprotein, 20 μ L of the protein solution (0.13 mg/mL) were deposited on a freshly evaporated gold coverslide, incubated for 30 min and rinsed thoroughly with MilliQ water. For the investigation of the influence of the ligand, 20 μ L of SCWP solution (1mg/mL) were deposited on a freshly evaporated gold coverslide, incubated for 30 min and rinsed with MilliQ water. Subsequently 30 μ L of the protein solution were added and incubated for 60 min. The measurements were performed using MSCT cantilevers (Bruker, AXS), which were individually calibrated using the equipartition theorem (60) resulting in a typical force constant of 16 pN/nm. Individual proteins were picked up from the surface by gently touching the coverslide with the AFM cantilever tip as to promote the nonspecific adhesion of the proteins to the cantilever and extended by retracting the cantilever with a constant pulling velocity of 400 nm/s. The resulting force-extension traces were fitted to the worm like chain (WLC) model of polymer elasticity (61) using a Mathematica script. We used a constant persistence length of 0.4 nm (40).

RESULTS

Stepwise mechanical unfolding of the S-layer protein SbsB. Fig. 2A,B show two typical force-extension curves corresponding to the unfolding of (I27)₂-SbsB-(I27)₃. We observe five consecutive unfolding events corresponding to the full extension of the five I27 domains at \sim 200 pN with a contour length increment of \sim 28 nm. The distribution of forces and increment in contour lengths for the mechanical unfolding of I27 are shown in Fig. 2G,H, which is in good agreement with published values (55) and can, therefore, be exclusively related to the mechanical unfolding of I27. Prior to the unfolding events corresponding to the I27 domains, a double peak is observed. This double peak at 86.8 ± 12.5 pN and 87.6 ± 10.8 pN

respectively, must be related to the unfolding of the S-layer protein SbsB (histograms are shown in Fig. 2D,F, $n=90$). The double peak corresponds to two consecutive unfolding intermediates in the structure of the protein, and once these intermediates are sequentially unfolded, the protein extends by 36.3 ± 5.2 nm (Fig. 2E, ΔL_2). The contour length increment between the two intermediates equals 12.1 ± 4.9 nm (Fig. 2C, ΔL_1). Upon the application of force the protein extends by various amounts without exhibiting features of mechanical stability, where this extension varies from almost zero up to 50 nm and is widely distributed. Based on the structural model of the S-layer protein (21), the complete unfolding of SbsB would lead to a contour length increase of ~ 330 nm, if the contour length per amino acid is taken to be 0.4 nm (62) (889 amino acids $\times 0.4$ nm = 25.2 nm end-to-end distance (35)). The maximum observed total contour length increase is much less than 330 nm, which indicates that the S-layer protein is not completely mechanically unfolded in the framework of the experiment. A lack of additional unfolding events following the peaks related to the unfolding of the five I27 domains strongly supports this hypothesis. Apparently, only certain domains of the protein are unfolded, while the double peak marks the fingerprint of these unfolding events. Such phenomenon has been observed previously for the mechanical unfolding of polyQ chains (63), where no initial extension could be observed. The authors propose that the inextensibility of polyQ chains might occur due to the formation of compact and highly stable conformations that are mechanically resilient. Moreover, a connection between this ensemble of mechanically resilient conformations and the assembly process of polyQ chains is postulated. The lack of an initial extension of S-layer proteins might be similarly explained through the existence of compact and highly stable conformations of S-layer proteins that are inextensible in the framework of the experiment. *Increase of the mechanical stability and an apparent conformational change of SbsB when bound to SCWP.* In Fig. 1B the set-up for this experiment is schematically shown. SCWP is responsible for anchoring the monomers *via* the N-terminal SLH domains to the rigid cell wall. The exact amino acids comprising the three SLH

domains of SbsB are not known yet, but the overall location at the N-terminal region could be revealed by various experimental methods (24,28) and is shown in a magnified view in Fig. 1B. This interaction is specific for *Geobacillus stearothermophilus* pV72/p2 and ensures that the crystalline protein layer is constrained to the cell surface. Fig. 3A,B show two typical force-extension traces corresponding to the unfolding of (I27)₂-SbsB-(I27)₃ in the presence of SCWP molecules. Again, the five consecutive peaks resulting from the unfolding of five I27 domains can be detected. Prior to the unfolding events related to the I27 domains, we observe again a double peak as for the protein monomers. Interestingly, an additional unfolding event appears in the presence of the ligand, which is marked with a green star in Fig. 3A,B. This unfolding event takes place posterior to the double peak and marks an additional intermediate structure of SbsB, which has not been observed for the unbound monomers. The increase in contour length between the first and second peak of the double peak is slightly larger than for the protein without SCWP, $\Delta L_1=17.3 \pm 4.2$ nm (Fig. 3C, ΔL_1). The contour length increment between the double peak and the consecutive new peak equals 44.9 ± 6.5 nm (Fig. 3E, ΔL_2) and the increase in contour length before the first I27 domain unfolds equals 32.8 ± 7.2 nm (Fig. 3G, ΔL_3). The force necessary to unfold the consecutive unfolding intermediates is about 110 pN for the two peaks of the double peak respectively and about 117 pN for the new peak and consequently of about 30% higher than for the unbound protein (Fig. 3D,F,H show the histograms for $n=64$). Hence, the binding of SbsB to its specific secondary cell wall polymer results in an increase in mechanical stability and to a change of the mechanical unfolding pathway revealed by an additional unfolding peak.

DISCUSSION

Our experiments at the single-molecule level demonstrate that the S-layer protein SbsB is not fully extensible under the presence of a pulling force. Apparently not the complete protein can be mechanically unfolded. A double-peak of about 87 pN serves as a mechanical fingerprint for the individual unfolding of certain domains of SbsB revealing a rather complex unfolding

pathway through distinct well-defined intermediates. The fact that only parts of the protein can be mechanically unfolded might be related to the formation of highly stable, compact and mechanically resilient conformations. The same phenomenon has been previously observed for the mechanical unfolding of polyQ chains (63). Nevertheless, the mechanical unfolding of SbsB results in a clear fingerprint related to at least two structurally closely located domains. Similar unfolding patterns have been reported for fibronectin type III modules (25), OspA (64), kinase domains (65), T4 lysozyme (66), and the maltose-binding protein (34). SbsB is expected to consist of fibronectin type III and Ig-like domains - structures of known mechanical stability - which allows for the assumption that the double peak originates from the unfolding of such domains in the protein. Thermally and chemically induced unfolding studies also indicated that SbsB is a multidomain protein exhibiting a complex unfolding behavior with a three-step unfolding having been observed (16). A contour length increase of about 12 nm indicates that the consecutive unfolding intermediates are closely located in the structure of the protein. Due to the lack of a pdb file for the atomistic structure of this S-layer protein, the transition states observed in the unfolding studies cannot be assigned to certain structural conformations. However, a mechanical stability of about 87 pN as well as the unfolding through distinct transition states allows for the hypothesis that SbsB is a multidomain protein consisting of domains exhibiting mechanical resistance. Comparable forces have been found for fibronectin (27,67) and tenascin (26), both extracellular matrix proteins, and again for a fibronectin type III domain (25). Hence, a mechanical stability of about 87 pN proves that S-layer proteins are extensible under the presence of a pulling force. The constant maintenance of a coherent S-layer lattice is important for acting as a barrier and well-defined molecular sieve (3,68). The unfolding pattern of individual S-layer proteins indicate a certain extensibility of S-layers, which plays an important role for their function as highly dynamic closed surface crystals (42,69,70). S-layers are anchored to the cell wall through the binding of species-specific SCWPs, where the binding takes place at the N-terminal region of the protein involving three SLH domains (71).

When individual proteins are mechanically unfolded in the presence of these ligands, an additional unfolding peak appears posterior to the double peak that was observed for the unbound proteins. Thus, we detect a change of the unfolding pathway upon ligand binding, which is a direct signature of a conformational change of the protein. Due to the lack of structural information of SbsB at an atomistic level, we are not able to relate the additional transition state to a specific protein conformation. However, the altered pathway upon ligand binding demonstrates that such conformational changes can be sampled using single-molecule force spectroscopy. S-layer proteins apparently adopt a different structure when bound to their specific ligands. It has been recently found that S-layer proteins adopt different conformations as monomers and when part of self-assembled structures (8,72). Obviously, such a conformational change also takes place when S-layer proteins bind to SCWPs. Recent studies on the mechanoenzymatics of titin kinase showed that mechanical force is able to induce a catalytically competent conformation in order to bind substrates (50). An additional peak was observed caused by the interaction with the ligand (51). Similar studies on mechano-sensitive proteins demonstrate that due to the extension and stretching of a protein, active intermediate conformations become accessible for the binding of specific substrates (41,52,53,73,74). In our experiments S-layer proteins bind SCWPs prior to any mechanically induced unfolding event. Thus, the observed conformational change takes place due to the binding of the ligand rather than due to extending the protein.

This finding allows for the hypothesis that S-layer monomers can be conformationally modified through the binding of SCWPs, which gives new insights into the behavior of S-layer proteins *in vivo*.

The morphogenesis of S-layers involves the translocation of monomers through the peptidoglycan-containing rigid cell wall matrix, the binding to SCWPs and the incorporation in the growing closed S-layer lattice in a defined orientation. The exact secretory pathway of S-layer proteins is not well understood yet. However, the high molecular mass and the complex tertiary structure of S-layer proteins leads one to assume that monomers are

transported through the cell wall in a rather unfolded conformation. Moreover, it has been shown that the peptidoglycan layer contains a pool of mobile S-layer monomers, which do not self-assemble within the peptidoglycan matrix in an intact cell (33). One reason for the proteins not to start crystallizing into layers might be related to their intermediate conformation. Our results provide strong evidence that the binding of SCWPs might induce a fundamental conformational change of the individual proteins, which might constitute an important step on their way to the cell surface.

Various single-molecule experiments have been conducted in order to investigate the influence of a bound ligand on the mechanical stability of different proteins (43-48). Independent of the binding site, it was observed that the mechanical stability of the protein increases in most of the cases, when the ligand is bound. Here we monitor the same phenomenon. The mechanical stability increases from about 87 pN up to about 110 pN, which represents an increase of about 30%. The additional unfolding event marks an even higher mechanical resistance of about 117 pN. Therefore, individual S-layer proteins, when bound to the cell surface, can bear mechanical forces of more than 110 pN without unfolding. This value might be even higher when the proteins are self-assembled due to a further stabilization of the structure. This enhanced mechanical stability arises despite the fact that the cell wall polymers do not directly interact with the key region that is supposed to be important for mechanical stability. The SCWPs bind to the N-terminal region of the protein, which is mainly composed of α -helices (21). α -helical structures are reported to unfold under the presence of much lower forces than observed in our experiments (34,75-77). Therefore, the mechanical stability is likely to arise due to β -

sheet structures, which are located closer to the C-terminal region, where the protein consists of Ig-like and fibronectin type III domains (21). Apparently a long-range coupling, as reported previously (45), can be also observed for S-layer proteins.

Our experiments provide the first demonstration of a change of the mechanical unfolding pathway due to the binding of a specific ligand prior to a force-induced extension of the protein, which serves as a direct indication of a conformational change of the protein. Moreover the mechanical stability of an S-layer protein could be probed at the single-molecule level, where an increase in mechanical stability is observed when the individual proteins are bound to their ligands. The observed conformational change of an S-layer protein together with experimental findings related to the behavior of S-layers within the peptidoglycan layer (33), provide a possible explanation of how the complex translocation of individual proteins through the cell wall might be controlled. S-layer monomers constitute an S-layer protein pool „waiting“ in the peptidoglycan matrix for the upcoming insertion into the S-layer on the surface of the growing and dividing cell. Due to the intermediate conformation of the individual proteins, they are not able to self-assemble within the cell. The binding of SCWPs induces a conformational change enabling the interaction with other S-layer proteins, where this intermolecular interaction leads to a final conformational change yielding the native structure within an S-layer. Finally, our results clearly demonstrate the mechanical unfolding through distinct well-defined intermediates and a mechanical stability comparable with other extracellular proteins, which identifies S-layers as not completely extensible protein layers able to bear the mechanical stress that prokaryotic cells permanently experience in nature.

REFERENCES

1. Sleytr, U. B., Messner, P., Pum, D., and Sára, M. (1988) *Crystalline bacterial cell surface layers*, Springer, Berlin
2. Messner, P., Schäffer, C., Egelseer, E.M., and Sleytr, U.B. (2010) Occurrence, structure, chemistry, genetics, morphogenesis, and functions of S-layers. in *Prokaryotic cell wall compounds - structure and biochemistry* (König, H., Claus, H., Varma, A. ed.), Springer, Berlin, Germany. pp 53-109
3. Sleytr, U. B., and Beveridge, T. J. (1999) *Trends Microbiol* **7**, 253-260

4. Müller, D. J., Baumeister, W., and Engel, A. (1999) *Proc Natl Acad Sci U S A* **96**, 13170-13174
5. Verbelen, C., Antikainen, J., Korhonen, T. K., and Dufrene, Y. F. (2007) *Ultramicroscopy* **107**, 1004-1011
6. Scheuring, S., Stahlberg, H., Chami, M., Houssin, C., Rigaud, J. L., and Engel, A. (2002) *Mol Microbiol* **44**, 675-684
7. Egelseer, E. M., Ilk, N., Pum, D., Messner, P., Schäffer, C., Schuster, B., and Sleytr, U. B. (2009) *S-Layers, Microbial, Biotechnological Applications*, John Wiley & Sons, Inc.
8. Chung, S., Shin, S. H., Bertozzi, C. R., and De Yoreo, J. J. (2010) *Proc Natl Acad Sci U S A* **107**, 16536-16541
9. Fagan, R. P., Albesa-Jove, D., Qazi, O., Svergun, D. I., Brown, K. A., and Fairweather, N. F. (2009) *Mol Microbiol* **71**, 1308-1322
10. Howorka, S., Sára, M., Wang, Y., Kuen, B., Sleytr, U. B., Lubitz, W., and Bayley, H. (2000) *J Biol Chem* **275**, 37876-37886
11. Kinns, H., Badelt-Lichtblau, H., Egelseer, E. M., Sleytr, U. B., and Howorka, S. (2010) *J Mol Biol* **395**, 742-753
12. Kinns, H., and Howorka, S. (2008) *J Mol Biol* **377**, 589-604
13. Norville, J. E., Kelly, D. F., Knight, T. F., Jr., Belcher, A. M., and Walz, T. (2007) *J Struct Biol* **160**, 313-323
14. Pavkov, T., Egelseer, E. M., Tesarz, M., Svergun, D. I., Sleytr, U. B., and Keller, W. (2008) *Structure* **16**, 1226-1237
15. Pavkov, T., Oberer, M., Egelseer, E. M., Sára, M., Sleytr, U. B., and Keller, W. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 1466-1468
16. Rünzler, D., Huber, C., Moll, D., Köhler, G., and Sára, M. (2004) *J Biol Chem* **279**, 5207-5215
17. Whitelam, S. (2010) *Phys Rev Lett* **105**, 088102
18. Horejs, C., Mitra, M. K., Pum, D., Sleytr, U. B., and Muthukumar, M. (2011) *J Chem Phys* **134**, 125103
19. Jaenicke, R., Welsch, R., Sára, M., and Sleytr, U. B. (1985) *Biol Chem Hoppe Seyler* **366**, 663-670
20. Kuen, B., Koch, A., Asenbauer, E., Sára, M., and Lubitz, W. (1997) *J Bacteriol* **179**, 1664-1670
21. Horejs, C., Pum, D., Sleytr, U. B., and Tscheliessnig, R. (2008) *J Chem Phys* **128**, 065106
22. Horejs, C., Pum, D., Sleytr, U. B., Peterlik, H., Jungbauer, A., and Tscheliessnig, R. (2010) *J Chem Phys* **133**, 175102
23. Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S., and Baumeister, W. (1994) *J Bacteriol* **176**, 1224-1233
24. Sára, M., Egelseer, E. M., Dekitsch, C., and Sleytr, U. B. (1998) *J Bacteriol* **180**, 6780-6783
25. Li, L., Huang, H. H., Badilla, C. L., and Fernandez, J. M. (2005) *J Mol Biol* **345**, 817-826
26. Oberhauser, A. F., Marszalek, P. E., Erickson, H. P., and Fernandez, J. M. (1998) *Nature* **393**, 181-185
27. Shaub, A. (1999) *Nat Cell Biol* **1**, E173-175
28. Mader, C., Huber, C., Moll, D., Sleytr, U. B., and Sára, M. (2004) *J Bacteriol* **186**, 1758-1768
29. Petersen, B. O., Sára, M., Mader, C., Mayer, H. F., Sleytr, U. B., Pabst, M., Puchberger, M., Krause, E., Hofinger, A., Duus, J., and Kosma, P. (2008) *Carbohydr Res* **343**, 1346-1358
30. Schäffer, C., and Messner, P. (2005) *Microbiology* **151**, 643-651
31. Sára, M., Dekitsch, C., Mayer, H. F., Egelseer, E. M., and Sleytr, U. B. (1998) *J Bacteriol* **180**, 4146-4153
32. Sleytr, U. B., and Messner, P. (1983) *Annu Rev Microbiol* **37**, 311-339
33. Breitwieser, A., Gruber, K., and Sleytr, U. B. (1992) *J Bacteriol* **174**, 8008-8015
34. Bertz, M., and Rief, M. (2008) *J Mol Biol* **378**, 447-458
35. Dietz, H., and Rief, M. (2004) *Proc Natl Acad Sci U S A* **101**, 16192-16197

