

# Protein modeling and molecular dynamics simulation of the two novel surfactant proteins SP-G and SP-H

Felix Rausch · Martin Schicht · Lars Bräuer · Friedrich Paulsen · Wolfgang Brandt

Received: 3 April 2014 / Accepted: 21 October 2014 / Published online: 9 November 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Surfactant proteins are well known from the human lung where they are responsible for the stability and flexibility of the pulmonary surfactant system. They are able to influence the surface tension of the gas–liquid interface specifically by directly interacting with single lipids. This work describes the generation of reliable protein structure models to support the experimental characterization of two novel putative surfactant proteins called SP-G and SP-H. The obtained protein models were complemented by predicted posttranslational modifications and placed in a lipid model system mimicking the pulmonary surface. Molecular dynamics simulations of these protein-lipid systems showed the stability of the protein models and the formation of interactions between protein surface and lipid head groups on an atomic scale. Thereby, interaction interface and strength seem to be dependent on orientation and posttranslational modification of the protein. The here presented modeling was fundamental for experimental localization studies and the simulations showed that SP-G and SP-H are theoretically able to interact with lipid systems and thus are members of the surfactant protein family.

**Keywords** Lipid layer · Molecular dynamics · Protein-lipid interactions · Protein modeling · Surfactant proteins

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-014-2513-0) contains supplementary material, which is available to authorized users.

F. Rausch (✉) · M. Schicht · L. Bräuer · F. Paulsen  
Institute of Anatomy, Department II, Friedrich Alexander University  
Erlangen-Nuremberg, Erlangen, Germany  
e-mail: felix.rausch@anatomie2.med.uni-erlangen.de

W. Brandt  
Department of Bioorganic Chemistry, Leibniz Institute of Plant  
Biochemistry, Halle, Germany

## Introduction

The direct contact of the lung surface with the air exposes this organ to numerous environmental dangers and pathogens. Apart from the physical damage, evaporation of the surface and the underlying tissue or possible infections of the lung by various pathogens are the biggest problems. To prevent these complications, the surface of the lung alveoli is covered by a complex mixture of lipids and proteins with dipalmitoylphosphatidylcholine (DPPC) as the major lipid component [1]. This mixture, called pulmonary surfactant, is essential for the normal respiratory mechanism. Complications within this mechanism cause severe diseases like the acute respiratory distress syndrome (ARDS) [2] or even complete respiratory failure [3, 4]. Surfactant proteins (SP) considerably influence characteristics and stability of this lipid system. Accordingly, the extensive investigation of SPs is of great interest to develop new therapies against diseases or aftercare medication for operation or transplantation patients of respiratory medicine. Four SPs are known so far, which differ significantly in their characteristics. Surfactant proteins A and D are members of the C-type lectin family, which show immunological properties [5, 6]. SP-A and SP-D can interact with carbohydrates on the surface of different bacteria, protozoans, fungi, and viruses which leads to an accelerated immune defense and opsonization [7, 8]. In contrast to that, the small and very hydrophobic proteins SP-B and SP-C are essential for the stability of lipid monolayers at air-fluid interfaces [9–11]. They can control the surface tension and fluidity of the layer and regulate the insertion of new lipids into an existing system. To achieve their full functionality, these proteins are modified highly posttranslationally [12, 13] and are able to interact with other surfactant proteins. For example, protein cooperation was demonstrated for SP-A and SP-B [14]. All four proteins were initially identified within pulmonary surfactant, but recently, they were also detected on the

eye surface and in different tissues of the ocular system [15, 16].

By means of whole genome sequencing and bioinformatic sequence analysis, two additional potential SPs named SP-G [17] and SP-H [18] could be identified. Their amino acid sequences have an identity of 23 % and can be found in the UniProt database (accession codes Q6UW10 and P0C7M3) [19]. Their length of 78 amino acids for SP-G and 94 amino acids for SP-H is too short to show any similarity to the group of huge and hydrophilic SPs (SP-A, SP-D). Their sequence length indicates that SP-G and SP-H belong to the SP group of small and hydrophobic proteins (SP-B, SP-C), but they do not share any domains with the members of this group and the sequence identities are very low (about 10 %). Unfortunately, there was no further information about these proteins available prior to the presented studies. Their 3D structure was not known, no characterization of the proteins was done and their detailed localization or function was still completely undiscovered. With these few facts about the proteins, choosing the right experimental work for their further characterization is very difficult. Fortunately, computational chemistry methods like 3D structure modeling or molecular dynamics (MD) simulations can help out in those situations. There are many studies reported in the literature where modeling and MD simulations led to new insights which could promote research and gave valuable suggestions for further experimental studies. MD simulations showed the detailed interaction of SP-B with different lipid species [20, 21] and demonstrated the orientation of SP-B in the vicinity of a lipid layer [22, 23]. For SP-C, the stability of the protein fold was shown [24] and an important role for the formation of bilayer reservoirs [25] was verified *in silico*. Furthermore, the cooperation of SP-B and SP-C in an MD simulation caused an increased fluidity of a membrane system [26] and was crucial for the preservation and formation of a stable lipid layer system on air-fluid surfaces [26, 27]. As a prerequisite for these protein-lipid simulations, the possibility to reproduce a protein-free monolayer system consisting of lung surfactant lipids in an MD simulation was also described in the literature [28]. Finally, the immunological activity of SP-D was also demonstrated by simulation studies investigating the binding affinity of different sugar moieties, including glycans presented on the surface of the *influenza A* virus [29, 30].

The aim of this work was the investigation of the novel and putative surfactant proteins SP-G and SP-H with computational chemistry methods to get first insights into their character and function. For this purpose, reliable protein structure models were generated and complemented with posttranslational modifications predicted by statistical tools. MD simulations were performed with these 3D models to find out if SP-G and SP-H are able to interact with single lipids or lipid layers and with that, show typical surfactant protein behavior. For the protein-lipid simulations, a basic DPPC lipid layer

system mimicking the lung surfactant was established. The findings obtained during the modeling and simulation process were used to design and support experimental studies, for example the generation of specific antibodies for SP-G and SP-H and the localization of both proteins in different tissues by immunohistochemical methods [31, 32].

## Methods

### Protein structure modeling and posttranslational modifications (PTMs)

The protein sequence identity of SP-G and SP-H to the already known surfactant proteins is only about 10 % and there are no other protein structures with a high sequence identity available in the PDB. For this reason, comparative modeling was not possible and the protein sequences were sent to the *ab initio* folding server ROBETTA [33]. This computationally expensive method was able to produce protein structure models for SP-G and SP-H with promising overall quality. The stereochemical quality was evaluated by PROCHECK [34] after minor model optimizations with YASARA [35, 36]. PROSA II [37] was used to determine the quality of the entire protein fold based on the statistical analysis of well resolved protein X-ray structures. Furthermore, the model quality was assessed with ERRAT [38] and PROQ [39]. To check the stability of the protein models, 20 ns MD simulations were performed with YASARA and the YASARA2 force field [36]. For the simulation, each protein model was placed separately in a water box with a physiological NaCl concentration of 0.9 %. The final models were deposited at the Protein Model DataBase PMDB [40] for public download and received the PMDB id PM0078341 for SP-G and PM0079092 for SP-H. Additionally, these final models for SP-G and SP-H were extended by posttranslational modifications (PTMs), which were predicted by sequence-based prediction tools. Different statistic-based programs were used from the ExPASy bioinformatics resource portal [41]. The protein sequences were scanned for acetylation, *N*-glycosylation, *O*-glycosylation with N-Acetylglucosamine (GlcNAc) or N-Acetylgalactosamine (GalNAc) and phosphorylation with NetAcet [42], NetNGlyc [43], NetOGlyc [44], YinOYang [45], and NetPhos [46], respectively. Furthermore, the possibility of palmitoyl chains bound to free cysteine side chains was checked by CSS-Palm [47]. Predicted modifications were added manually to the protein structure models, followed by an energy minimization in YASARA. The final modified protein models were deposited at the Protein Model DataBase PMDB [40] as well and received the PMDB id PM0078342 for SP-G and PM0079093 for SP-H. For more details about the protein modeling procedure and PTM

prediction process, please see the respective papers for SP-G [31] and SP-H [32].

### DPPC simulation system setup

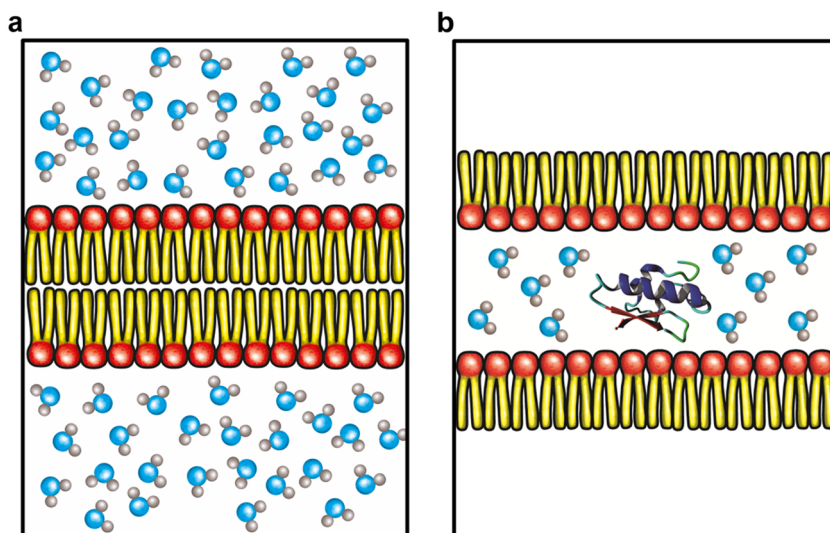
To simulate the SP-G and SP-H models in a natural environment, a basic DPPC lipid layer system was established. DPPC is the most abundant lipid in the pulmonary surfactant [48, 49] and for MD simulations described in the literature, DPPC-only lipid layers are often used to investigate different aspects of lung surfactant research [26, 50–52]. All simulations in this work were carried out with the GROMACS package version 4.5.4 [53, 54] and the united-atom G53a6 force field [55]. The standard parameter set of the force field for DPPC was slightly modified after Kukol [56] to produce a reliable lipid system. The initial bilayer consisting of 128 DPPC molecules per layer was built with the CELLmicrocosmos MembraneEditor 2.2 [57]. The bilayer was placed in the center of a simulation box and solvated with water (Fig. 1a). A simulation of 75 ns length indicated that the chosen lipid parameters and simulation settings are able to reproduce a stable lipid bilayer system. The MD simulation was performed with the Nosé-Hoover thermostat [58, 59] at 323 K and the Parrinello-Rahman barostat [60, 61] with semi-isotropic coupling and a reference pressure of 1 bar. The LINCS constraint algorithm [62, 63] was used to fix the stretching of all bonds, allowing a time step of 4 fs. Electrostatic interactions were calculated with the particle mesh Ewald (PME) algorithm [64, 65] as implemented in GROMACS with a cutoff at 1.2 nm, the van der Waals potential was switched off between 1.2 and 1.3 nm. The neighbor list was updated every five steps, energy and pressure dispersion correction was applied. The last 25 ns of the simulation were used to calculate area and volume per lipid, lateral diffusion coefficient and area compressibility. In order to estimate the simulation quality, these values were compared to literature data (area and volume per lipid [66], lateral diffusion

coefficient [67], and area compressibility [66, 68]). The last snapshot of this 75 ns MD simulation was used to build the DPPC monolayer system. The membrane layer with the lipids 1–128 was rotated by 180 degrees so that the polar lipid head groups were facing each other. Afterward, the layers were separated from each other generating space between the lipid head groups. Two systems were generated, one with lipid layers approx. 6.5 nm apart (hereafter referred to as “small system”) and one with approx. 9.5 nm space between the DPPC layers (hereafter referred to as “big system”). Both systems were placed in a simulation box with the lipid layers parallel to the x-y-plane. The z dimension of the box was set big enough to generate a 4–5 nm vacuum phase between the hydrophobic lipid tails due to the applied periodic boundary conditions. The space between the lipid head groups was filled with water molecules. A 25 ns MD simulation was performed to equilibrate the monolayer systems and check their stability. The compressibility of the systems in z direction was set to zero for these simulations to preserve the vacuum layer between the lipid tails. Apart from that, the simulation settings were identical to the bilayer calculations. The resulting monolayer systems were used to build the initial protein-lipid simulation layouts by placing the protein models in the water phase between the lipid head groups (Fig. 1b).

### SP-G and SP-H simulation in lipid environment

All four protein models (SP-G and SP-H without and with PTMs, respectively) were equilibrated by a 20 ns MD simulation in a water box with the G53a6 force field [55]. For this purpose, the force field was further modified with parameters for the attached PTM residues, namely phosphorylated serine, threonine and tyrosine, palmitoylated cysteine, serine or threonine residues that are *O*-glycosylated with GlcNAc or GalNAc and *N*-glycosylated asparagine. The *N*-glycosylation residue consists of a pentasaccharide core with two GlcNAc and three mannose

**Fig. 1** Representation of the simulation box layout for (a) the equilibration of a DPPC bilayer in water and (b) the protein-lipid simulations with two DPPC monolayers, which enclose a water phase with the protein model. Lipid head groups are shown in red and lipid carbon chains in yellow. Blue balls with gray hydrogen atoms depict regions filled with water. The protein backbone is shown in ribbon representation



moieties (–GlcNAc-GlcNAc-mannose-(mannose)<sub>2</sub>). The parameters for these residues were taken from original building blocks of the G53a6 force field (for example glucose or mannose building block) and combined with standard amino acid building blocks to describe the whole modified residue. Missing values for the connection between those parts were complemented manually with parameter sets also available from the original force field. A derivation of novel force field parameters was not necessary. In the case of the phosphorylated amino acids, parameters were taken from the G43a1p force field [69]. The equilibrated protein models were placed in arbitrary orientations in the water phase between the DPPC monolayers. This resulted in six different starting orientations per model, each system containing only one copy of the respective protein model. From these six starting orientations per model, four systems were built based on the “small system” and two were based on the “big system”. As a special feature, in one starting structure based on the “small system” for each modified protein, the model was manually positioned in a way that the palmitoylated cysteine residues are interacting with the lipid layer. That is, for the SP-G model with PTMs the palmitoyl moiety of Cys76 is in contact with the DPPC 1–128 layer and the palmitoylated Cys45 and Cys56 are already interacting with the DPPC 129–256 layer for the modified SP-H model at simulation start. Hydrogens were added to all structures according to pH 7 with an automated routine implemented in YASARA [70]. All 24 starting orientations (simulation systems) were neutralized with counter ions (Na<sup>+</sup>/Cl<sup>-</sup>) and submitted to a 250 ps equilibration run with NVT ensemble and the Berendsen thermostat at 323 K, followed by a 250 ps equilibration run with NPT ensemble and the Berendsen thermostat at 323 K and barostat at 1 bar. Afterward, a 50 ns production run was performed for all 24 orientations. The LINCS constraint algorithm [62, 63] was applied on all bonds involving hydrogens and the simulation time step was set to 2 fs. The Nosé-Hoover thermostat [58, 59] at 323 K and the Parrinello-Rahman barostat [60, 61] with semi-isotropic coupling and a reference pressure of 1 bar were used for temperature and pressure coupling. Similar to the monolayer equilibration MD, the compressibility in z dimension was set to zero to maintain the simulation box layout. Electrostatic interactions were calculated with a cutoff at 1.2 nm with the particle mesh Ewald (PME) algorithm [64, 65], the van der Waals potential was switched off between 1.2 and 1.3 nm. The neighbor list was updated every 10 steps and no dispersion correction was applied. Trajectories of the system were saved every 10 ps.