36. Garcia-Manyes, S., Brujic, J., Badilla, C. L., and Fernandez, J. M. (2007) *Biophys J* **93**, 2436-2446
37. Garcia-Manyes, S., Dougan, L., Badilla, C. L., Brujic, J., and Fernandez, J. M. (2009) *Proc Natl Acad Sci U S A* **106**, 10534-10539
38. Kellermayer, M. S., Smith, S. B., Granzier, H. L., and Bustamante, C. (1997) *Science* **276**, 1112-1116
39. Perez-Jimenez, R., Garcia-Manyes, S., Ainarapu, S. R., and Fernandez, J. M. (2006) *J Biol Chem* **281**, 40010-40014
40. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997) *Science* **276**, 1109-1112
41. Puchner, E. M., and Gaub, H. E. (2009) *Curr Opin Struct Biol* **19**, 605-614
42. Glaubert, A. M., and Sleytr, U. B. (1975) *J Ultrastruct Res* **50**, 103-116
43. Ainarapu, S. R., Li, L., Badilla, C. L., and Fernandez, J. M. (2005) *Biophys J* **89**, 3337-3344
44. Cao, Y., Balamurali, M. M., Sharma, D., and Li, H. (2007) *Proc Natl Acad Sci U S A* **104**, 15677-15681
45. Cao, Y., Yoo, T., Zhuang, S., and Li, H. (2008) *J Mol Biol* **378**, 1132-1141
46. Hann, E., Kirkpatrick, N., Kleanthous, C., Smith, D. A., Radford, S. E., and Brockwell, D. J. (2007) *Biophys J* **92**, L79-81
47. Junker, J. P., Hell, K., Schlierf, M., Neupert, W., and Rief, M. (2005) *Biophys J* **89**, L46-48
48. Wang, C. C., Tsong, T. Y., Hsu, Y. H., and Marszalek, P. E. (2011) *Biophys J* **100**, 1094-1099
49. Cao, Y., Li, Y. D., and Li, H. (2011) *Biophys J* **100**, 1794-1799
50. Puchner, E. M., Alexandrovich, A., Kho, A. L., Hensen, U., Schafer, L. V., Brandmeier, B., Grater, F., Grubmuller, H., Gaub, H. E., and Gautel, M. (2008) *Proc Natl Acad Sci U S A* **105**, 13385-13390
51. Puchner, E. M., and Gaub, H. E. (2010) *Angew Chem Int Ed Engl* **49**, 1147-1150
52. Baneyx, G., Baugh, L., and Vogel, V. (2001) *Proc Natl Acad Sci U S A* **98**, 14464-14468
53. Vogel, V., and Sheetz, M. P. (2009) *Curr Opin Cell Biol* **21**, 38-46
54. Vogel, V., Thomas, W. E., Craig, D. W., Krammer, A., and Baneyx, G. (2001) *Trends Biotechnol* **19**, 416-423
55. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. (1999) *Proc Natl Acad Sci U S A* **96**, 3694-3699
56. Oberhauser, A. F., Hansma, P. K., Carrion-Vazquez, M., and Fernandez, J. M. (2001) *Proc Natl Acad Sci U S A* **98**, 468-472
57. Sambrook, J., Russel, D.W. (2001) *Molecular cloning: a laboratory manual*, 3 ed., Cold Spring Harbor Laboratory Press, New York
58. Jarosch, M., Egelseer, E. M., Huber, C., Moll, D., Mattanovich, D., Sleytr, U. B., and Sára, M. (2001) *Microbiology* **147**, 1353-1363
59. Ries, W., Hotzy, C., Schocher, I., Sleytr, U. B., and Sára, M. (1997) *J Bacteriol* **179**, 3892-3898
60. Florin, E. L., Rief, M., Lehmann, H., Ludwig, M., Dornmair, C., Moy, V. T., and Gaub, H. E. (1995) *Biosensors and Bioelectronics* **10**, 895-901
61. Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994) *Science* **265**, 1599-1600
62. Carrion-Vazquez, M., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (1999) *Proc Natl Acad Sci U S A* **96**, 11288-11292
63. Dougan, L., Li, J., Badilla, C. L., Berne, B. J., and Fernandez, J. M. (2009) *Proc Natl Acad Sci U S A* **106**, 12605-12610
64. Hertadi, R., Gruswitz, F., Silver, L., Koide, A., Koide, S., Arakawa, H., and Ikai, A. (2003) *J Mol Biol* **333**, 993-1002
65. Greene, D. N., Garcia, T., Sutton, R. B., Gernert, K. M., Benian, G. M., and Oberhauser, A. F. (2008) *Biophys J* **95**, 1360-1370
66. Peng, Q., and Li, H. (2008) *Proc Natl Acad Sci U S A* **105**, 1885-1890
67. Oberhauser, A. F., Badilla-Fernandez, C., Carrion-Vazquez, M., and Fernandez, J. M. (2002) *J Mol Biol* **319**, 433-447

68. Sára, M., and Sleytr, U. B. (1987) *J Bacteriol* **169**, 4092-4098
69. Sleytr, U. B., Sára, M., Pum, D., and Schuster, B. (2005) Crystalline bacterial cell surface layers (S-layers): a versatile self-assembly system. . in *Supramolecular polymers* (Ciferri, A. ed.), CRC Press, Taylor and Francis group, Boca Raton, Florida. pp 583-616
70. Sleytr, U. B., and Messner, P. (1989) Self-assembly of crystalline bacterial cell surface layers (S-layers). in *Electron Microscopy of Subcellular Dynamics* (Plattner, H. ed.), CRC Press, Boca Raton, Florida. pp 13-31
71. Sára, M. (2001) *Trends Microbiol* **9**, 47-49; discussion 49-50
72. Horejs, C., Gollner, H., Pum, D., Sleytr, U. B., Peterlik, H., Jungbauer, A., and Tscheliessnig, R. (2011) *ACS Nano* **5**, 2288-2297
73. del Rio, A., Perez-Jimenez, R., Liu, R., Roca-Cusachs, P., Fernandez, J. M., and Sheetz, M. P. (2009) *Science* **323**, 638-641
74. Johnson, C. P., Tang, H. Y., Carag, C., Speicher, D. W., and Discher, D. E. (2007) *Science* **317**, 663-666
75. Rief, M., Pascual, J., Sáraste, M., and Gaub, H. E. (1999) *J Mol Biol* **286**, 553-561
76. Junker, J. P., Ziegler, F., and Rief, M. (2009) *Science* **323**, 633-637
77. Junker, J. P., and Rief, M. (2009) *Proc Natl Acad Sci U S A* **106**, 14361-14366

FOOTNOTES

* We greatly appreciate the gift of the modified pET3 vector containing the I27 domains from Dr. David Brockwell and we would like to thank José Luis Toca-Herrera for making the contact. We would like to thank Andrea Scheberl for her help to purify the expressed polypeptides and Jacqueline Friedmann and Birgit Kainz for helpful discussions. This work has been supported by the Air Force projects FA9550-09-1-0342, FA9550-10-1-0223, by the Austrian Science Fund Projects P19047-B12 and P21954-B20 to Christina Schäffer and by the Federal Ministry of Economy, Family and Youth (BMWFJ), the Federal Ministry of Traffic, Innovation and Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG. C.H. holds a DOC-ffORTE fellowship of the Austrian Academy of Sciences.

The abbreviations used are: S-layer, surface layer; NMR, nuclear magnetic resonance; SLH, surface layer homologous; AFM, atomic force microscopy; SCWP, secondary cell wall polymer; PCR, polymerase chain reaction; WLC, worm-like chain; pN, pico Newton; pdb, protein data bank; aa, amino acid.

FIGURE LEGENDS

Fig. 1. (A) Schematic representation of the polyprotein (I27)₂-SbsB-(I27)₃. The structural model of the S-layer protein SbsB as calculated by computer simulations (21) and small-angle X-ray scattering (22) is shown in cyan. The exact location of secondary structure elements is not known. The S-layer protein is flanked by five I27 modules of the muscle protein titin. (B) Schematic representation of the interaction between the N-terminal SLH domains (aa1-aa208) and a secondary cell wall polymer, where one repeating unit of the polymer is shown.

Fig. 2. (A) (B) Typical force-extension traces for (I27)₂-SbsB-(I27)₃ showing five consecutive unfolding peaks for the five I27 domains (WLC fit in orange) and one double peak for the S-layer protein SbsB (WLC fit in blue and red respectively). (C) (D) (E) (F) Histograms of the contour lengths increments and unfolding forces of the double peak resulting from the unfolding of SbsB. The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 12.1 \pm 4.9$ nm (C) and between the second peak of the double peak and the first I27 peak $\Delta L_2 = 36.3 \pm 5.2$ nm (E). The mean unfolding force of the first peak is 86.8 ± 12.5 pN (D) and 87.6 ± 10.8 pN (F) respectively (n = 90). (G) (H) Histograms of the contour length increase and the unfolding forces of the I27 domains. The mean contour length increment is 28.14 ± 2.2 nm and the mean unfolding force 197.89 ± 16.2 pN for n=400.

Fig. 3. Mechanical unfolding of the S-layer protein SbsB when bound to its specific SCWP. (A) (B) Typical force-extension traces. The five consecutive peaks (WLC fit in orange) at about 200 pN correspond to the unfolding of the five I27 domains. The three peaks prior to the I27 peaks correspond to SbsB (blue, red and green WLC fit). The additional peak that can be observed is marked with a green star. The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 17.3 \pm 4.2$ nm (C), between the second peak of the double peak and the third peak $\Delta L_2 = 44.9 \pm 6.5$ nm (E) and between the third peak and the first I27 peak $\Delta L_3 = 32.8 \pm 7.2$ nm (G). The unfolding forces are higher than for the unbound protein, namely $F_1 = 110.1 \pm 9.1$ pN (D), $F_2 = 109.9 \pm 9.8$ pN (F) and $F_3 = 117.4 \pm 9.7$ pN (H).