### MD simulation analysis

The analysis of the MD simulation results and trajectories was done with tools included in GROMACS [53, 54]. The overall energy, pressure, temperature, and box dimensions for the calculation of the area per lipid were extracted from the energy file by “g\_energy”. Furthermore, the introduction of two

energy groups “PROTEIN” and “DPPC” in the simulation settings allowed the calculation of the approximate protein-lipid interacting energy with respect to the force field parameters. The protein behavior was observed by root mean square deviation (RMSD) and root mean square fluctuation (RMSF) calculated with “g\_rms” and “g\_rmsf”. Finally, “do\_dssp” allowed determining major changes in the protein secondary structure during a simulation. For the visualization of the systems and results, VMD [71] and YASARA [35, 36] were used.

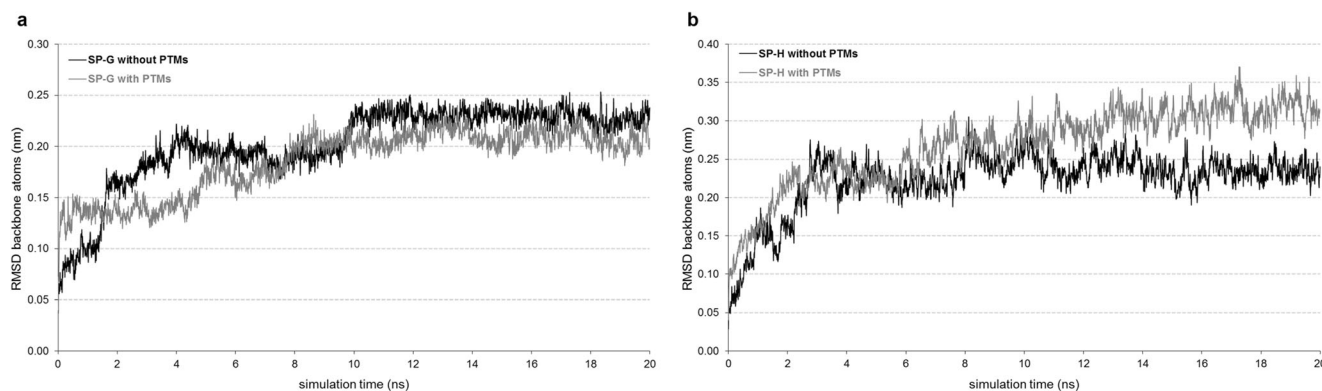
## Results

### Protein structure modeling and posttranslational modifications

The SP-G and SP-H protein models from ROBETTA initially showed very promising quality. Only minor optimizations and an MD refinement with YASARA were needed to achieve satisfactory results in structure validation tools. PROSA II shows a clearly negative plot for the whole SP-G structure model and the combined Z-score of –6.16 is close to the average value for proteins of this length (–7.77). PROCHECK determined 95.5 % of the 78 amino acids with a dihedral angle in the most favored regions of the Ramachandran plot. The ERRAT overall quality factor is 100 %, and PROQ calculated an LGscore of 3.579 and a MaxSub score of 0.141, which indicate a “very good” and “fairly good” model, respectively. Altogether, this suggests a reliable SP-G model structure, which shows no structure similarity to the already known surfactant proteins.

For the model of SP-H, the PROSA II plot is also completely negative and the Z-score (–5.72) is in acceptable distance to the length-dependent average value (–8.0), indicating a native-like fold of the model. In addition, the Ramachandran plot shows 94 % of the 94 amino acids with a dihedral angle in the most favored regions, implying a very high stereochemical quality. The overall quality factor of ERRAT is 93 %. The PROQ LGscore of 1.804 and MaxSub score of 0.131 indicate a “fairly good” model. Summarizing the model quality evaluations reveals a reliable protein structure model for SP-H, which also does not resemble the fold of one of the already known surfactant proteins.

Both protein models were subjected to a 20 ns MD simulation in a water box with YASARA to determine the model stability. The analysis of the root mean square deviation (RMSD) of the protein backbone atoms revealed that both protein models reach a stable conformation within a reasonable simulation time (Fig. 2a, black plot for SP-G, Fig. 2b, black plot for SP-H). The secondary structure element percentages of 47 % helix, 19 % sheet, and 34 % coil for SP-G



**Fig. 2** Plots of the protein backbone atoms RMSD (in nm) as a measure for the protein model stability for (a) SP-G and (b) SP-H during a 20 ns molecular dynamics simulation. Plots for the models without

posttranslational modifications (PTMs) are shown in black, plots for models with PTMs are shown in gray

and 50 % helix, 8 % sheet, and 42 % coil for SP-H remain unchanged during the simulation, which also indicates a stable protein fold. The protein models resulting from these simulations were completed by posttranslational modifications, which were determined by various statistic-based online prediction tools. For the manual attachment of the modifications, only predictions with high probability were considered and in the case of more than one predicted modifications for a position, only the modification with the highest probability was added to the protein model. According to Table 1a, two phosphorylations, three *O*-glycosylations with GlcNAc, one palmitoylation, and one *N*-glycosylation were added to the SP-G model. For the SP-H sequence, six phosphorylation sites, six *O*-glycosylations (two GlcNAc and four GalNAc) as well as two palmitoylations were predicted and attached to the protein model as stated in Table 1b. After the manual addition of the posttranslational modifications, the protein models were submitted to a 20 ns MD simulation in YASARA to check the influence of the attached modifications on the protein model stability in comparison to the unmodified models. Again, the RMSD values show a stable protein structure for SP-G (Fig. 2a, gray plot) and SP-H (Fig. 2b, gray plot), and no unfolding or major loss of secondary structure elements are visible.

With this, two model variants for each protein (with and without PTMs) were obtained, which maintain their good model quality during MD simulations and therefore allow the initial characterization of the 3D structures of SP-G and SP-H. Furthermore, they are suitable for computational chemistry studies in a lipid environment.

#### Preparation of the DPPC simulation system

The trajectories of the last 25 ns of the 75 ns DPPC bilayer MD simulation with the modified Gromos53a6 force field were used to calculate typical bilayer characteristics (Table 2). The volume per lipid settling at  $1.221 \text{ nm}^3$  is very similar to the experimental literature value of  $1.232 \text{ nm}^3$ . The lateral diffusion coefficient of  $9.2 \times 10^{-8} \text{ cm}^2/\text{s}$  nearly matches the experimental value of  $9.7 \times 10^{-8} \text{ cm}^2/\text{s}$ . The area compressibility of  $533 \text{ mN/m}$  is far off the experimental value of  $231 \text{ mN/m}$ , but is within the typical range of reported values for MD simulations ( $200\text{--}600 \text{ mN/m}$ ). As the primary criteria for a stable bilayer system, the area per lipid was calculated. In this simulation, it shows only minor fluctuations and remains stable at a level of about  $0.625 \text{ nm}^2$  (Fig. 3). This is very close to the experimentally determined value reported in the literature of  $0.64 \text{ nm}^2$  (blue line in Fig. 3). Altogether, this suggests

**Table 1** Predicted posttranslational modifications and their sequence positions in the (a) SP-G sequence and (b) SP-H sequence. Present modification types are phosphorylation (PHOS), palmitoylation (PALM),