Figure 1

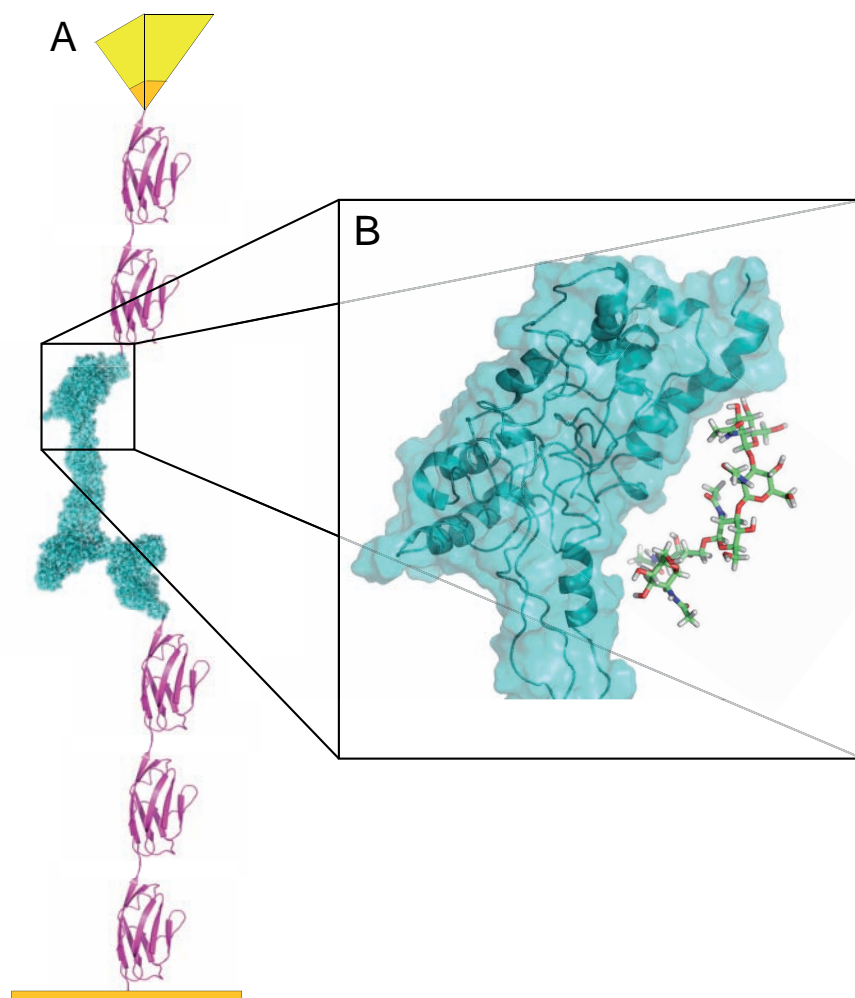


Figure 2

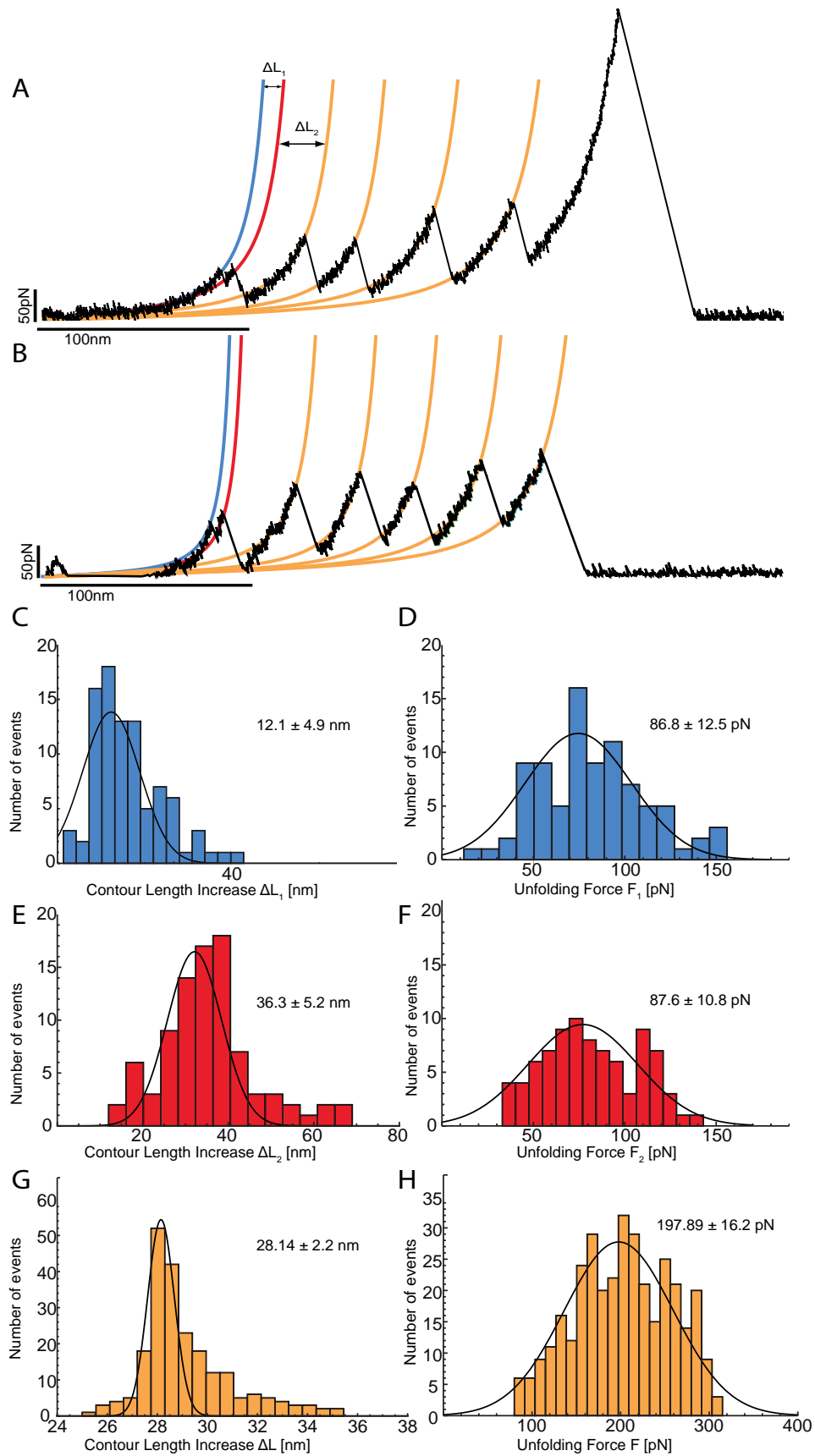
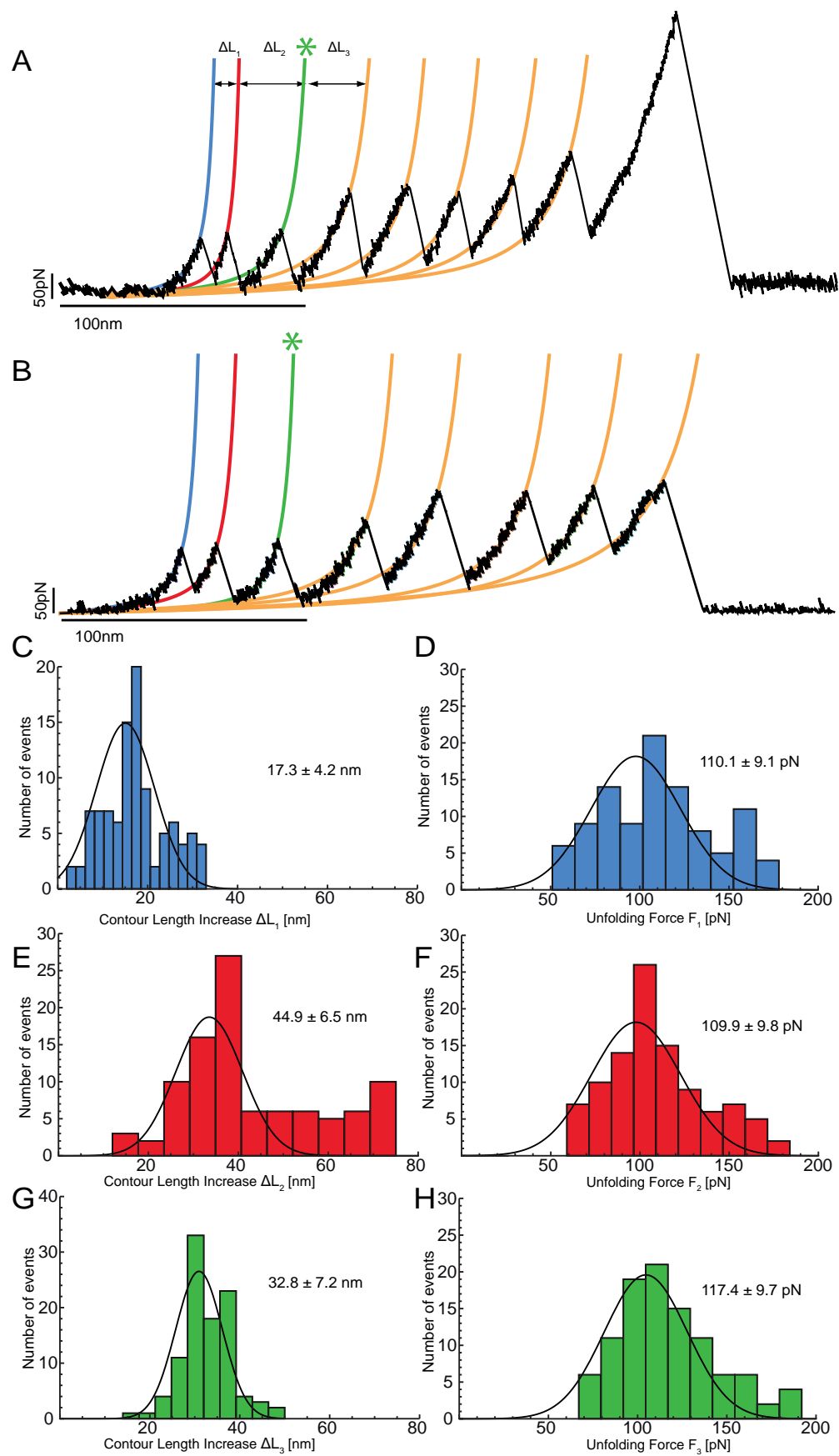


Figure 3



Extract of the book chapter: Nanobiotechnology with S-Layer Proteins as Building Blocks

In: *Progress in Molecular Biology and Translational Science: Molecular Assembly in Natural and Engineered Systems*. (Ed. Howorka, S., Elsevier, Amsterdam (2011), in press)

Uwe B. Sleytr, Bernhard Schuster, Eva M. Egelseer, Dietmar Pum, Christine Horejs, Rupert Tscheliessnig, Nicola Ilk

Department of NanoBiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

Molecular Modelling and Computer Simulations

Over the past decades, a vast amount of studies addressed the three-dimensional (3D) structure determination of S-layer proteins, which led to a considerable knowledge about the distribution of amino acids on S-layer lattices, the structure-function relationship, molecular mechanisms of the self-assembly process and even structural details of some S-layer species.^{80; 82; 83; 173; 174; 175; 176; 177} However, experimental structure determination techniques, e.g. NMR and X-ray crystallography pose problems due to the size and crystallization characteristics of S-layer proteins, as in solution they form crystallized monomolecular layers rather than isotropic three-dimensional crystals. The dissolved proteins immediately interact in order to form small oligomers, which provide the nucleation seed for the formation of large layers.¹⁷⁸ Additionally some S-layer proteins do not fold into their native tertiary structure as monomers in solution but rather condense into amorphous clusters in an extended conformation. Only when assembled into the lattice structure, they restructure into their native conformation.^{173; 179} Thus, 3D reconstructions were limited to truncated or mutated forms of the proteins. The combination of molecular simulations and low resolution experimental techniques, e.g. small angle X-ray scattering (SAXS) and TEM offers an alternative to determine the atomistic structure of unmodified native S-layer proteins and self-assembled lattices. The folding of

small protein domains and of entire proteins can be monitored by reverse and steered molecular dynamics simulations. However, in order to facilitate the equilibration process, secondary structure elements taken from homologous protein models, have to be implemented first. The calculated 3D model of the entire protein can be consequently verified by a systematic exploration of the free energy, by a reverse Monte Carlo simulation based on scattering contrast data obtained by SAXS, or by three-dimensional density distribution data as calculated by TEM. Following this approach, the structural models of the S-layer proteins SbsB from *G. stearotheophilus* pV72/p2 (p1 lattice symmetry)^{81; 178} and of the unit cell of SbpA from *Ly. sphaericus* CCM 2177 (p4 lattice symmetry)¹⁷⁹ could be calculated. Additionally, based on the model of the S-layer protein SbsB, the molecular mechanisms guiding the self-assembly into monomolecular sheets exhibiting a p1 lattice symmetry could be analyzed using Monte Carlo simulations.¹⁸⁰ The structure prediction of the S-layer protein SbsB is shown in Fig. 7. The protein is split into structurally meaningful domains based on homology searches, secondary structure and domain predictions. To obtain three-dimensional coordinates, a pre-modeling by fold recognition is performed. Molecular dynamics simulations were processed with each part and consequently the domains were joined and the whole structure was equilibrated in vacuum (Fig.7 A B C D). The resulting structure was analyzed by pulling parts of the protein along a chosen reaction coordinate and the protein was deformed to quantify the stability, which is expressed as the potential of mean force (Fig. 7F). The final structural model of SbsB is shown in detail in Fig.7 E. The structure could be verified and additionally refined by SAXS studies, where the monomeric structure and self-assemblies were investigated. The analysis is based on a fractal mean potential, which describes best the behavior of S-layers in solution.¹⁷⁸

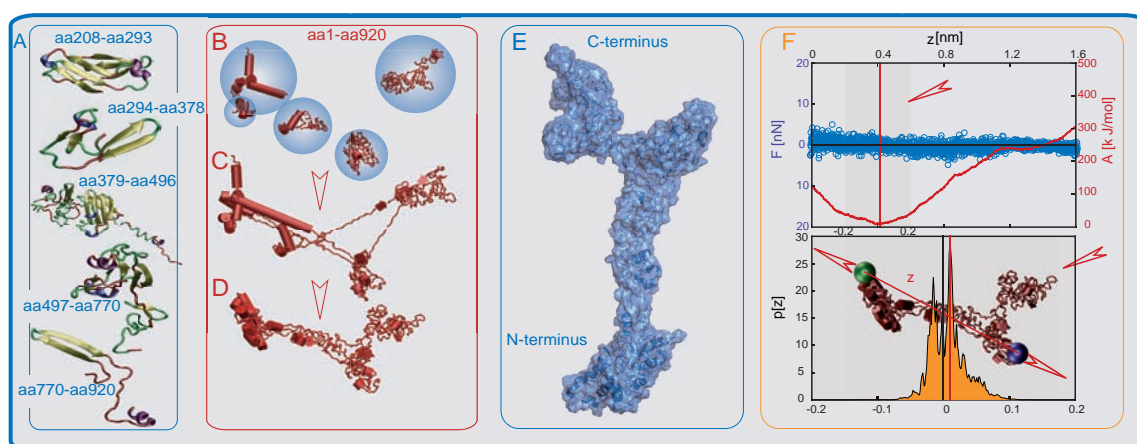


Fig. 7 (A) 3D models of the single domains of the S-layer protein SbsB created by fold recognition. Yellow arrows: beta-sheets, violet strands: alpha-helices, green line: turns, and red line: coils. Modelling method: (B) the individual domains were equilibrated in water spheres at 310K, (C) joined in vacuum, and (D) the final structure was obtained by molecular dynamics simulations. (E) Structural model of SbsB. The protein is L-shaped, where the L is formed by the C-terminal domains. The N-terminus contains the SLH domains and is mainly made up of alpha-helices. (F) Structural analysis of the monomer structure by a calculation of the global free energy. The protein was deformed along the reaction coordinate z . Mean force values F are indicated by open blue circles. The red full line gives the potential of mean force A , which has a clear minimum at $z=0$. Orange body gives the local density probability distribution $p[z]$. The model of the protein is given as an insert, the reaction coordinate is indicated, and the green and blue spheres indicate fixed regions. Figure modified after⁸¹ with friendly permission of the American Institute of Physics.

The reconstruction of the 3D structure of an SbpA unit cell is based on a similar approach. An intermediate structural model was calculated by fold recognition and molecular dynamics simulations. The resulting 3D model of an SbpA tetramer is shown in Fig.8 A,B,C. In this case, three-dimensional density data facilitated the modeling process which were obtained by tilting studies and inverse Fourier transform using transmission electron microscopy. Regions of high and low electron density contrast were identified by SAXS studies (Fig.8 D), where those with high

contrast were classified as non-interacting and those with low contrast as interacting with the other monomers in the tetramer. The resulting 3D model of an SbpA tetramer is shown in Fig.8.