*O*-glycosylation with GlcNAc (*O*-GLYC) or GalNAc (*O*-GAL), and *N*-glycosylation with a pentasaccharide core consisting of two GlcNAc and three mannose moieties (*N*-GLYC)

a							
Sequence position SP-G	Ser17	Asn37	Tyr40	Ser62	Ser70	Cys76	Thr78
Modification	PHOS	<i>N</i> -GLYC	PHOS	<i>O</i> -GLYC	<i>O</i> -GLYC	PALM	<i>O</i> -GLYC
b							
Sequence position SP-H	Ser32	Ser39	Cys45	Thr55	Cys56	Thr66	Thr69
Modification	PHOS	<i>O</i> -GLYC	PALM	PHOS	PALM	<i>O</i> -GAL	<i>O</i> -GAL
Sequence position SP-H	Thr75	Ser78	Ser80	Ser82	Ser83	Ser84	Thr93
Modification	<i>O</i> -GAL	<i>O</i> -GLYC	PHOS	PHOS	PHOS	PHOS	<i>O</i> -GAL

**Table 2** Comparison between characteristics for a DPPC bilayer reported in the literature and values obtained from simulations in this work

Value	Literature	Simulation
Volume/lipid (nm <sup>3</sup> )	1.232 [66]	1.221
Lateral diffusion coefficient (cm <sup>2</sup> /s)	$9.7 \cdot 10^{-8}$ [67]	$9.2 \cdot 10^{-8}$
Area compressibility (mN/m)	200-600 [68]	533

that the chosen force field parameters and simulation settings are able to reproduce a stable DPPC bilayer correctly and can be used for further studies.

### Protein-lipid MD simulation analysis

The protein model started to interact with the lipid layer in all 24 performed MD simulations. However, the results after 50 ns show a high diversity of protein parts that are responsible for the protein-lipid interactions. As can be seen from the final trajectory overlay of all six simulations per model (Fig. 4a, c, e, and g), no specific interaction site or “consensus orientation” can be identified for any of the four models. To pick a representative result for each case, the protein-lipid interaction strength as calculated by the force field was used as major criterion and the protein stability measured by RMSD of the backbone atoms was checked. Appendix Fig. 7 and Appendix Fig. 8 show the protein-lipid interaction energy and RMSD plots for all performed simulations.

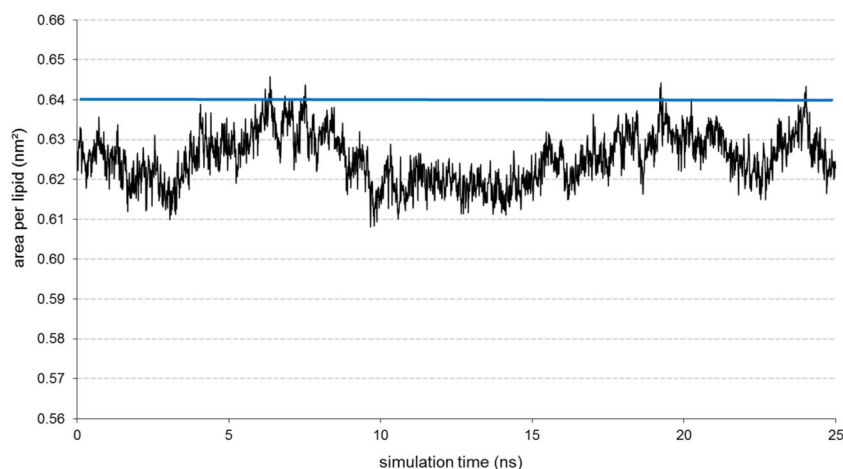
In the orientation with the most negative interaction energy for the SP-G model without PTMs (Fig. 4b), the N-terminus (1–14) and the residues of  $\alpha$ -helix 41–58 are mostly responsible for the protein-lipid contact. The first interactions establish after six ns, as visible in the interaction energy plot (Fig. 5a, black plot). After 30 ns, the interaction energy is essentially stable at a value of about  $-1100 \text{ kJ mol}^{-1}$ . The protein backbone RMSD plot for this simulation is not completely equilibrated, but nearly constant with only minor fluctuations after

25 ns. This indicates a stable protein structure (Fig. 5b, black plot). A closer investigation of the protein-lipid interaction site (Fig. 6a) reveals that there is only a small number of amino acid side chains interacting with the lipids. In the final simulation snapshot, three hydrogen bonds and four polar interactions between protein side chains and lipid phosphate or choline moieties are responsible for a moderate fixation of the protein on the lipid surface.

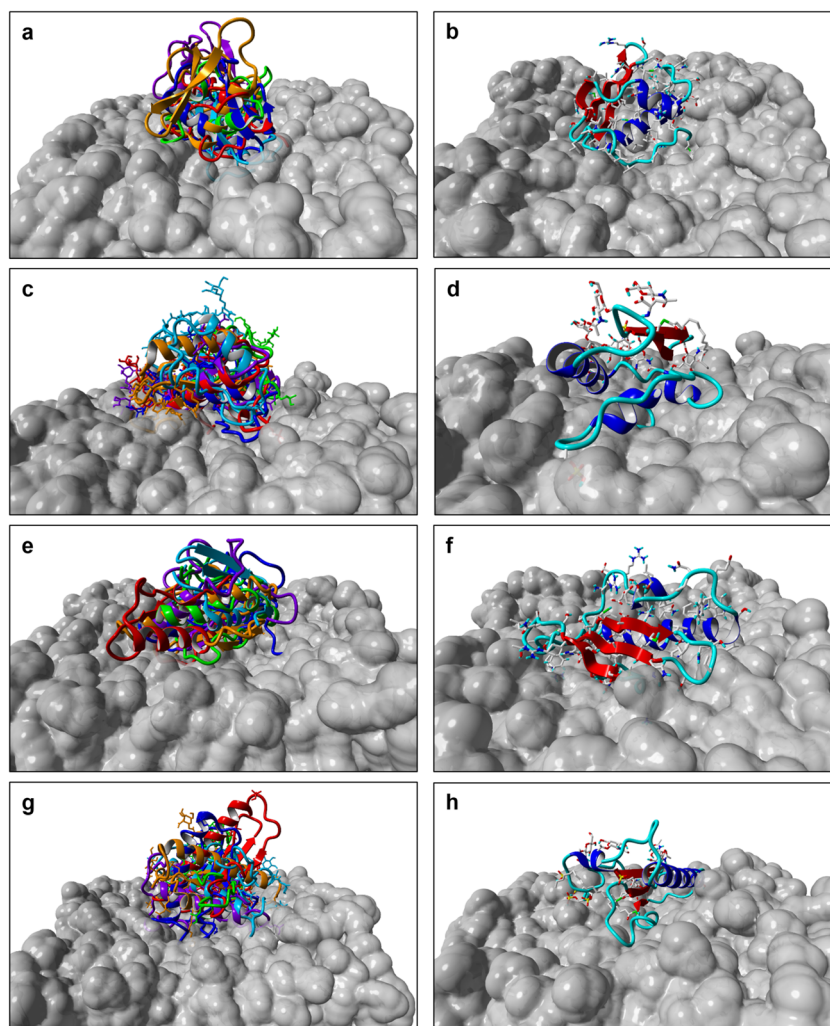
For the SP-G model with PTMs and most negative interaction energy, mainly the 18 N-terminal residues as well as amino acids 29–43 are in contact with the lipid layer (Fig. 4d). First protein-lipid interactions are visible after three ns of MD (Fig. 5a, gray plot) and increase quickly thereafter. Unfortunately, the interaction energy is not stable at the end of the simulation and may have been even stronger if the simulation had proceeded. The fact that the RMSD plot does not equilibrate after 50 ns (Fig. 5b, gray plot), reflects this as well. Conformational changes of the protein to adapt to the layer surface and optimize atomic interactions cause fluctuations in both graphs. However, the interaction energy of about  $-1800 \text{ kJ mol}^{-1}$  at the end of the simulation with PTMs attached to the SP-G model is already significantly stronger than the energy observed for the best SP-G model simulation without PTMs. This low interaction energy is also apparent from the protein-lipid interaction site (Fig. 6b). Compared to the results of the unmodified SP-G model, the number of interacting amino acids is increased (nine instead of five, interactions of Gly2, Ser3, and Glu46 are not shown in Fig. 6b due to clarity reasons). Hydrogen bonds are the dominant interaction type and Lys31 alone is responsible for interactions to fatty acid carbonyl groups of three different lipids. However, only one modified residue (phosphorylated Ser17) is interacting with a lipid, all other PTMs are interacting with the water phase.