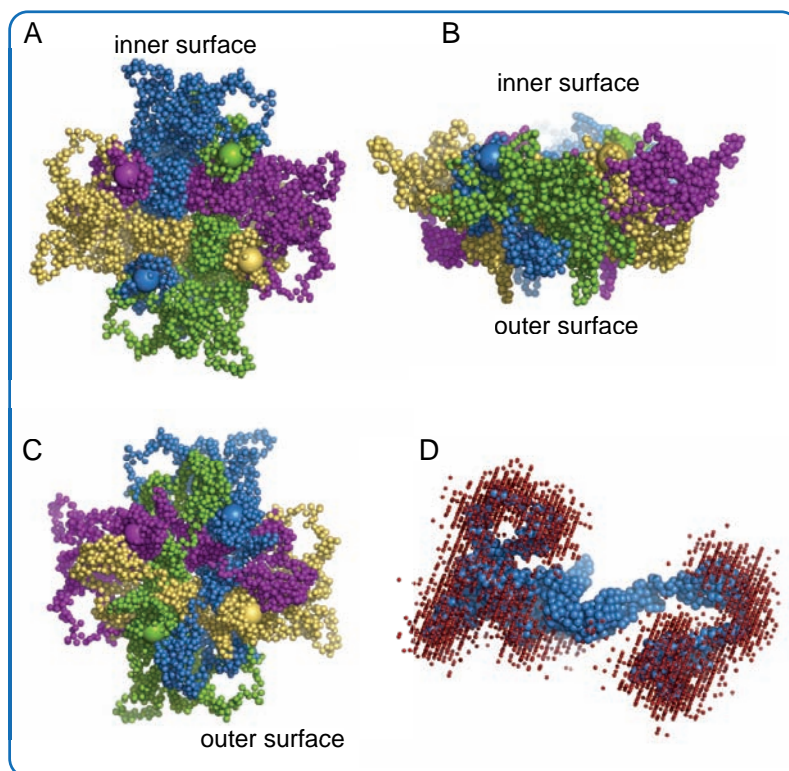


Fig. 8 Three-dimensional structure of the SbpA unit cell. Every monomer in the tetramer is illustrated in a different colour. The proteins are interlocked into each other. (A) Inner surface of the tetramer, which anchors the protein on the cell surface. The N-termini are represented by magnified beads and are accessible on the surface. (B) Side view of the unit cell. (C) Outer surface of the tetramer, which is exposed to the surroundings of the cell. The C-termini are also accessible and marked as magnified beads. (D) Scattering clusters (red beads) of one SbpA monomer as determined by SAXS and a Monte Carlo algorithm. The scattering clusters represent regions of high electron density contrast, where those domains in the protein, which do not show high contrast, are related to interacting or overlapping parts in the tetramer.

The structural model of the S-layer protein SbsB in combination with Monte Carlo simulations was used to study the functional protein self-assembly into S-layers¹⁸⁰.

Using a coarse-grained model, the specific interactions between two protein monomers in solution were investigated to determine the ground-state conformations, which lead to a p1 lattice symmetry. Consequently, the calculated energies of the interactions between two proteins were used to study the large-scale self-assembly by means of lattice Monte Carlo simulations as schematically shown in Fig. 9. Only very few and mainly hydrophobic amino acids, located on the surface of the monomer, are responsible for the formation of the highly anisotropic protein lattice, which is in excellent agreement with known experimental results.

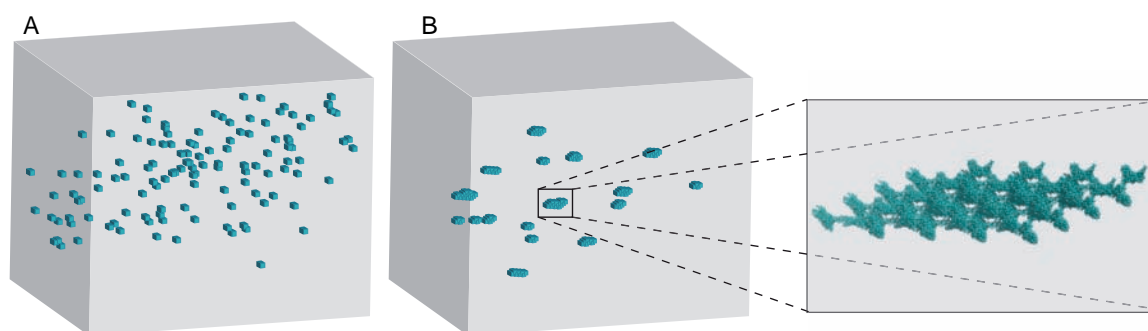


Fig. 9 Representation of the lattice Monte Carlo simulations of the large-scale self-assembly of the S-layer protein SbsB. Energy values were taken from Monte Carlo simulations of the interaction of two monomers. Proteins are represented as cubes. (A) Initial configuration: cubes are randomly distributed in the simulation box. (B) Competitive growth study. Multiple sheets start to grow during initial period. Magnified view of the corresponding S-layer sheet using a coarse-grained model. Figure modified after¹⁸⁰ with friendly permission of the American Institute of Physics.

80. Pavkov, T., Egelseer, E. M., Tesarz, M., Svergun, D. I., Sleytr, U. B. & Keller, W. (2008). The structure and binding behavior of the bacterial cell surface layer protein SbsC. *Structure* **16**, 1226-37.
81. Horejs, C., Pum, D., Sleytr, U. B. & Tscheliessnig, R. (2008). Structure prediction of an S-layer protein by the mean force method. *J Chem Phys* **128**, 065106 (1-11).
82. Kinns, H., Badelt-Lichtblau, H., Egelseer, E. M., Sleytr, U. B. & Howorka, S. (2010). Identifying assembly-inhibiting and assembly-tolerant sites in the SbsB S-layer protein from *Geobacillus stearothermophilus*. *J Mol Biol* **395**, 742-53.
83. Howorka, S., Sára, M., Wang, Y., Kuen, B., Sleytr, U. B., Lubitz, W. & Bayley, H. (2000). Surface-accessible residues in the monomeric and assembled forms of a bacterial surface layer protein. *J Biol Chem* **275**, 37876-86.

173. Chung, S., Shin, S. H., Bertozzi, C. R. & De Yoreo, J. J. (2010). Self-catalyzed growth of S layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc Natl Acad Sci U S A* **107**, 16536-41.
174. Fagan, R. P., Albesa-Jove, D., Qazi, O., Svergun, D. I., Brown, K. A. & Fairweather, N. F. (2009). Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol Microbiol* **71**, 1308-22.
175. Norville, J. E., Kelly, D. F., Knight, T. F., Jr., Belcher, A. M. & Walz, T. (2007). 7Å projection map of the S-layer protein sbpA obtained with trehalose-embedded monolayer crystals. *J Struct Biol*.
176. Pavkov, T., Oberer, M., Egelseer, E. M., Sara, M., Sleytr, U. B. & Keller, W. (2003). Crystallization and preliminary structure determination of the C-terminal truncated domain of the S-layer protein SbsC. *Acta Crystallogr D Biol Crystallogr* **59**, 1466-8.
177. Whitelam, S. (2010). Control of pathways and yields of protein crystallization through the interplay of nonspecific and specific attractions. *Phys Rev Lett* **105**, 088102.
178. Horejs, C., Pum, D., Sleytr, U. B., Peterlik, H., Jungbauer, A. & Tscheliessnig, R. (2010). Surface layer protein characterization by small angle x-ray scattering and a fractal mean force concept: from protein structure to nanodisk assemblies. *J Chem Phys* **133**, 175102.
179. Horejs, C., Gollner, H., Pum, D., Sleytr, U. B., Peterlik, H., Jungbauer, A. & Tscheliessnig, R. (2011). Atomistic structure of monomolecular surface layer self-assemblies: toward functionalized nanostructures. *ACS Nano* **5**, 2288-2297.
180. Horejs, C., Mitra, M. K., Pum, D., Sleytr, U. B. & Muthukumar, M. (2011). Monte Carlo study of the molecular mechanisms of surface-layer self-assembly. *J Chem Phys* **134**, 125103-1.

3 Summary and Conclusions

Molecular self-assembly - one of nature's most fascinating strategies to organize matter on a large-scale - can be only understood by studying the correlation between structural details and resulting mesoscopic morphologies. Surface layers demonstrate a unique model self-assembly system in order to understand the crucial structural features leading to highly ordered two-dimensional protein crystals, where their function is only released by the self-assembly process itself. Just as fascinating these proteins are, as complex is the elucidation of structural details and molecular mechanisms due to their intrinsic function itself. The self-assembly into two-dimensional structures as well as the remarkable molecular weight complicate the application of conventional structure determination methods and demands an alternative - probably unconventional - approach. Computer simulations have entered the scientific world some decades ago and have taken over many open questions that could not be solved by means of experimental methods. Additionally, simulations opened up the way for looking at the details of molecular processes with a resolution still much better than any experimental method has achieved so far. Simulation techniques also highly contributed in studying fundamental theoretical questions, especially regarding the nature of protein folding and protein-protein interactions. However, the quality of computer simulations still depends on experimental and theoretical input, where only a combination of both approaches can pave the way for the determination of protein structures from the scratch and the investigation of the molecular mechanisms that underlie protein-protein interactions.

In this work, such a combinational approach has been applied in order to illuminate the three-dimensional structure of a surface layer protein and consequently, investigate the self-assembly process into S-layers. Molecular dynamics simulations of protein folding, small-angle X-ray scattering studies, high-resolution transmission electron microscopy, AFM-based single-molecule force spectroscopy and finally, Monte Carlo simulations together with a coarse-graining procedure were combined, where new theoretical concepts were suggested in order to describe S-layer proteins in solution. This approach yielded three-dimensional structural models of two different S-layer proteins, which could be subsequently used in order to analyze the specific intra- and intermolecular interactions. New insights could be gathered regarding the structural composition and architecture, which helped to explain and better understand experimental findings regarding location, type, and distribution of amino acids in the S-layer lattice. The kinetics and the geometry could be described by an investigation of the small-angle X-ray scattering patterns of S-layers in solution, where different structural conformations of monomers and self-assemblies, respectively were revealed. A fractal geometry as well as a fractal mean potential could be derived in order to describe the interactions between S-layer proteins and self-assembled sheets in solution, where the immediate initiation of the self-assembly process into small oligomeric structures could be also shown. Such a fractal

approach, based on a mean potential assumption, might open up new avenues for the analysis of critical systems in solution. AFM-based single-molecule force spectroscopy measurements revealed the individual unfolding pathways of one single S-layer protein. A complex unfolding behavior through distinct well-defined intermediates uncovering a multi-domain structure could be observed with a high mechanical stability comparable to other extracellular proteins. The same experimental technique was used in order to study the interaction between an S-layer protein and its specific binding partner in the cell wall - the secondary cell wall polymer. Interestingly, a change of the unfolding pathway indicating an altered conformation of the protein when bound to the cell wall polymer was detected. The observed conformational adaption gave new insights into the complex translocation of S-layer proteins through the rigid cell wall, where apparently, a restructuring or refolding process due to the binding of secondary cell wall polymers on the one hand, and the interaction with neighboring S-layer proteins on the other hand, guides the way of an S-layer protein from an intermediate monomeric state to the incorporation into a self-assembled S-layer on the cell surface. Finally, Monte Carlo simulations using a coarse-grained protein model and a modified statistical potential were applied in order to investigate the large-scale self-assembly of S-layer proteins. Mainly hydrophobic interactions and only few amino acids enable the specific interactions leading to a defined lattice structure, where various experimental studies could be explained by the simulated morphology of the S-layer sheet.

This work constitutes the first theoretical study of S-layer proteins by means of computer simulations at an atomistic level and provides fundamental new insights regarding the structural architecture and the specific interactions of S-layer proteins. Together with the long research history, the results of this work constitute an important step on the way to better understand surface layers, and more generally, the self-assembly strategies of biomolecules and eventually, the successful implementation of computer simulations for the study of large protein systems.

Bibliography

- [1] G. M. Whitesides and B. Grzybowski. Self-assembly at all scales. *Science*, 295:2418–2421, 2002.
- [2] G. M. Whitesides and M. Boncheva. Beyond molecules: self-assembly of mesoscopic and macroscopic components. *Proc. Natl. Acad. Sci. U.S.A.*, 99:4769–4774, 2002.
- [3] Jacob N. Israelachvili. *Intermolecular and surface forces*. Academic Press, London, 1991.
- [4] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, and W. M. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.
- [5] R. Lipowsky. The conformation of membranes. *Nature*, 349:475–481, 1991.
- [6] E. Sackmann. *Biological membranes architecture and function*. Handbook of Biological Physics: Structure and dynamics of membranes - from cells to vesicles. Elsevier, 1995.
- [7] D.G. Angelescu and P. Linse. Viruses as supramolecular self-assemblies: modelling of capsid formation and genome packaging. *Soft Matter*, 4:1981–1990, 2008.
- [8] M. Muthukumar, C. K. Ober, and E. L. Thomas. Competing interactions and levels of ordering in self-organizing polymeric materials. *Science*, 277:1225–1232, 1997.
- [9] J. T. Chen, E. L. Thomas, C. K. Ober, and G. Mao. Self-assembled smectic phases in rod-coil block copolymers. *Science*, 273:343–346, 1996.
- [10] D. K. Schwartz. Mechanisms and kinetics of self-assembled monolayer formation. *Annu. Rev. Phys. Chem.*, 52:107–137, 2001.
- [11] U. B. Sleytr. Regular arrays of macromolecules on bacterial cell walls: structure, chemistry, assembly, and function. *Int. Rev. Cytol.*, 53:1–62, 1978.
- [12] U. B. Sleytr. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. *Nature*, 257:400–402, 1975.
- [13] U. B. Sleytr, P. Messner, D. Pum, and M. Sára. Crystalline bacterial cell surface layers (S-layers). *Ang. Chem. Int. Ed.*, 38:1034–1054, 1999.
- [14] M. Sára and U. B. Sleytr. S-layer proteins. *J. Bacteriol.*, 182:859–868, 2000.
- [15] E.M. Egelseer, N. Ilk, D. Pum, P. Messner, C. Schäffer, B. Schuster, and U.B. Sleytr. *S-layers, microbial, biotechnological applications*. Encyclopedia of Industrial Biotechnology. John Wiley and Sons, Inc., 2009.

- [16] M. McCullagh, T. Prytkova, S. Tonzani, N. D. Winter, and G. C. Schatz. Modeling self-assembly processes driven by nonbonded interactions in soft materials. *J. Phys. Chem. B*, 112:10388–10398, 2008.
- [17] U.B. Sleytr, M. Sára, S. Küpcü, and P. Messner. Structural and chemical characterization of S-layers of selected strains of *Bacillus stearothermophilus* and *Desulfotomaculum nigrificans*. *Arch. Microbiol.*, 146:19–24, 1986.
- [18] C. Schäffer and P. Messner. The structure of secondary cell wall polymers: how gram-positive bacteria stick their cell walls together. *Microbiology*, 151:643–651, 2005.
- [19] M. Sára. Conserved anchoring mechanisms between crystalline cell surface S-layer proteins and secondary cell wall polymers in gram-positive bacteria? *Trends Microbiol.*, 9:47–49, 2001.
- [20] S. Mesnage, T. Fontaine, T. Mignot, M. Delepierre, M. Mock, and A. Fouet. Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *EMBO J.*, 19:4473–4484, 2000.
- [21] P. Messner, K. Steiner, K. Zarschler, and C. Schäffer. S-layer nanoglycobiology of bacteria. *Carbohydr. Res.*, 343:1934–1951, 2008.
- [22] U. B. Sleytr. *Crystalline cell surface layers (S-layers)*. Encyclopedia of Microbiology. Elsevier, 2009.
- [23] S. Sotiropoulou, S.S. Mark, E.R. Angert, and C.A. Batt. Nanoporous S-layer protein lattices. A biological ion gate with calcium selectivity. *J. Phys. Chem. C*, 111:13232–13237, 2007.
- [24] P. Messner, D. Pum, and U.B. Sleytr. Characterization of the ultrastructure and the self-assembly of the surface layer of *Bacillus stearothermophilus* strain NRS 2004/3a. *J. Ultrastruct. Mol. Struct. Res.*, 97:73–88, 1986.
- [25] U. B. Sleytr, B. Schuster, EM. Egelseer, D. Pum, C. Horejs, R. Tscheliessnig, and N. Ilk. *Nanobiotechnology with S-layers as building blocks*. Progress in Molecular Biology and Translational Science: Molecular Assembly in Natural and Engineered Systems. Elsevier, 2011.
- [26] S. Howorka, M. Sára, Y. Wang, B. Kuen, U. B. Sleytr, W. Lubitz, and H. Bayley. Surface-accessible residues in the monomeric and assembled forms of a bacterial surface layer protein. *J. Biol. Chem.*, 275:37876–37886, 2000.