The best simulation with the SP-H model without PTMs (Fig. 4f) shows a huge contact area between protein and lipids. In detail, especially the 27 N-terminal and nine C-terminal

**Fig. 3** Plot of the area per lipid for each DPPC molecule in a bilayer patch (in nm<sup>2</sup>) with 128 lipids during a 25 ns molecular dynamics simulation to verify the chosen parameters and simulation settings. The blue line denotes the experimental literature value for a DPPC molecule in a bilayer of  $0.64 \text{ nm}^2$  at 323 K [66]



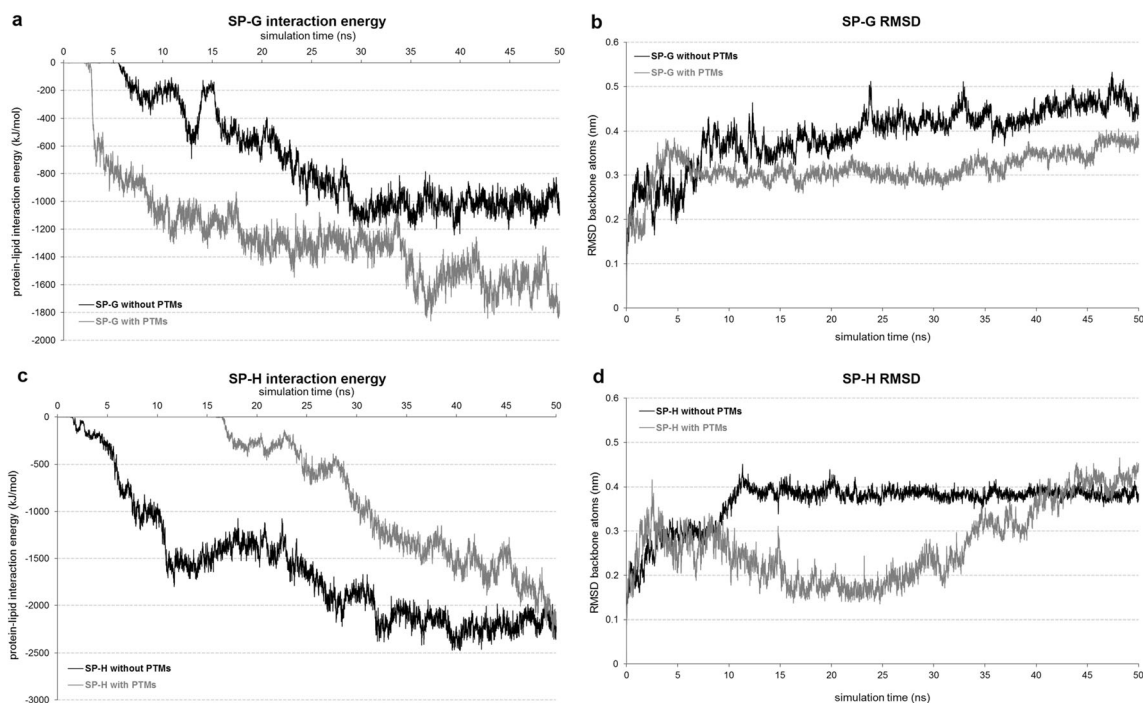
**Fig. 4** Resulting structures of MD simulations of the SP-G and SP-H models in a lipid environment. The left column shows the final trajectories of all six performed simulations (orientations) for each model superimposed in one figure, separately for the SP-G model without (a) and with (c) PTMs, and the SP-H model without (e) and with (g) PTMs. The DPPC lipids are shown as a gray surface, only the colored protein backbones and atoms of the PTMs (for c and g) are shown, each orientation with a different color. The right column shows the final simulation snapshots of the orientations with the most negative protein-lipid interaction energy, separately for the SP-G model without (b) and with (d) PTMs, and the SP-H model without (f) and with (h) PTMs. The protein backbones are shown in ribbon representation ( $\alpha$ -helices in blue,  $\beta$ -sheets in red, coil in cyan) and the DPPC lipid layer as gray surface. Atoms of the amino acid side chains (in b and f) and PTMs (in d and h) are shown without aliphatic hydrogens in a stick representation



amino acids are in close contact with the lipid layer. Accordingly, the interaction energy plot shows a steady decrease following the very early first contact at two ns until reaching a plateau after 40 ns at circa  $-2300 \text{ kJ mol}^{-1}$  (Fig. 5c, black plot). The protein model, meanwhile, is strikingly stable in this simulation. There are no major fluctuations of the RMSD plot later than 10 ns and the model can be denoted as equilibrated after 20 ns (Fig. 5d, black plot). The reason for the model stability could be the numerous interactions between amino acids and lipid head groups, which fix the protein on the lipid surface (Fig. 6c). Positively charged amino acid side chains form three of nine observed interactions (Arg2, Gln23, Glu27, Met88, and Leu89 are not shown in Fig. 6c due to clarity reasons) and serve as “anchors” in the ester bond region of the lipid layer.

The SP-H model with PTMs and most negative interaction energy also shows a large contact area with the N-terminus and C-terminal residues (with phosphorylations) being very important (Fig. 4h). In this case, the residues 32–51 also form numerous interactions. The first interaction energy between

protein and lipids can be spotted after 16 ns (Fig. 5c, gray plot) and is quickly decreasing to a value comparable to the simulation without PTMs ( $-2300 \text{ kJ mol}^{-1}$ ). Unfortunately, it is clearly not stable at the end of the calculation, but stronger than for all of the other five simulations with the SP-H model and PTMs. This instability is also reflected in the RMSD plot (Fig. 5d, gray plot), which shows significant fluctuations until the end of the simulation at 50 ns. The extension of this simulation until 100 ns showed a stable interaction energy at about  $-2300 \text{ kJ mol}^{-1}$  and an equilibrated protein model with respect to the RMSD after 60 ns (data not shown). This indicates that there is nearly no difference in the best interaction energy between the SP-H model without and with PTMs. However, the strong positive side chain interactions observed in the protein-lipid interaction of the model without PTMs are absent for the model with PTMs (Fig. 6d). This is compensated by a significant increase of the interaction count from nine to 14 (interactions of Arg24, Trp28, Leu31, Thr42, Arg49, Glu50, glycosylated Ser39, and Ala94 are not shown in Fig. 6d due to clarity reasons). Furthermore, two



**Fig. 5** Protein-lipid interaction energy and RMSD plots of the 50 ns MD simulations. Only the results for the orientation with the most negative protein-lipid interaction energy are shown at the end of the simulation for each protein model. The plots for the protein-lipid interaction energy (in

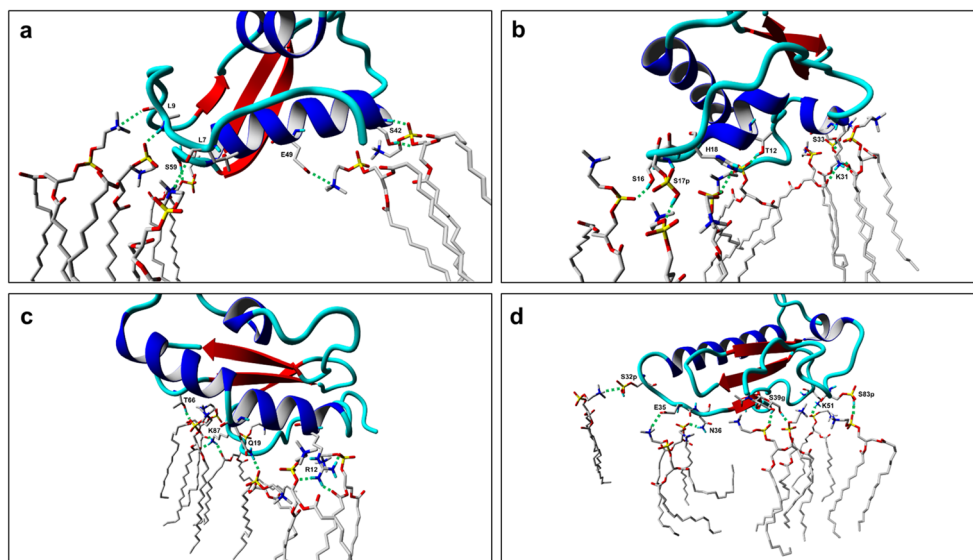
kJ mol<sup>-1</sup>) and protein backbone atoms RMSD (in nm) are shown for SP-G (a and b) and SP-H (c and d), respectively. Curves for models without PTMs are shown in black, for those with PTMs in grey

glycosylated residues (Ser39 and Thr93) and two phosphorylated amino acids (Ser32 and Ser83) contribute to the protein-lipid interaction energy.

The fluctuation analysis of each protein residue during the simulation (RMSF) for all 24 orientations (Appendix Fig. 9) indicates a reduced fluctuation of protein parts in general, which are interacting with the polar lipid head groups. This is due to hydrogen bonds and ionic interactions not only of the

amino acid side chain atoms, but also of protein backbone atoms with the lipid head groups. Polar PTMs like phosphorylations or glycosylations enhance this effect. In contrast, these PTMs increase the fluctuation of their attached protein parts if they are oriented toward the water phase. The area per lipid was also monitored for all simulations, but there was no case where the binding of the protein model to the lipid layer introduced any significant change in the area per lipid plot.

**Fig. 6** Detailed representation of the protein-lipid interaction sites after 50 ns MD for the lowest interaction energy poses for the SP-G model without (a) and with (b) PTMs as well as the SP-H model without (c) and with (d) PTMs. The protein is shown in ribbon representation. Amino acids and lipids are shown without aliphatic hydrogens in stick representation. Green dashed lines indicate interactions between amino acid sidechains and lipids. Labels show amino acid type, number, and modification (“p” for phosphorylation, “g” for glycosylation)





For all simulations, this value reaches approximately  $0.54 \text{ nm}^2$ , with a fluctuation of about  $\pm 0.02 \text{ nm}^2$ , which can be ascribed to the MD methodology.

## Discussion

Although there were no proteins with already known 3D structure and high sequence homology available for comparative modeling, structure models for SP-G and SP-H were obtained by ab initio protein structure prediction using ROBETTA. Common evaluation tools and a 20 ns MD simulation showed the good quality and stability of the models. This demonstrates that ROBETTA is able to produce valuable models also for practically oriented studies, besides the excellent performance in structure modeling contests (CASP) [72].

In the literature, the high impact of posttranslational modifications (PTMs) on the stability and function of surfactant proteins is a well-known fact [12, 13]. To consider this for the putative surfactant proteins SP-G and SP-H, their models were extended with PTMs obtained by sequence-based prediction tools. Although a conclusive experimental evidence of the determined and attached modifications is still pending, the reliability of the applied prediction algorithms is in general between 75 and 93 % [42–47].

The final models for SP-G and SP-H without and with PTMs were used to perform 24 MD simulations in a lipid environment. A typical feature of surfactant proteins is their ability to interact with lipids, as reported by previous studies especially for SP-B and SP-C [25, 26, 73, 74]. Correspondingly, the SP-G and SP-H models were simulated in the presence of a DPPC monolayer. This meets the current understanding of the pulmonary surfactant layout and DPPC as major lipid component of the pulmonary surfactant [48, 49] was already shown to adequately reproduce the surfactant system of the lung in simulations [26, 50–52]. Parameters and settings for MDs with similar systems were extensively studied in the literature and needed only minor adaptations for the PTMs attached to the protein models. All calculations were performed at a temperature of 323 K, which is above the phase transition temperature of DPPC at 314 K [75, 76]. This ensured that the lipid system was present in the biologically relevant fluid  $L_\alpha$  state instead of the more ordered gel or subgel state of a lipid layer [77]. To estimate the influence of the higher temperature on the protein stability, MD simulations of all models were performed in a water box at 298 K. The calculations showed no significant changes in stability or structure of the protein models (data not shown). In contrast to other studies, the lipid layer system for this work was built from scratch to obtain a lipid layer patch with the appropriate dimensions for the protein sizes. Literature values for comparable systems were reproduced successfully.

The 24 performed simulations mostly showed the stability of the protein model fold in the RMSD plots and demonstrated the influence of the PTMs on the secondary structure. Bulky modifications like the *N*- or also *O*-glycosylations can introduce flexibility to their connected protein region due to their rapidly changing hydrogen bonding partners (i.e., water molecules) in free solution. On the other hand, they can significantly stabilize a protein region when they form mostly hydrogen bonds with the polar head groups of DPPC molecules. This demonstrates the influence of the PTMs on the stability and interaction potential of both proteins.

Most of the interactions were established between polar amino acid side chains or PTMs and the polar head groups of the lipid molecules. Nearly no contact of protein parts with the hydrophobic lipid tail region was observed. The results of the simulations showed no direct impact of the protein-lipid interaction on the layer stability or lipid ordering. The literature suggests that longer simulations in the microsecond range may be required to observe protein mediated events like lipid layer folding or lipid vesicle fusion [26, 74]. Such long simulations would be computationally too expensive for the here used united atom approach. A method called coarse-grained simulations [78, 79] with reduced complexity developed especially for long-term simulations would be the technique of choice for future experiments. For this, knowledge about the 3D protein structure is very important and a required input, because currently the most commonly used MARTINI coarse grained force field is unable to model conformational changes of a protein [78, 80]. The simulation results of this work demonstrate the stability of the protein fold in most cases, even during the formation of interactions between protein and lipid layer. Therefore, the here performed calculations provide the requirements for coarse-grained simulations.

Although the protein models were between 1.5 and 3.5 nm apart from the lipid layer at the simulation start, they began to interact mostly within 25 ns of simulation, in some cases already after less than five ns. This process was traceable by monitoring the protein-lipid interaction energy. In this way, it was possible to discriminate between different interaction scenarios and visualize the influence of polar amino acids and PTMs on the interaction strength. However, the here used energies calculated based on force field parameters can only give a rough estimation of in vivo energies, since the accuracy of force fields reproducing intermolecular (i.e., non-bonded) interaction energies is limited [81]. For more detailed insights, advanced computational chemistry techniques like semi-empirical [82] or QM/MM methods [83], or experimental studies like the isothermal titration calorimetry (ITC [84]) would be advantageous. However, the fact that all 24 performed simulations showed a clear interaction between protein model and lipid layer strongly supports the hypothesis that SP-G and SP-H are indeed able to interact with lipids and may be capable of surface-regulatory features.