- [27] H. Kinns and S. Howorka. The surface location of individual residues in a bacterial S-layer protein. *J. Mol. Biol.*, 377:589–604, 2008.
- [28] H. Kinns, H. Badelt-Lichtblau, E. M. Egelseer, U. B. Sleytr, and S. Howorka. Identifying assembly-inhibiting and assembly-tolerant sites in the SbsB S-layer protein from *Geobacillus stearothermophilus*. *J. Mol. Biol.*, 395:742–753, 2010.
- [29] T. Pavkov, M. Oberer, E. M. Egelseer, M. Sára, U. B. Sleytr, and W. Keller. Crystallization and preliminary structure determination of the C-terminal truncated domain of the S-layer protein SbsC. *Acta Crystallogr. D Biol. Crystallogr.*, 59:1466–1468, 2003.
- [30] T. Pavkov, E. M. Egelseer, M. Tesarz, D. I. Svergun, U. B. Sleytr, and W. Keller. The structure and binding behavior of the bacterial cell surface layer protein SbsC. *Structure*, 16:1226–1237, 2008.
- [31] M. Kroutil, T. Pavkov, R. Birner-Gruenberger, M. Tesarz, U. B. Sleytr, E. M. Egelseer, and W. Keller. Towards the structure of the C-terminal part of the S-layer protein SbsC. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 65:1042–1047, 2009.
- [32] J. E. Norville, D. F. Kelly, T. F. Knight, A. M. Belcher, and T. Walz. 7Å projection map of the S-layer protein SbpA obtained with trehalose-embedded monolayer crystals. *J. Struct. Biol.*, 160:313–323, 2007.
- [33] R. P. Fagan, D. Albesa-Jove, O. Qazi, D. I. Svergun, K. A. Brown, and N. F. Fairweather. Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol. Microbiol.*, 71:1308–1322, 2009.
- [34] M. Sára and U. B. Sleytr. Charge distribution on the S-layer of *Bacillus stearothermophilus* NRS 1536/3c and importance of charged groups for morphogenesis and function. *J. Bacteriol.*, 169:2804–2809, 1987.
- [35] D. Pum, M. Sára, and U. B. Sleytr. Structure, surface charge, and self-assembly of the S-layer lattice from *Bacillus coagulans* E38-66. *J. Bacteriol.*, 171:5296–5303, 1989.
- [36] M. Sára and U. B. Sleytr. Relevance of charged groups for the integrity of the S-layer from *Bacillus coagulans* E38-66 and for molecular interactions. *J. Bacteriol.*, 175:2248–2254, 1993.
- [37] A. E. Lopez, S. Moreno-Flores, D. Pum, U. B. Sleytr, and J. L. Toca-Herrera. Surface dependence of protein nanocrystal formation. *Small*, 6:396–403, 2010.

- [38] S. Moreno-Flores, A. Kasry, H. J. Butt, C. Vavilala, M. Schmittl, D. Pum, U. B. Sleytr, and J. L. Toca-Herrera. From native to non-native two-dimensional protein lattices through underlying hydrophilic/hydrophobic nanoprotusions. *Angew. Chem. Int. Ed. Engl.*, 47:4707–4710, 2008.
- [39] D. Rünzler, C. Huber, D. Moll, G. Kohler, and M. Sára. Biophysical characterization of the entire bacterial surface layer protein SbsB and its two distinct functional domains. *J. Biol. Chem.*, 279:5207–5215, 2004.
- [40] J. L. Toca-Herrera, S. Moreno-Flores, J. Friedmann, D. Pum, and U. B. Sleytr. Chemical and thermal denaturation of crystalline bacterial S-layer proteins: an atomic force microscopy study. *Microsc. Res. Tech.*, 65:226–234, 2004.
- [41] S. Chung, S. H. Shin, C. R. Bertozzi, and J. J. De Yoreo. Self-catalyzed growth of S-layers via an amorphous-to-crystalline transition limited by folding kinetics. *Proc. Natl. Acad. Sci. U.S.A.*, 107:16536–16541, 2010.
- [42] C. Horejs, D. Pum, U. B. Sleytr, H. Peterlik, A. Jungbauer, and R. Tscheliessnig. Surface layer protein characterization by small angle X-ray scattering and a fractal mean force concept: from protein structure to nanodisk assemblies. *J. Chem. Phys.*, 133:175102, 2010.
- [43] C. Horejs, H. Gollner, D. Pum, U. B. Sleytr, H. Peterlik, A. Jungbauer, and R. Tscheliessnig. Atomistic structure of monomolecular surface layer self-assemblies: toward functionalized nanostructures. *ACS Nano*, 5:2288–2297, 2011.
- [44] B. Kuen, A. Koch, E. Asenbauer, M. Sára, and W. Lubitz. Molecular characterization of the *Bacillus stearothermophilus* PV72 S-layer gene sbsB induced by oxidative stress. *J. Bacteriol.*, 179:1664–1670, 1997.
- [45] N. Ilk, C. Völlenkle, E. M. Egelseer, A. Breitwieser, U. B. Sleytr, and M. Sára. Molecular characterization of the S-layer gene, sbpA, of *Bacillus sphaericus* CCM 2177 and production of a functional S-layer fusion protein with the ability to recrystallize in a defined orientation while presenting the fused allergen. *Appl. Environ. Microbiol.*, 68:3251–3260, 2002.
- [46] M. Sára, E. M. Egelseer, C. Dekitsch, and U. B. Sleytr. Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. *J. Bacteriol.*, 180:6780–6783, 1998.
- [47] C. Mader, C. Huber, D. Moll, U. B. Sleytr, and M. Sára. Interaction of the crystalline bacterial cell surface layer protein SbsB and the secondary cell wall polymer of *Geobacillus stearothermophilus* PV72 assessed by real-time surface plasmon resonance biosensor technology. *J. Bacteriol.*, 186:1758–1768, 2004.

- [48] B. O. Petersen, M. Sára, C. Mader, H. F. Mayer, U. B. Sleytr, M. Pabst, M. Puchberger, E. Krause, A. Hofinger, J. Duus, and P. Kosma. Structural characterization of the acid-degraded secondary cell wall polymer of *Geobacillus stearothermophilus* PV72/p2. *Carbohydr. Res.*, 343:1346–1358, 2008.
- [49] D. Moll, C. Huber, B. Schlegel, D. Pum, U. B. Sleytr, and M. Sára. S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl. Acad. Sci. U.S.A.*, 99:14646–14651, 2002.
- [50] N. Ilk, S. Küpcü, G. Moncayo, S. Klimt, R. C. Ecker, R. Hofer-Warbinek, E. M. Egelseer, U. B. Sleytr, and M. Sára. A functional chimaeric S-layer-enhanced green fluorescent protein to follow the uptake of S-layer-coated liposomes into eukaryotic cells. *Biochem. J.*, 379:441–448, 2004.
- [51] H. Badelt-Lichtblau, B. Kainz, C. Völlenklee, E. M. Egelseer, U. B. Sleytr, D. Pum, and N. Ilk. Genetic engineering of the S-layer protein SbpA of *Lysinibacillus sphaericus* CCM 2177 for the generation of functionalized nanoarrays. *Bioconjug. Chem.*, 20:895–903, 2009.
- [52] M. Delcea, R. Krastev, T. Gutberlet, D. Pum, U.B. Sleytr, and J.L. Toca-Herrera. Thermal stability, mechanical properties and water content of bacterial protein layers recrystallized on polyelectrolyte multilayers. *Soft Matter*, 4:1414–1421, 2008.
- [53] C. Huber, J. Liu, E. M. Egelseer, D. Moll, W. Knoll, U. B. Sleytr, and M. Sára. Heterotetramers formed by an S-layer-streptavidin fusion protein and core-streptavidin as a nanoarrayed template for biochip development. *Small*, 2:142–150, 2006.
- [54] C. Völlenklee, S. Weigert, N. Ilk, E. Egelseer, V. Weber, F. Loth, D. Falkenhagen, U. B. Sleytr, and M. Sára. Construction of a functional S-layer fusion protein comprising an immunoglobulin G-binding domain for development of specific adsorbents for extracorporeal blood purification. *Appl. Environ. Microbiol.*, 70:1514–1521, 2004.
- [55] B. Schuster, P. C. Gufler, D. Pum, and U. B. Sleytr. Interplay of phospholipase A2 with S-layer-supported lipid monolayers. *Langmuir*, 19:3393–3397, 2003.
- [56] J. Tang, A. Ebner, H. Badelt-Lichtblau, C. Völlenklee, C. Rankl, B. Kraxberger, M. Leitner, L. Wildling, H. J. Gruber, U. B. Sleytr, N. Ilk, and P. Hinterdorfer. Recognition imaging and highly ordered molecular templating of bacterial S-layer nanoarrays containing affinity-tags. *Nano Lett.*, 8:4312–4319, 2008.
- [57] C.B. Anfinsen, E. Haber, M. Sela, and F.H.Jr. White. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.*, 47:1309–1314, 1961.

- [58] K.A. Dill and H.S. Chan. From Levinthal to pathways to funnels. *Nature Struct. Biol.*, 4:10–19, 1997.
- [59] C. Levinthal. Are there pathways for protein folding? *J. Chim. Phys.*, 65:44–45, 1968.
- [60] R. Zwanzig, A. Szabo, and B. Bagchi. Levinthal’s paradox. *Proc. Natl. Acad. Sci. U.S.A.*, 89:20–22, 1992.
- [61] K.A. Dill. Theory for the folding and stability of globular proteins. *Biochem.*, 24:1501–1509, 1985.
- [62] K.A. Dill, S.B. Ozkan, T.R. Weikl, J.D. Chodera, and V.A. Voelz. The protein folding problem: when will it be solved. *Curr. Opin. Struct. Biol.*, 17:342–346, 2007.
- [63] J.D. Bryngelson and P.G. Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, 84:7524–7528, 1987.
- [64] J.D. Bryngelson and P.G. Wolynes. A simple statistical field theory of heteropolymer collapse with application to protein folding. *Biopolymers*, 30:177–188, 1990.
- [65] F.H. Stillinger and T.A. Weber. Hidden structure in liquids. *Phys. Rev. A*, 25:978–989, 1982.
- [66] H. Frauenfelder, S.G. Sligar, and P.G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.
- [67] S. Tanaka and H.A. Scheraga. Model of protein folding: Incorporation of one-dimensional short-range (Ising) model into a three-dimensional model. *Proc. Natl. Acad. Sci. U.S.A.*, 74:1320–1323, 1977.
- [68] N. Gō and H. Taketomi. Respective roles of short- and long-range interactions in protein folding. *Proc. Natl. Acad. Sci. U.S.A.*, 75:559–563, 1978.
- [69] N. Gō. Protein folding as a stochastic process. *J. Statist. Phys.*, 30:413–423, 1983.
- [70] A. Sali, E. Shakhnovich, and M. Karplus. How does a protein fold? *Nature*, 369:248–251, 1994.
- [71] K.A. Dill, S. Bromberg, K. Yue, K.M. Fiebig, D.P. Yee, P.D. Thomas, and H.S. Chan. Principles of protein folding - a perspective from simple exact models. *Protein Sci.*, 4:561–602, 1995.
- [72] J. King. Deciphering the rules of protein folding. *Chem. Eng. News*, 67:32–54, 1989.

- [73] T. Lazaridis and M. Karplus. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science*, 278:1928–1931, 1997.
- [74] J.N. Onuchic and P.G. Wolynes. Theory of protein folding. *Curr. Opin. Struct. Biol.*, 14:70–75, 2004.
- [75] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, and P.G. Wolynes. Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins Struct. Funct. Genet.*, 21:167–195, 1995.
- [76] T.R. Sosnick and D. Barrick. The folding of single domain proteins - have we reached a consensus? *Curr. Opin. Struct. Biol.*, 21:12–24, 2011.
- [77] S. Garcia-Manyes, L. Dougan, C.L. Badilla, J. Brujić, and J.M. Fernandez. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl. Acad. Sci. U.S.A.*, 106:10534–10539, 2009.
- [78] J. Brujić, R.I. Hermans, K.A. Walther, and J.M. Fernandez. Single-molecule force spectroscopy reveals signatures of glassy dynamics in the energy landscape of ubiquitin. *Nature Phys.*, 2:282–286, 2006.
- [79] M. Karplus and J. Kuryian. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.*, 102:6679–6685, 2005.
- [80] D. Petrey and B. Honig. Protein structure prediction: inroads to biology. *Molec. Cell*, 20:811–819, 2005.
- [81] W.F. Van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D.P. Geerke, A. Glättli, P.H. Hünenberger, M.A. Kastenholtz, C. Oostenbrink, M. Schenk, D. Trzesniak, N.F.A. van der Vegt, and H.B. Yu. Biomolecular modeling: goals, problems, perspectives. *Ang. Chem. Int. Ed.*, 45:4064–4092, 2006.
- [82] P.L. Freddolino, C.B. Harrison, Y. Liu, and K. Schulten. Challenges in protein-folding simulations. *Nat. Phys.*, 6:751–758, 2010.
- [83] Y. Duan and P.A. Kollman. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282:740–744, 1998.
- [84] C. Simmerling, B. Strockbine, and A.E. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.*, 124:11258–11259, 2010.
- [85] C.D. Snow, B. Zagrovic, and V.S. Pande. The Trp cage: Folding kinetics and unfolded state topology via molecular dynamics simulations. *J. Am. Chem. Soc.*, 124:14548–14549, 2002.

- [86] V.S. Pande, I. Baker, J. Chapman, S.P. Elmer, S. Khaliq, S.M. Larson, Y.M. Rhee, M.R. Shirts, C.D. Snow, E.J. Sorin, and B. Zagrovic. Atomistic protein folding simulations on the submilisecond time scale using worldwide distributed computing. *Biopolymers*, 68:91–109, 2003.
- [87] H. Lei, C. Wu, H. Liu, and Y. Duan. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, 104:4925–4930, 2007.
- [88] H. Lei and Y. Duan. Ab initio folding of albumin binding domain from all-atom molecular dynamics simulation. *J. Phys. Chem. B*, 111:5458–5463, 2007.
- [89] V.A. Voelz, G.R. Bowman, K. Beauchamp, and V.S. Pande. Molecular simulation of ab initio protein folding for a milisecond folder NTL9(1-39). *J. Am. Chem. Soc.*, 132:1526–1528, 2010.
- [90] J. Hsin, J. Strumpfer, E.H. Lee, and K. Schulten. Molecular origin of the hierarchical elasticity of titin: simulation, experiment and theory. *Annu. Rev. Biophys.*, 40:187–203, 2011.
- [91] J. Moult. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.*, 15:285–289, 2005.
- [92] W.R. Pearson and M.L. Sierk. The limits of protein sequence comparison? *Curr. Opin. Struct. Biol.*, 15:254 – 260, 2005.
- [93] S. Kaczanowski and P. Zielenkiewicz. Why similar protein sequences encode similar three-dimensional structures? *Theor. Chem. Acc.*, 125:643–650, 2010.
- [94] D. Baker and A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93–96, 2001.
- [95] C. Chothia and A.M. Lesk. The relation between the divergence of sequence and structure in proteins. *EMBO J.*, 5:823–826, 1986.
- [96] L.J. McGuffin. *Protein fold recognition and threading*. Computational structural biology: methods and applications. World Scientific Publishing Co.Pte.Ltd., Singapore, 2008.
- [97] J.U. Bowie, R. Lüthi, and D. Eisenberg. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253:164–170, 1991.
- [98] D.T. Jones, W.R. Taylor, and J.M. Thornton. A new approach to fold recognition. *Nature*, 358:86–89, 1992.