Although both proteins were annotated to the surfactant protein family due to bioinformatics prediction [17, 18], their actual family membership was questionable on the basis of the available data. The results of this work provide several indications that SP-G and SP-H are indeed surfactant proteins. Their high grade of modification is similar to the already known surfactant proteins. Apart from polar modifications like phosphorylations and glycosylations, they also show hydrophobic modifications. This could allow SP-G and SP-H to present an amphiphilic protein surface, as is typical for surfactant proteins [12, 13]. Previous attempts to produce specific antibodies for localization studies failed. However, with the here obtained knowledge about the 3D structure and modification pattern, it was possible to identify PTM-free protein surface regions. Their use as antigen peptides led to specific antibodies for SP-G and SP-H. The successful production of these antibodies on the one hand indicated a high reliability of the protein models and on the other hand allowed localization studies. Immunohistochemical staining showed that SP-G [31] and SP-H [32] are present in tissues of the human lung and eye, mostly membrane associated. These are tissues where the already known surfactant proteins are also present and play a crucial role [5–11, 15, 16]. Furthermore, the antibodies allowed first functional studies, which showed that inflammatory cytokines could influence the SP-H expression [32]. This could indicate an immunoregulatory function of SP-H comparable to SP-A and SP-D [5, 6]. Finally, the simulations showed the potential of SP-G and SP-H to interact with lipid systems as described for SP-B and SP-C [9–11]. Altogether, these points strongly support the hypothesis that SP-G and SP-H are indeed part of the surfactant protein family.

## Conclusions

With the help of ab initio protein structure prediction it was possible to obtain 3D models for the two putative surfactant proteins SP-G (SFTA2) and SP-H (SFTA3), although there are no homologue proteins with already known 3D structure available. Common quality assessment tools indicated a native-like fold of the proteins models and molecular dynamics simulations demonstrated the stability of the SP-G and SP-H model fold. The models were extended by posttranslational modifications (PTMs), because the literature states the high importance of PTMs for the function of the already known surfactant proteins. Sequence-based prediction tools indicated numerous phosphorylations, glycosylations, and palmitoylations for SP-G and SP-H, which were manually added to the protein models and did not influence the overall model stability in MD simulations.

Previous attempts to obtain specific antibodies for SP-G and SP-H failed due to the lack of knowledge about the three-

dimensional protein structure. The models obtained in this work revealed sequence parts on the surface of the proteins without any PTM, which are suitable antigens for the production of specific antibodies. In this way, the computational modeling significantly promoted experimental work, because the antibodies allowed the first localization of SP-G and SP-H in different cell tissues where other SPs are also present. Furthermore, they could be used in first functional studies [31, 32].

To mimic the basic properties of the pulmonary surfactant, a simulation system containing a DPPC lipid monolayer was established. This system was used to study the characteristics of the SP-G and SP-H model without and with PTMs in their natural environment in 24 MD simulations over a time of 50 ns each. Although the strength of the interactions and contact areas on the protein surface were dependent on the starting structure and attached PTMs, all performed simulations indicated a high potential of SP-G and SP-H to interact with a lipid system. Furthermore, the calculation results suggest that position and conformation of PTMs could be responsible for an amphiphilic character of both proteins, as described for the already known surfactant proteins. The high theoretical lipid interaction potential determined by the presented simulations could be used to support and discuss the outcome of experimental characterization and localization studies [31, 32] which suggest that SP-G and SP-H are indeed part of the surfactant protein family.

**Acknowledgments** This research was supported by the Deutsche Forschungsgemeinschaft (DFG, <http://www.dfg.de/>) given to Wolfgang Brandt (grant: BR 1329/12-1) and Lars Bräuer (grant BR 3681/2-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Sylvia Dyczek for proof-reading of the manuscript.

## References

1. Yu SH, Possmayer F (2003) Lipid compositional analysis of pulmonary surfactant monolayers and monolayer-associated reservoirs. *J Lipid Res* 44:621–629
2. Griese M (1999) Pulmonary surfactant in health and human lung diseases: state of the art. *Eur Respir J* 13:1455–1476
3. Gortner L, Hilgendorff A (2004) Surfactant-associated proteins B and C: molecular biology and physiologic properties. *Z Geburtshilfe Neonatol* 208:91–97
4. Halliday HL (2008) Surfactants: past, present and future. *J Perinatol* 28(Suppl 1):S47–S56
5. Wright JR (2005) Immunoregulatory functions of surfactant proteins. *Nat Rev Immunol* 5:58–68
6. Kishore U, Greenhough TJ, Waters P, Shrive AK, Ghai R et al (2006) Surfactant proteins SP-A and SP-D: structure, function and receptors. *Mol Immunol* 43:1293–1315
7. Ferguson JS, Voelker DR, McCormack FX, Schlesinger LS (1999) Surfactant protein D binds to Mycobacterium tuberculosis bacilli and lipoarabinomannan via carbohydrate-lectin interactions resulting in

- reduced phagocytosis of the bacteria by macrophages. *J Immunol* 163:312–321
8. Hartshorn KL, Crouch E, White MR, Colamussi ML, Kakkanatt A et al. (1998) Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. *Am J Physiol* 274:L958–L969
  9. Robertson BV, Lambert MG (1992) Pulmonary surfactant: from molecular biology to clinical practice. Elsevier, Amsterdam
  10. Yu SH, Possmayer F (1990) Role of bovine pulmonary surfactant-associated proteins in the surface-active property of phospholipid mixtures. *Biochim Biophys Acta* 1046:233–241
  11. Notter RH, Shapiro DL, Ohning B, Whitsett JA (1987) Biophysical activity of synthetic phospholipids combined with purified lung surfactant 6000 dalton apoprotein. *Chem Phys Lipids* 44:1–17
  12. Voorhout WF, Veenendaal T, Haagsman HP, Weaver TE, Whitsett JA et al. (1992) Intracellular processing of pulmonary surfactant protein-B in an endosomal lysosomal compartment. *Am J Physiol* 263:L479–L486
  13. Glasser SW, Korfhagen TR, Perme CM, Pilot-Matias TJ, Kister SE et al. (1988) Two SP-C genes encoding human pulmonary surfactant proteolipid. *J Biol Chem* 263:10326–10331
  14. Kobayashi T, Nitta K, Takahashi R, Kurashima K, Robertson B et al. (1991) Activity of pulmonary surfactant after blocking the associated proteins SP-A and SP-B. *J Appl Physiol* 71:530–536
  15. Bräuer L, Johl M, Borgermann J, Pleyer U, Tsokos M et al. (2007) Detection and localization of the hydrophobic surfactant proteins B and C in human tear fluid and the human lacrimal system. *Curr Eye Res* 32:931–938
  16. Bräuer L, Kindler C, Jäger K, Sel S, Nölle B et al. (2007) Detection of surfactant proteins A and D in human tear fluid and the human lacrimal system. *Invest Ophthalmol Vis Sci* 48:3945–3953
  17. Zhang Z, Henzel WJ (2004) Signal peptide prediction based on analysis of experimentally verified cleavage sites. *Protein Sci* 13:2819–2824
  18. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C et al (2003) The DNA sequence and analysis of human chromosome 14. *Nature* 421:601–607
  19. Consortium U (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40:D71–D75
  20. Lee H, Kandasamy SK, Larson RG (2005) Molecular dynamics simulations of the anchoring and tilting of the lung-surfactant peptide SP-B1-25 in palmitic acid monolayers. *Biophys J* 89:3807–3821
  21. Bertani P, Vidovic V, Yang TC, Rendell J, Gordon LM et al. (2012) Orientation and depth of surfactant protein B C-terminal helix in lung surfactant bilayers. *Biochim Biophys Acta* 1818:1165–1172
  22. Kandasamy SK, Larson RG (2005) Molecular dynamics study of the lung surfactant peptide SP-B1-25 with DPPC monolayers: insights into interactions and peptide position and orientation. *Biophys J* 88:1577–1592
  23. Kim HI, Kim H, Shin YS, Beegle LW, Jang SS et al. (2010) Interfacial reactions of ozone with surfactant protein B in a model lung surfactant system. *J Am Chem Soc* 132:2254–2263
  24. Kovacs H, Mark AE, Johansson J, van Gunsteren WF (1995) The effect of environment on the stability of an integral membrane helix: molecular dynamics simulations of surfactant protein C in chloroform, methanol and water. *J Mol Biol* 247:808–822
  25. Baoukina S, Monticelli L, Amrein M, Tieleman DP (2007) The molecular mechanism of monolayer-bilayer transformations of lung surfactant from molecular dynamics simulations. *Biophys J* 93:3775–3782
  26. Duncan SL, Larson RG (2010) Folding of lipid monolayers containing lung surfactant proteins SP-B(1–25) and SP-C studied via coarse-grained molecular dynamics simulations. *Biochim Biophys Acta* 1798:1632–1650
  27. Zhang H, Zhang SA, Lu CH, Jia TQ, Wang ZG et al. (2011) Single-photon fluorescence enhancement in IR144 by phase-modulated femtosecond pulses. *Chem Phys Lett* 503:176–179
  28. Javanainen M, Monticelli L, Bernardino de la Serna J, Vattulainen I (2010) Free volume theory applied to lateral diffusion in Langmuir monolayers: atomistic simulations for a protein-free model of lung surfactant. *Langmuir* 26:15436–15444
  29. Zhang JL, Zheng QC, Zhang HX (2010) Unbinding of glucose from human pulmonary surfactant protein D studied by steered molecular dynamics simulations. *Chem Phys Lett* 484:338–343
  30. van Eijk M, Rynkiewicz MJ, White MR, Hartshorn KL, Zou X et al. (2012) A unique sugar-binding site mediates the distinct anti-influenza activity of pig surfactant protein D. *J Biol Chem* 287:26666–26677
  31. Rausch F, Schicht M, Paulsen F, Ngueya I, Brauer L et al. (2012) “SP-G”, a putative new surfactant protein–tissue localization and 3D structure. *PLoS One* 7:e47789
  32. Schicht M, Rausch F, Finotto S, Mathews M, Mattil M, Schubert M, Koch B, Traxdorf M, Bohr C, Worlitzsch D, Brandt W, Garreis F, Sel S, Paulsen F, Bräuer L (2014) SFTA3, a novel protein of the lung — 3D-structure, characterization and immune activation. *Eur Respir J* 44:447–456
  33. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32:W526–W531
  34. Laskowski RA, MacArthur DS, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26:283–291
  35. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins* 47:393–402
  36. Krieger E, Joo K, Lee J, Raman S, Thompson J et al. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. *Proteins* 77(Suppl 9):114–122
  37. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17:355–362
  38. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519
  39. Wallner B, Elofsson A (2003) Can correct protein models be identified? *Protein Sci* 12:1073–1086
  40. Castrignano T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A (2006) The PMDB Protein Model Database. *Nucleic Acids Res* 34:D306–D309
  41. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G et al (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 40:W597–W603
  42. Kiemer L, Bendtsen JD, Blom N (2005) NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics* 21:1269–1270
  43. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4:1633–1649
  44. Julenius K, Molgaard A, Gupta R, Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15:153–164
  45. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 310–322
  46. Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294:1351–1362
  47. Ren J, Wen L, Gao X, Jin C, Xue Y et al (2008) CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Eng Des Sel* 21:639–644
  48. Goerke J (1998) Pulmonary surfactant: functions and molecular composition. *Biochim Biophys Acta* 1408:79–89
  49. Veldhuizen R, Nag K, Orgeig S, Possmayer F (1998) The role of lipids in pulmonary surfactant. *Biochim Biophys Acta* 1408:90–108

50. Knecht V, Muller M, Bonn M, Marrink SJ, Mark AE (2005) Simulation studies of pore and domain formation in a phospholipid monolayer. *J Chem Phys* 122:024704
51. Mohammad-Aghaie D, Mace E, Sennoga CA, Seddon JM, Bresme F (2010) Molecular dynamics simulations of liquid condensed to liquid expanded transitions in DPPC monolayers. *J Phys Chem B* 114:1325–1335
52. Rose D, Rendell J, Lee D, Nag K, Booth V (2008) Molecular dynamics simulations of lung surfactant lipid monolayers. *Biophys Chem* 138:67–77
53. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE et al. (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718
54. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447
55. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) 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
56. Kukol A (2009) Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *J Chem Theory Comput* 5:615–626
57. Sommer B, Dingersen T, Gamroth C, Schneider SE, Rubert S et al. (2011) CELLmicrocosmos 2.2 MembraneEditor: a modular interactive shape-based software approach to solve heterogeneous membrane packing problems. *J Chem Inf Model* 51:1165–1182
58. Nose S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52:255–268
59. Hoover WG (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys Rev A* 31:1695–1697
60. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52:7182–7190
61. Nose S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50:1055–1076
62. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472
63. Hess B (2008) P-LINCS: A parallel linear constraint solver for molecular simulation. *J Chem Theory Comput* 4:116–122
64. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092
65. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H et al (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593
66. Nagle JF, Tristram-Nagle S (2000) Structure of lipid bilayers. *Biochim Biophys Acta* 1469:159–195
67. Konig S, Pfeiffer W, Bayerl T, Richter D, Sackmann E (1992) Molecular-dynamics of lipid bilayers studied by incoherent quasi-elastic neutron-scattering. *J Phys II* 2:1589–1615
68. Anézo C, de Vries AH, Höltje H-D, Tieleman DP, Marrink S-J (2003) Methodological issues in lipid bilayer simulations. *J Phys Chem B* 107:9424–9433
69. Smith GR (2002) G43a1 force field modified to contain phosphorylated Ser, Thr and Tyr. GROMACS User Contributions. [http://www.gromacs.org/Downloads/User\\_contributions/Force\\_fields](http://www.gromacs.org/Downloads/User_contributions/Force_fields). Accessed 2 Oct 2014.
70. Krieger E, Dunbrack RL Jr, Hooft RW, Krieger B (2012) Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods Mol Biol* 819:405–421
71. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:27–38
72. Kryshchukovych A, Fidelis K, Moulton J (2013) CASP10 results compared to those of previous CASP experiments. *Proteins*. doi:10.1002/prot.24448
73. Freites JA, Choi Y, Tobias DJ (2003) Molecular dynamics simulations of a pulmonary surfactant protein B peptide in a lipid monolayer. *Biophys J* 84:2169–2180
74. Baoukina S, Tieleman DP (2010) Direct simulation of protein-mediated vesicle fusion: lung surfactant protein B. *Biophys J* 99:2134–2142
75. Nagle JF (1993) Area/lipid of bilayers from NMR. *Biophys J* 64:1476–1481
76. Biltonen RL, Lichtenberg D (1993) The use of differential scanning calorimetry as a tool to characterize liposome preparations. *Chem Phys Lipids* 64:129–142
77. Kranenburg M, Smit B (2005) Phase behavior of model lipid bilayers. *J Phys Chem B* 109:6553–6563
78. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 111:7812–7824
79. Tozzini V (2005) Coarse-grained models for proteins. *Curr Opin Struct Biol* 15:144–150
80. Bradley R, Radhakrishnan R (2013) Coarse-grained models for protein-cell membrane interactions. *Polymers* 5:890–936
81. Almlöf M, Brandsdal BO, Åqvist J (2004) Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J Comput Chem* 25:1242–1254
82. Bredow T, Jug K (2005) Theory and range of modern semiempirical molecular orbital methods. *Theor Chem Accounts* 113:1–14
83. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 48:1198–1229
84. Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* 11:560–566