- [99] T. Schlick. *Pursuing Laplace's vision of modern computers*. Mathematical applications to biomolecular structure and dynamics, IMA Volumes in Mathematics and its Applications. Springer, New York, 1996.
- [100] A.R. Leach. *Molecular modelling: principles and applications*. Addison Wesley Longman Limited, Singapore, 2001.
- [101] T. Hansson, C. Oostenbrink, and W.F. van Gunsteren. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.*, 12:190–196, 2002.
- [102] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Clarendon Press, New York, 1989.
- [103] J.W. Ponder and D.A. Case. Force fields for protein simulations. *Adv. Protein Chem.*, 66:27–86, 2003.
- [104] S.A. Adcock and J.A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106:1589–1615, 2006.
- [105] P.K. Weiner and P.A. Kollmann. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.*, 2:287–303, 1981.
- [106] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.
- [107] W.L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1671, 1988.
- [108] L.D. Schuler, X. Daura, and W.F. van Gunsteren. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.*, 22:1205–1218, 2001.
- [109] C. Oostenbrink, A. Villa, A.E. Mark, and W.F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25:1656–1676, 2004.
- [110] P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369:253–287, 1921.
- [111] I.G. Tironi, R. Sperb, P.E. Smith, and W.F. van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.*, 102:5451–5459, 1995.

- [112] D. Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, 1987.
- [113] D. Trzesniak, A.-P.E. Kunz, and W.F. van Gunsteren. A comparison of methods to compute the potential of mean force. *ChemPhysChem*, 8:162–169, 2007.
- [114] T. Rodinger and R. Pomes. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr. Opin. Struc. Biol.*, 15:164–170, 2005.
- [115] W.F. van Gunsteren, X. Daura, and A.E. Mark. Computation of free energy. *Hel. Chim. Acta*, 85:3113–3129, 2002.
- [116] J.G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3:300–313, 1935.
- [117] G.M. Torrie and J.P. Valleau. Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling. *J. Comput. Phys.*, 23:187–199, 1977.
- [118] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *J. Chem. Phys.*, 119:3559–3566, 2003.
- [119] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.
- [120] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master equation approach. *Phys. Rev. Lett.*, 78:2690–2693, 1997.
- [121] J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco, and C. Bustamante. Equilibrium information from nonequilibrium measurements is an experimental test of Jarzynski’s equality. *Science*, 296:1832–1835, 2002.
- [122] W.F. Van Gunsteren, T.C. Beutler, F. Fraternali, P.M. King, Mark A.E., and P.E. Smith. *Computation of free energy in practice: choice of approximations and accuracy limiting factors*. Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications. Escom Science Publishers, Leiden, 1993.
- [123] T. Mulders, P. Kruger, W. Swegat, and J. Schlitter. Free energy as the potential of mean constraint force. *J. Chem. Phys.*, 104:4869–4870, 1996.
- [124] W. K. den Otter and W. J. Briels. The calculation of free-energy differences by constrained molecular-dynamics simulations. *J. Chem. Phys.*, 109:4139–4146, 1998.
- [125] M. Sprik and G. Ciccotti. Free energy from constrained molecular dynamics. *J. Chem. Phys.*, 109:7737–7744, 1998.

- [126] E.A. Carter, G. Ciccotti, J.T. Hynes, and R. Kapral. Constrained reaction coordinates dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156:472–477, 1989.
- [127] J. Hénin and C. Chipot. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.*, 121:2904–2914, 2004.
- [128] E. Darve, D. Rodriguez-Gomez, and A. Pohorille. Adaptive biasing force methods for scalar and vector free energy calculations. *J. Chem. Phys.*, 128:144120–1–13, 2008.
- [129] C. Chipot and J. Hénin. Exploring the free-energy landscape of a short peptide using an average force. *J. Chem. Phys.*, 123:244906–1–6, 2005.
- [130] P.L. Freddolino, S. Park, B. Roux, and K. Schulten. Force field bias in protein folding simulations. *Biophys. J.*, 96:3772–3780, 2009.
- [131] P. Wolynes. Water mediated interactions in protein folding and structure. *Biophys. J.*, 100:5a, 2011.
- [132] P. Langevin. On the theory of Brownian motion. *C. R. Acad. Sci.*, 146:530–533, 1908.
- [133] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1093, 1953.
- [134] V. Tozzini. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.*, 15:144–150, 2005.
- [135] Y. Ueeda, H. Taketomi, and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulation. A three-dimensional lattice model of lysozyme. *Biopolymers*, 17:1531–1548, 1978.
- [136] K.F. Lau and K.A. Dill. A lattice statistical mechanics model for the conformational and sequence spaces of proteins. *Macromolecules*, 22:3986–3997, 1989.
- [137] K. Yue, K.M. Fiebig, P.D. Thomas, H.S. Chan, E.I. Shakhnovich, and K.A. Dill. A test of lattice protein folding algorithms. *Proc. Natl. Acad. Sci. U.S.A.*, 92:325–329, 1995.
- [138] S. Tanaka and H.A. Scheraga. Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules*, 9:945–950, 1976.

- [139] R.L. Jernigan and I. Bahar. Structure-derived potentials and protein simulations. *Curr. Opin. Struct. Biol.*, 6:195–209, 1996.
- [140] P.D. Thomas and K.A. Dill. An iterative method for extracting energy-like quantities from protein structures. *Proc. Natl. Acad. Sci. U.S.A.*, 93:11628–11633, 1996.
- [141] A. Godzik, A. Kolinski, and J. Skolnick. Knowledge-based potentials for protein folding: what can we learn from protein structures? *Structure*, 4:363–366, 1996.
- [142] J. Miyazawa and R.L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, 256:623–644, 1996.
- [143] J. Skolnick, L. Jaroszewski, A. Kolinski, and A. Godzik. Derivation and testing of pair potentials for protein folding: when is the quasichemical approximation correct. *Protein Sci.*, 6:676–688, 1997.
- [144] M.R. Betancourt and D. Thirumalai. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.*, 8:361–369, 2007.
- [145] C. Horejs, M.K. Mitra, D. Pum, U.B. Sleytr, and M. Muthukumar. Monte Carlo study of the molecular mechanisms of surface-layer protein self-assembly. *J. Chem. Phys.*, 134:125103–1–11, 2011.
- [146] P.D. Thomas and K.A. Dill. Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.*, 257:457–469, 1996.
- [147] C. Horejs, D. Pum, U. B. Sleytr, and R. Tscheliessnig. Structure prediction of an S-layer protein by the mean force method. *J. Chem. Phys.*, 128:065106, 2008.
- [148] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26:1781–1802, 2005.
- [149] A. Guinier and G. Fournet. *Small-angle scattering of X-rays*. Wiley, New York, 1955.
- [150] O. Glatter and O. Kratky. *Small angle X-ray scattering*. Academic Press, London, 1982.
- [151] D.I. Svergun and M.H.J. Koch. Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.*, 66:1735–1782, 2003.

- [152] J. Lipfert and S. Doniach. Small-angle X-ray scattering from RNA, proteins and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.*, 36:307–327, 2007.
- [153] M.H.J. Koch, P. Vachette, and D.I. Svergun. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Quart. Rev. Biophys.*, 36:147–227, 2003.
- [154] M.V. Petoukhov and D.I. Svergun. Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr. Op. Struct. Biol.*, 17:562–571, 2007.
- [155] C.D. Putnam, M. Hammel, G.L. Hura, and J.A. Tainer. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quart. Rev. Biophys.*, 40:191–285, 2007.
- [156] P. Debye. Zerstreuung von Röntgenstrahlen. *Ann. Phys.*, 46:809–823, 1915.
- [157] P. Debye and A.M. Bueche. Scattering by an inhomogeneous solid. *J. Appl. Phys.*, 20:518–521, 1949.
- [158] G. Porod. The X-ray small angle scattering of close packed colloid systems. *Kolloid-Z.*, 124:83–114, 1951.
- [159] O. Glatter. The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Cryst.*, 12:166–175, 1979.
- [160] P. Debye. Molecular-weight determination by light scattering. *J. Phys. Chem.*, 51:18–32, 1947.
- [161] J.S. Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. *Adv. Colloid Interface Sci.*, 70:171–210, 1997.
- [162] A. Tardieu, A. Le Verge, M. Malfois, F. Bonneté, S. Finet, M. Riès-Kautt, and L. Belloni. Proteins in solution: from X-ray scattering intensities to interaction potentials. *J. Cryst. Growth*, 196:193–203, 1999.
- [163] J.P. Hansen and I.R. McDonald. *Theory of simple liquids*. Academic Press, Elsevier, 2006.
- [164] G. Nägele. On the dynamics and structure of charge-stabilized suspensions. *Phys. Rep.*, 272:215–372, 1996.
- [165] L.S. Ornstein and F. Zernike. Accidental deviations of density and opalescence at the critical point of a single substance. *Proc. Acad. Sci. Amsterdam*, 17:793, 1914.

- [166] D.A. McQuarrie. *Statistical mechanics*. Harper Collins Publisher, 1976.
- [167] J.L. Lebowitz and J.K. Percus. Mean spherical model for lattice gases with extended hard cores and continuum fluids. *Phys. Rev.*, 144:251–258, 1966.
- [168] W.G. Madden and S.A. Rice. The mean spherical approximation and effective pair potentials in liquids. *J. Chem. Phys.*, 72:4208–4215, 1980.
- [169] J.K. Percus and G.J. Yevivk. Analysis of classical statistical mechanics by means of collective coordinates. *Phys. Rev.*, 110:1–13, 1958.
- [170] F.J. Rogers and D.A. Young. New, thermodynamically consistent, integral equation for simple fluids. *Phys. Rev. A*, 30:999–1007, 1984.
- [171] G. Zerah and J.-P. Hansen. Self-consistent integral equations for fluid pair distribution functions: another attempt. *J. Chem. Phys.*, 84:2336–2343, 1985.
- [172] J. Chihara. Integral equations for fluids with long-range and short-range potentials. *Progr. Theor. Phys.*, 50:409–423, 1973.
- [173] H.B. Stuhrmann. Interpretation of small-angle scattering functions of dilute solutions and gases. A representation of the structures related to a one-particle scattering function. *Acta Cryst. A*, 26:297–306, 1970.
- [174] D.I. Svergun. Mathematical methods in small-angle X-ray scattering. *J. Appl. Cryst.*, 24:485–492, 1991.
- [175] D.I. Svergun. Restoring three-dimensional structure of biopolymers from solution scattering. *J. Appl. Cryst.*, 30:792–797, 1997.
- [176] D.I. Svergun, V.V. Volkov, M.B. Kozin, H.B. Stuhrmann, C. Barberato, and M.H.J. Koch. Shape determination from solution scattering of biopolymers. *J. Appl. Cryst.*, 30:798–802, 1997.
- [177] P. Chacón, F. Morán, J.F. Díaz, E. Pantos, and J.M. Andreu. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.*, 74:2760–2775, 1998.
- [178] D.I. Svergun. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.*, 76:2879–2886, 1999.
- [179] D.I. Svergun, M.V. Petoukhov, and M.H.J. Koch. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.*, 80:2946–2953, 2001.
- [180] B.B. Mandelbrot. *Fractals, Form and Dimension*. Freeman, San Francisco, 1977.

- [181] B.B. Mandelbrot. *The fractal geometry of nature*. Freeman, San Francisco, 1982.
- [182] A. Coniglio, L. De Arcangelis, and H.J. Herrmann. Fractals and multifractals: applications in physics. *Physica A*, 157:21–30, 1989.
- [183] D. Avnir, O. Biham, D. Lidar, and O. Malcai. Is the geometry of nature fractal? *Science*, 279:39–40, 1998.
- [184] M. Sahimi. Flow phenomena in rocks: from continuum models to fractals, percolation, cellular automata, and simulated annealing. *Rev. Mod. Phys.*, 65:1393–1534, 1993.
- [185] M. Daniel, S. Baskar, and M.M. Latha. Fractal dimension and tertiary structure of proteins. *Phys. Scr.*, 60:270–276, 1999.
- [186] H.J. Stapleton, J.P. Allen, C.P. Flynn, D.G. Stinson, and S.R. Kurtz. Fractal form of proteins. *Phys. Rev. Lett.*, 45:1456–1459, 1980.
- [187] J.S. Helman, A. Coniglio, and C. Tsallis. Fractons and fractal structure of proteins. *Phys. Rev. Lett.*, 53:1195–1197, 1984.
- [188] R. Elber and M. Karplus. Low-frequency modes in proteins: use of effective-medium approximation to interpret the fractal dimension observed in electron-spin relaxation measurements. *Phys. Rev. Lett.*, 56:394–397, 1986.
- [189] A.C. Castellano, M. Barteri, A. Bianconi, E. Borghi, L. Cassiano, M. Castagnola, S. Della Longa, and A. La Monaca. X-ray small angle scattering of the human transferrin protein aggregates. *Biophys. J.*, 64:520–524, 1993.
- [190] T.A. Witten and L.M. Sander. Diffusion-limited aggregation. *Phys. Rev. B*, 27:5686–5697, 1983.
- [191] J.E. Martin and A.J. Hurd. Scattering from fractals. *J. Appl. Cryst.*, 20:61–87, 1987.
- [192] T. Freltoft and J.K. Kjems. Power-law correlations and finite-size effects in silica particle aggregates studied by small-angle neutron scattering. *Phys. Rev. B*, 33:269–275, 1986.
- [193] S.-H. Chen and J. Teixeira. Structure and fractal dimension of protein-detergent complexes. *Phys. Rev. Lett.*, 57:2583–2586, 1986.
- [194] J. Teixeira. Small-angle scattering by fractal systems. *J. Appl. Cryst.*, 21:781–785, 1988.
- [195] R. Blender and W. Dieterich. Debye-Hückel theory on random fractals. *J. Phys. A: Math. Gen.*, 19:L785–L790, 1986.

- [196] S.H. Lee and J.B. Hubbard. Electrolytes constrained on fractal structures: Debye-Hückel theory. *J. Chem. Phys.*, 98:1504–1508, 1993.
- [197] J.G. Kirkwood. Statistical mechanics of liquid solutions. *Chem. Rev.*, 19:275–307, 1936.
- [198] D. Chandler. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.*, 68:2959–2970, 1977.
- [199] S. Vembu. Fourier transformation of the n-dimensional radial delta function. *Quart. J. Math. Oxford*, 12:165–168, 1961.
- [200] L.E. Blumenson. A derivation of n-dimensional spherical coordinates. *Am. Math. Monthly*, 67:63–66, 1960.
- [201] H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H.E. Gaub. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Biol.*, 4:524–530, 2000.
- [202] E.M. Puchner and H.E. Gaub. Force and function: probing proteins with AFM-based force spectroscopy. *Curr. Opin. Struct. Biol.*, 19:605–614, 2009.
- [203] T.E. Fisher, A.F. Oberhauser, M. Carrion-Vazquez, P.E. Marszalek, and J.M. Fernandez. The study of protein mechanics with the atomic force microscope. *Trends. Biochem. Sci.*, 24:379–384, 1999.
- [204] T.E. Fisher, P.E. Marszalek, A.F. Oberhauser, M. Carrion-Vazquez, and J.M. Fernandez. The micro-mechanics of single molecules studied with atomic force microscopy. *J. Physiol.*, 520:5–14, 1999.
- [205] C. Bustamante. Entropic elasticity of lamda-phage DNA. *Science*, 265:1599–1600, 1994.
- [206] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, and H.E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 276:1109–1112, 1997.
- [207] M. Rief, M. Gautel, A. Schemmel, and H.E. Gaub. The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. *Biophys. J.*, 75:3008–3014, 1998.
- [208] P.E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A.F. Oberhauser, K. Schulten, and J. Fernandez. Mechanical unfolding intermediates in titin modules. *Nature*, 402:100–103, 1999.
- [209] M. Carrion-Vazquez, P.E. Marszalek, A.F. Oberhauser, and J.M. Fernandez. Atomic force microscopy captures length phenotypes in single proteins. *Proc. Natl. Acad. Sci. USA*, 96:11288–11292, 1999.

- [210] A.F. Oberhauser, P.E. Marszalek, H.P. Erickson, and J.M. Fernandez. The molecular elasticity of the extracellular matrix protein tenascin. *Nature*, 393:181–185, 1998.
- [211] R. Perez-Jimenez, S. Garcia-Manyes, S.R.K. Ainarapu, and J.M. Fernandez. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *J. Biol. Chem.*, 281:40010–40014, 2006.
- [212] M. Rief, J. Pascual, M. Saraste, and H.E. Gaub. Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J. Mol. Biol.*, 286:553–561, 1999.
- [213] M. Kim, K. Abdi, G. Lee, M. Rabbi, W. Lee, M. Yang, C.J. Schofield, V. Bennett, and P.E. Marszalek. Fast and forceful refolding of stretched alpha-helical solenoid proteins. *Biophys. J.*, 98:3086–3092, 2010.
- [214] A. Sarkar, S. Caamano, and J.M. Fernandez. The elasticity of individual titin pevk exons measured by single molecule atomic force microscopy. *J. Biol. Chem.*, 280:6261–6264, 2005.
- [215] D. Müller, M. Kessler, F. Oesterhelt, C. Möller, D. Oesterhelt, and H.E. Gaub. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys. J.*, 83:3587–3588, 2002.
- [216] A.F. Oberhauser, P.K. Hansma, M. Carrion-Vazquez, and J.M. Fernandez. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. USA*, 98:468–472, 2001.
- [217] H. Li, A.F. Oberhauser, S.D. Redick, M. Carrion-Vazquez, H.P. Erickson, and J.M. Fernandez. Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc. Natl. Acad. Sci. USA*, 98:10682–10686, 2001.
- [218] P.M. Williams, S.B. Fowler, R.B. Best, J.L. Toca-Herrera, K.A. Scott, A. Steward, and J. Clarke. Hidden complexity in the mechanical properties of titin. *Nature*, 422:446–449, 2003.
- [219] J.M. Fernandez and H. Li. Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science*, 303:1674–1678, 2004.
- [220] H. Dietz and M. Rief. Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. USA*, 101:16192–16197, 2004.
- [221] L. Li, H. Han-Li Huang, C.L. Badilla, and J.M. Fernandez. Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J. Mol. Biol.*, 345:817–826, 2005.

- [222] M. Schlierf and M. Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. *Biophys. J.*, 90:L33–L35, 2006.
- [223] M. Mickler, R.I. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc. Natl. Acad. Sci. USA*, 104:20268–20273, 2007.
- [224] Q. Peng and H. Li. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. USA*, 105:1885–1890, 2008.
- [225] D.N. Greene, T. Garcia, R.B. Sutton, K.M. Gernert, G.M. Benian, and A.F. Oberhauser. Single-molecule force spectroscopy reveals a stepwise unfolding of *Caenorhabditis elegans* giant protein kinase domains. *Biophys. J.*, 95:1360–1370, 2008.
- [226] M. Bertz and M. Rief. Mechanical unfoldons as building blocks of the maltose-binding protein. *J. Mol. Biol.*, 378:447–458, 2008.
- [227] S. Garcia-Manyes, J. Brujić, C.L. Badilla, and J.M. Fernandez. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys. J.*, 93:2436–2446, 2007.
- [228] S. Garcia-Manyes, J. Liang, R. Szoszkiewicz, T.-L. Kuo, and J.M. Fernandez. Force-activated reactivity switch in a biomolecular chemical reaction. *Nat. Chem.*, 1:236–242, 2009.
- [229] E. Evans and K. Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72:1541–1555, 1997.
- [230] I. Tinoco Jr. and C. Bustamante. The effect of force on thermodynamics and kinetics of single molecule reactions. *Biophys. Chem.*, 101-102:513–533, 2002.
- [231] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. USA*, 98:3658–3661, 2001.
- [232] D. Keller, D. Swigon, and C. Bustamante. Relating single-molecule measurements to thermodynamics. *Biophys. J.*, 84:733–738, 2003.
- [233] D.D.L. Minh. Multidimensional potentials of mean force from biased experiments along a single coordinate. *J. Phys. Chem. B*, 111:4137–4140, 2007.
- [234] O.K. Dudko, S. Hummer, and A. Szabo. Theory, analysis and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. USA*, 105:15755–15760, 2008.

- [235] G.E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721–2726, 1999.
- [236] E.A. Shank, C. Cecconi, J.W. Dill, S. Marquese, and C. Bustamante. The folding topology of a protein is controlled by its chain topology. *Nature*, 465:637–641, 2010.
- [237] S. Arrhenius. On the reaction rate of the inversion of non-refined sugar upon souring. *Z. Phys. Chem.*, 4:226–248, 1889.
- [238] G.I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200:618–627, 1978.
- [239] M. Schlierf, H. Li, and J.M. Fernandez. The unfolding kinetics of ubiquitin captured with single-molecule force clamp techniques. *Proc. Natl. Acad. Sci. USA*, 101:7299–7304, 2004.
- [240] L. Dougan, G. Feng, H. Lu, and J.M. Fernandez. Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc. Natl. Acad. Sci. USA*, 105:3185–3190, 2008.
- [241] A.P. Wiita, S.R. Ainarapu, H.H. Huang, and J.M. Fernandez. Force-dependent chemical kinetics of disulfide bond reduction observed with single-molecule techniques. *Proc. Natl. Acad. Sci. USA*, 103:7222–7227, 2006.
- [242] K.S.R. Ainarapu, A.P. Wiita, L. Dougan, E. Ufferud, and J.M. Fernandez. Single-molecule force spectroscopy measurements of bond elongation during a biomolecular reaction. *J. Am. Chem. Soc.*, 130:6479–6487, 2008.
- [243] O.K. Dudko, A.E. Filippov, J. Klafter, and M. Urbakh. Beyond the conventional description of dynamic force spectroscopy of adhesion bonds. *Proc. Natl. Acad. Sci. USA*, 100:11378–11381, 2003.
- [244] O.K. Dudko, G. Hummer, and A. Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96:108101, 2006.
- [245] G. Hummer and A. Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophys. J.*, 85:5–15, 2003.
- [246] T.-L. Kuo, S. Garcia-Manyes, J. Li, I. Barel, H. Lu, B.J. Berne, M. Urbakh, J. Klafter, and J.M. Fernandez. Probing static disorder in Arrhenius kinetics by single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. USA*, 107:11336–11340, 2010.
- [247] S.R.K. Ainarapu, L. Li, C.L. Badilla, and J.M. Fernandez. Ligand binding modulates the mechanical stability of dihydrofolate reductase. *Biophys. J.*, 89:3337–3344, 2005.

- [248] J.P. Junker, K. Hell, M. Schlierf, W. Neupert, and M. Rief. Influence of substrate binding on the mechanical stability of mouse dihydrofolate reductase. *Biophys. J.*, 89:L46–L48, 2005.
- [249] E. Hann, N. Kirkpatrick, C. Kleanthous, D.A. Smith, S.E. Radford, and D.J. Brockwell. The effect of protein complexation on the mechanical stability of Im9. *Biophys. J.*, 92:L79–L81, 2007.
- [250] Y. Cao, M.M. Balamurali, D. Sharma, and H. Li. A functional single-molecule binding assay via force spectroscopy. *Proc. Natl. Acad. Sci. USA*, 104:15677–15681, 2007.
- [251] Y. Cao, T. Yoo, S. Zhuang, and H. Li. Protein-protein interaction regulates proteins’ mechanical stability. *J. Mol. Biol.*, 378:1132–1141, 2008.
- [252] C.-C. Wang, T.-Y. Tsong, Y.-H. Hsu, and P.E. Marszalek. Inhibitor increases the mechanical stability of staphylococcal nuclease. *Biophys. J.*, 100:1094–1099, 2011.
- [253] F. Qian, W. Wei, G. Germino, and A. Oberhauser. The nanomechanics of polycystin-1 extracellular region. *J. Biol. Chem.*, 280:40723–40730, 2005.
- [254] E.M. Puchner, A. Alexandrovich, A.L. Kho, U. Hensen, L.V. Schäfer, B. Brandmeier, F. Gräter, H. Grubmüller, H.E. Gaub, and M. Gautel. Mechanoenzymatics of titin kinase. *Proc. Natl. Acad. Sci. USA*, 105:13385–13390, 2008.
- [255] J.P. Junker and M. Rief. Single-molecule force spectroscopy distinguishes target binding modes of calmodulin. *Proc. Natl. Acad. Sci. USA*, 106:14361–14366, 2009.
- [256] J.P. Junker, F. Ziegler, and M. Rief. Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science*, 323:633–637, 2009.
- [257] E.M. Puchner and H.E. Gaub. Exploring the conformation-regulated function of titin kinase by mechanical pump and probe experiments with single molecules. *Ang. Chem. Int. Ed.*, 49:1147–1150, 2010.
- [258] U.B. Sleytr and T.J. Beveridge. Bacterial S-layers. *Trends Microbiol.*, 7:253–260, 1999.
- [259] U.B. Sleytr and A.M. Glaupert. Analysis of regular arrays of subunits on bacterial surfaces; evidence for a dynamic process of assembly. *J. Ultrastruct. Res.*, 50:103–116, 1975.
- [260] D.J. Müller, W. Baumeister, and A. Engel. Controlled unzipping of a bacterial surface layer with atomic force microscopy. *Proc. Natl. Acad. Sci. USA*, 96:13170–13174, 1999.

- [261] S. Scheuring, H. Stahlberg, M. Chami, C. Houssin, J.-L. Rigaud, and A. Engel. Charting and unzipping the surface layer of *Corynebacterium glutamicum* with the atomic force microscope. *Mol. Microbiol.*, 44:675–684, 2002.
- [262] C. Verbelen, J. Antikainen, T.K. Korhonen, and Y.F. Dufrêne. Exploring the molecular forces within and between CbsA S-layer proteins using single molecule force spectroscopy. *Ultramicroscopy*, 107:1004–1011, 2007.
- [263] L. Dougan, L. Li, C.L. Badilla, B.J. Berne, and J.M. Fernandez. Single homopolypeptide chains collapse into mechanically rigid conformations. *Proc. Natl. Acad. Sci. USA*, 106:12605–12610, 2009.
- [264] A. Breitwieser, K. Gruber, and U.B. Sleytr. Evidence for an S-layer protein pool in the peptidoglycan of *Bacillus stearothermophilus*. *J. Bacteriol.*, 174:8008–8015, 1992.
- [265] H. Gollner. Dreidimensionale Rekonstruktion elektronenmikroskopischer Aufnahmen kristalliner Proteinmonoschichten. Diploma Thesis, Department for Nanobiotechnology, University of Natural Resources and Life Sciences, 2006.

List of Figures

1	Surface layer proteins. (A) Electron micrograph of a freeze-etched preparation of <i>Desulfotomaculum nigrificans</i> showing a regular S-layer lattice. ¹⁷ (B) Recrystallized patches of the S-layer protein SbpA from <i>Lysinibacillus sphaericus</i> CCM2177 on a silicon solid support visualized by atomic force microscopy. (C-F) Different lattices symmetries: (C) p1, (D) p2, (E) p4, (F) p6.	2
2	Self-assembly products of negatively stained S-layer proteins in solution visualized by transmission electron microscopy. ²⁴	3
3	(A) 3D models of the single domains of the S-layer protein SbsB created by fold recognition. Yellow arrows: beta-sheets, violet strands: alpha-helices, green line: turns, and red line: coils. Modeling method: (B) the individual domains were equilibrated in water spheres at 310K, (C) joined in vacuum, and (D) the final structure was obtained by molecular dynamics simulations. (E) Structural model of SbsB. The protein is L-shaped, where the L is formed by the C-terminal domains. The N-terminus contains the SLH domains and is mainly made up of alpha-helices. (F) Structural analysis of the monomer structure by a calculation of the global free energy. The protein was deformed along the reaction coordinate ζ . Mean force values MF are indicated by open blue circles. The red full line gives the potential of mean force PMF, which has a clear minimum at $\zeta = 0$. Orange body gives the local density probability distribution $P[\zeta]$. The model of the protein is given as an insert, the reaction coordinate is indicated, and the green and blue spheres indicate fixed regions.	17
4	Schematic representation of a SAXS experiment. An incident X-ray beam with a wave vector of $ \vec{k}_i = \frac{2\pi}{\lambda}$ is elastically scattered by atoms in solution and the scattered beam with $ \vec{k}_s = \frac{2\pi}{\lambda}$ is recorded as a function of the scattering angle θ . The scattered beam emerges from electrons that resonate with the frequency of the X-rays and consequently emit secondary waves, which interfere with each other. The scattering process is described by a Fourier transformation from the coordinates of the atoms ζ of the molecule to the reciprocal space of the scattering vectors $Q = \vec{k}_s - \vec{k}_i$. The scattering amplitudes $A(Q)$ are functions of the momentum transfer $ Q = \frac{4\pi \sin \theta}{\lambda}$ and represent the Fourier transform of the electron density distribution $\Delta\rho(\zeta)$. The phase change $\Delta\varphi = Q \cdot \zeta$ depends on the electron coordinates ζ , but is not detectable. Therefore, the detected scattering intensity $I(Q) = A(Q) ^2$ contains information about the electron density distribution of the sample.	18

5	Small-angle X-ray scattering experiments provide information about the electron density distribution of a molecule and therefore enable the determination of the shape, the mass and distances (left figure). In order to obtain a more detailed structural information different approaches might be applied possibly leading to the model of an elephant in a snake (right figure). This example just emphasizes the complex process of calculating a unique structure based on X-ray scattering data due to the fact that a variety of different structures may have the same scattering pattern. Figures from: <i>Le petit prince (The little prince)</i> , Antoine de Saint-Exupéry, 1943.	23
6	Schematic representation of a small-angle X-ray scattering instrument. An X-ray source, such as a synchrotron, a rotating anode or a sealed tube, produces X-rays of different wavelengths. An optical system selects appropriate wavelengths and focuses the beam on the detector plane. The sample is placed on a sample stage in a vacuum chamber, which can be moved in x- and y-direction. The unscattered initial beam (marked in red) is blanked by the beamstop. The scattered X-rays are detected by a two-dimensional detector.	26
7	Small-angle X-ray scattering set-up that was used for the investigation of S-layers in this work. SAXS was performed with Cu Ka radiation from a rotating anode generator equipped with a pinhole camera and an area detector. The SAXS measurements were performed at the Institute of Dynamics of Condensed Systems, Faculty of Physics, University of Vienna, Austria under the supervision of Herwig Peterlik.	26
8	SAXS data of (A) the monomeric solution and (B) the self-assemblies are given by gray filled circles and slopes for the fractal dimension are inserted as black lines: (A) $D=2.4$ and (B) $D=2.9$. The dashed blue line is the analytical fit P^D for (A) $P^{2.4}$ and (B) $P^{2.9}$. The full red line is the complete numerical reconstruction using a Monte Carlo algorithm generalized to fractal dimensions.	33
9	Numerical reconstruction of the SAXS scattering intensities of the monomeric sample. High electron densities are indicated by red beads. The structural model as calculated by molecular dynamics simulations is shown in blue.	33

- 10 (A) Schematic of the principle of an atomic force microscope. A sharp tip is mounted at the end of a flexible cantilever. Forces acting between the tip and the sample cause a deflection of the cantilever, which is detected by a laser beam. A piezo scanner enables a movement in x-, y- and z-direction. (B) Principle of AFM-based single-molecule force spectroscopy. A protein construct - here an S-layer protein embedded within five titin I27 domains is exemplarily shown - is stretched through the adsorption of one terminus to the AFM tip and consequent pulling. By exerting mechanical force the polyprotein is stochastically unfolded resulting in force-extension curves as shown in (C). The mechanical unfolding of proteins or protein domains at constant velocity causes a saw-tooth like force-extension pattern revealing a wealth of information regarding the structure, mechanical stability and unfolding pathways, where the number of peaks corresponds to the number of unfolded domains. Such characteristic response to mechanical force can be described by different entropic elasticity models, where here the worm-like chain model has been applied.²⁰⁵ 35
- 11 (A) JPK Nano Wizard I set-up. This atomic force microscope was used for all single-molecule force spectroscopy experiments presented in this work. (B) Polyprotein $(I27)_2 - SbsB - (I27)_3$ set-up: the S-layer protein SbsB is embedded within five I27 domains of the human muscle protein titin. The enlarged view shows the proposed interaction site between the N-terminal region of the S-layer protein and a secondary cell wall polymer (SCWP). SCWPs represent the specific binding partners for S-layer proteins within the peptidoglycan layer of the cell wall and anchor the proteins to the cell surface. Experiments were conducted independently using first the polyprotein alone and second, the protein bound to SCWP, in a constant-velocity mode. 37
- 12 Typical force-extension traces for the unfolding of the polyprotein $(I27)_2 - SbsB - (I27)_3$ for the unbound protein (A) and the protein bound to its ligand SCWP (B). (A) The force-extension trace shows five consecutive unfolding peaks for the five I27 domains (WLC fit in purple) and one double peak for the S-layer protein SbsB (WLC fit in blue). The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 12.1 \pm 4.9$ nm and between the second peak of the double peak and the first I27 peak $\Delta L_2 = 36.3 \pm 5.2$ nm. The mean unfolding force of the first peak is 86.8 ± 12.5 pN and 87.6 ± 10.8 pN, respectively ($n = 90$). The mean contour length increment between the consecutive I27 domains is 28.14 ± 2.2 nm and the mean unfolding force 197.89 ± 16.2 pN for $n=400$. (B) An additional peak is observed when the S-layer protein is bound to SCWP (marked by a blue star). The contour length increase between the first and the second peak of the double peak is $\Delta L_1 = 17.3 \pm 4.2$ nm, between the second peak of the double peak and the third peak $\Delta L_2 = 44.9 \pm 6.5$ nm and between the third peak and the first I27 peak $\Delta L_3 = 32.8 \pm 7.2$ nm. The unfolding forces are higher than for the unbound protein, namely $F_1 = 110.1 \pm 9.1$ pN, $F_2 = 109.9 \pm 9.8$ pN and $F_3 = 117.4 \pm 9.7$ pN. 38

13	Atomistic and coarse-grained structure of the S-layer protein SbpA as obtained by molecular dynamics simulations. The termini are marked in red. In the coarse-grained model every amino acid is represented by a single bead, where consecutive beads are linked by an harmonic potential.	40
14	Reconstruction of the three-dimensional density distribution of one SbpA unit cell calculated by inverse Fourier transform of single sections of SbpA self-assembly products. ²⁶⁵ A transmission electron microscope was used to obtain sections of the unit cell by performing tilting measurements. In the top left corner, an electron micrograph of one unit cell is shown. The density boundaries are systematically increased (starting in the top left corner). A superposition of all sections results in a three-dimensional unit cell.	41
15	Three-dimensional structure of the SbpA unit cell. Every monomer in the tetramer is illustrated in a different color. The proteins are interlocked into each other. (A) Inner surface of the tetramer, which anchors the protein to the cell surface. The N-termini are represented by magnified beads and are accessible on the surface. (B) Side view of the unit cell. (C) Outer surface of the tetramer, which is exposed to the surroundings of the cell. The C-termini are also accessible and marked as beads. (D) Scattering clusters (red beads) of one SbpA monomer as determined by SAXS and a Monte Carlo algorithm. The scattering clusters represent regions of high electronic contrast, where those domains in the protein, which do not show high electronic contrast, are related to interacting or overlapping parts in the tetramer.	42
16	Coarse-grained model of the S-layer protein SbsB. Each amino acid residue is represented by a single bead of 0.65 nm diameter, which is located at the center of mass. Negatively charged beads are colored in red and positively charged beads are colored in blue. The protein is treated as a rigid body.	43
17	Results of the Monte Carlo simulation of the interaction of two monomers. (A) (C) Two stable dimer conformations could be identified, where they are interacting in (A) y-direction and (C) x-direction, respectively. (B) (D) Corresponding dimer conformations, where the proteins are represented by unit cubes to be used in the lattice Monte Carlo simulation of the large-scale assembly process. Blue colored faces represent the interaction directions.	45
18	Morphology of the self-assembled S-layer lattice as obtained by lattice Monte Carlo simulations. (A) Lattice with p1 symmetry made up of 16 monomers, represented by the coarse-grained model. The magnified view shows the details of one pore. Hydrophilic residues are colored in blue, hydrophobic ones in red. (B) Cross section of the lattice. The outer surface is composed of the C-terminal region (blue), the inner surface of the N-terminal region (red).	46

List of Abbreviations

ABF	adaptive biasing force
AFM	atomic force microscope
CD	circular dichroism
Ig	immunoglobulin like
kDa	kilo Daltons
keV	kilo electron Volt
MF	mean force
NMR	nuclear magnetic resonance
PMF	potential of mean force
pN	pico Newtons
S-layer	surface layer
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SCWP	secondary cell wall polymer
SLH	S-layer homologous
UV	ultra violet
WLC	worm like chain

List of Symbols and Physical Constants

\mathcal{L}	Lagrange function
∇	Nabla operator
\mathcal{H}	Hamiltonian
\mathcal{F}	Fourier transform
λ	Lagrangian multiplier
J	Jacobian determinant
Y_{lm}	spherical harmonics
j_l	spherical Bessel functions
G	Green's function
K_ν	Bessel functions of the second kind
J_ν	Bessel functions of the first kind
δ	Delta function
Θ	Heaviside function
Γ	Gamma function
L_c	contour length
M_{ij}	Miyazawa-Jernigan statistical potential
κ	Debye length
ℓ_B	Bjerrum length
$k, k_B = 1.3806504(24) \times 10^{-23} JK^{-1}$	Boltzmann constant
$h = 6.62606896(33) \times 10^{-34} Js$	Planck constant

Curriculum vitae

Christine-Maria Horejs

Address:

Suedtirolerplatz 2/23

1040 Vienna

Austria

Tel.: +43-660-1201456

Mail: christine.horejs@gmail.com

Nationality: Austria

Birth: January 13th 1983

Gender: female

single, no children

EDUCATION

Sep 2011	PhD expected
Apr – May 2010	Visiting researcher at the Polymer Science and Engineering Center at Prof. Muthukumar's lab at the University of Massachusetts, Amherst, USA.
Sep 2009 – Jan 2010	Visiting researcher at the Polymer Science and Engineering Center at Prof. Muthukumar's lab at the University of Massachusetts, Amherst, USA.
2009	Recipient of a DOC-fORTE fellowship of the Austrian Academy of Sciences
2008	Start PhD (Dr.nat.techn.) , Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
2007	Master of Science (DI) , passed with distinction
2006	Master thesis at the Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
2006	Start Master Biotechnology , University of Natural Resources and Life Sciences, Vienna, Austria
2005	Bachelor of Science , passed with distinction
2005	Bachelor thesis at the Institute for Chemical and Energy Engineering, University of Natural Resources and Life Sciences, Vienna, Austria
2003	Change curriculum (due to university legislation): Bachelor of Food- and Biotechnology , University of Natural Resources and Life Sciences, Vienna, Austria
2001	Start diploma study Food- and Biotechnology , University of Natural Resources and Life Sciences, Vienna, Austria

2001	Matura passed with distinction
1993-2001	Bundesgymnasium Babenbergerring , Wr.Neustadt, Austria
1989-1993	Volksschule Felixdorf, Austria

THESES

Bachelor thesis: Molekulardynamische Simulation der Adsorption von Benzol aus einer wässrigen Lösung, 2005, Institute for Chemical and Energy Engineering, University of Natural Resources and Life Sciences, Vienna, Austria, supervised by Dr.Rupert Tscheliessnig and Prof. Dr.Johann Fischer

Master thesis: Structure prediction of an S-layer protein by the mean force method, 2007, Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, supervised by Dr.Dietmar Pum and Dr.Rupert Tscheliessnig

Dissertation: The structure and assembly of surface layer proteins: A combined approach of *in silico* and experimental methods. Department for Nanobiotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, supervised by Dr.Dietmar Pum and Dr.Rupert Tscheliessnig

WORK EXPERIENCE

2005 – 2006	Student Work, Department of Material Sciences and Process Engineering, Institute for Chemical and Energy Engineering University of Natural Resources and Life Sciences, Vienna, Austria Prof. Dr.Johann Fischer, Dr. Rupert Tscheliessnig Molecular dynamic simulations of the adsorption from aqueous solutions
July-August 2003	Microbiological laboratory, Baxter Bioscience, Vienna, QC-Mikrobiology Validation, SOP
July-August 2002	Microbiological laboratory, Baxter Bioscience, Vienna, QC-Microbiology Principles of microbiological quality control, identification of microorganisms, microbiological water control, monitoring, documentation

INTERNATIONAL EXPERIENCE

Apr – May 2010	Visiting researcher at the Polymer Science and Engineering Center at Prof. Muthukumar's lab at the University of Massachusetts, Amherst, USA
Sep 2009 – Jan 2010	Visiting researcher at the Polymer Science and Engineering Center at Prof. Muthukumar's lab at the University of Massachusetts, Amherst, USA.
Mar - June 2006	Ceepus semester abroad, Akademia Rolnicza (Agricultural University), Wroclaw, Poland Institute for Food- and Biotechnology
2005	3 weeks language course Polish at the school for polish language and culture Jagiellonen-University, Krakow, Poland

PUBLICATIONS

Horejs C., Mitra M.K., Pum D., Sleytr UB., Muthukumar M. (2011): Monte Carlo study of the molecular mechanisms of surface-layer protein self-assembly. *J. Chem. Phys.*, 134: 125103

Horejs C., Gollner H., Pum D., Sleytr UB., Peterlik H., Jungbauer A., Tscheliessnig, R. (2011): Atomistic structure of monomolecular surface layer self-assemblies: toward functionalized nanostructures. *ACS Nano*, 5: 2288

Horejs C., Pum D., Sleytr UB., Peterlik H., Jungbauer A., Tscheliessnig R. (2010): Surface layer protein characterization by SAXS and a fractal mean force concept: from protein structure to nanodisk assemblies. *J. Chem. Phys.*, 133: 175102

Horejs C., Pum D., Sleytr UB., Tscheliessnig R. (2008): Structure prediction of an S-layer protein by the mean force method. *J. Chem. Phys.*, 128: 065106

Horejs C., Ristl R., Sleytr UB., Tscheliessnig R., Pum D. (2011): Single-molecule force spectroscopy reveals the individual mechanical unfolding pathways of an S-layer protein. (revision resubmitted to the *Journal of Biological Chemistry*)

Sleytr UB., Schuster B., Egelseer EM., Pum D., **Horejs C.**, Tscheliessnig R., Ilk N. Nanobiotechnology with S-layer proteins as building blocks. In *Progress in molecular biology and translational Science: Molecular assembly in natural and engineered systems*. Ed. Howorka S., Elsevier, Amsterdam (2011)

CONFERENCE PROCEEDINGS

Horejs C., Mitra M.K., Pum D., Sleytr UB., Muthukumar M. (2011): Monte Carlo study of the molecular mechanisms of surface-layer protein self-assembly. *Biophys. J.* 100: 386a

Horejs C., Pum D., Sleytr UB., Tscheliessnig R. (2010): SAXS: Structure verification of an S-layer protein using a fractal mean force potential. *Biophys. J.* 98: 760a

Tscheliessnig R., Ueberbacher, R., **Horejs C.**, Peterlik, H., Jungbauer, A. (2010): Structural analysis of proteins by SAXS and a mean force concept. *SPICA2010, Elsevier*

TALKS

SAXS: Structure verification of an S-layer protein using a fractal mean potential. 54th Biophysical Society Meeting in San Francisco, California, USA. Session: Protein Structure. 2010

POSTER PRESENTATIONS

Monte Carlo study of the molecular mechanisms of S-layer protein self-assembly. 55th Biophysical Society Meeting in Baltimore, Maryland, USA. 2011

Assembly and modification of S-layer proteins studied by molecular dynamic simulation and force spectroscopy. 53rd Biophysical Society Meeting in Boston, Massachusetts, USA. 2009

S-layer proteins investigated by single-molecule force spectroscopy and *in silico* experiments. XI. Annual Linz Winter Workshop in Linz, Austria. 2009

LANGUAGES

German (mother tongue), English (fluent), Polish (advanced), proficiency advanced in Latin

REFERENCES

Uwe B. Sleytr
Department for Nanobiotechnology
University of Natural Resources and Life Sciences
Muthgasse 11
1190 Wien
uwe.sleytr@boku.ac.at
+43-1-476542200

Dietmar Pum
Department for Nanobiotechnology
University of Natural Resources and Life Sciences
Muthgasse 11
1190 Wien
dietmar.pum@boku.ac.at
+43-1-47654220

Murugappan Muthukumar
Department of Polymer Science and Engineering
University of Massachusetts
120 Governor's Drive
Amherst, MA, 01003
muthu@polysci.umass.edu
+1-413-545-0082

Rupert Tscheliessnig
Institute for Biotechnology
University of Natural Resources and Life Sciences
Muthgasse 18
1190 Wien
rupert.tscheliessnig@boku.ac.at
+43-1-476546227