Supplementary Information

Location of the Hydrophobic Surfactant Proteins, SP-B and SP-C, in Fluid-Phase Bilayers

Ryan W. Loney[1], Sergio Panzuela[2,3], Jespar Chen[4], Zimo Yang[4], Jonathan R. Fritz[4], Zackary Dell[4], Valentina Corradi[2], Kamlesh Kumar[1], D. Peter Tieleman[2], Stephen B. Hall[1*], Stephanie A. Tristram-Nagle[4*]

[1]Pulmonary and Critical Care Medicine, Oregon Health & Science University, Portland, Oregon 97239, USA
[2]Centre for Molecular Simulation and Department of Biological Sciences, University of Calgary, Calgary AB T2N 1N4, Canada
[3]Department of Theoretical Condensed Matter Physics, Universidad Autónoma de Madrid, E-28049, Madrid, Spain
[4]Biological Physics Group, Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
*Corresponding authors

Additional Methods.

MD Simulations. SP-B has been identified as a member of the saposin-like family of proteins. The homology of two amino acid sequences is commonly expressed in terms of their sequence identity, defined as the number of identical residues divided by the length for the shorter of the two compared sequences. The sequence identity among the saposin-like proteins is generally low.[1] For *bt* SP-B compared with the human saposins, values range from 20% (SapA) to 27% (SapC) (Fig. S1). The

presence, however, of conserved cysteine residues engaged in intra-monomeric disulfide bridges, along with the series of predicted amphipathic helices, has generally been accepted as defining a family of homologous proteins.[1] These criteria classify SP-B as a member of that family.

```
                        8 11                                35          46
bt_SP-B_P15781   FPIPIPYCWLCRTLIKRIQAVIP-----KGVLAMTVAQVCHVVPLL-VGGICQCLVERYS
hs_SP-B_P07988   FPIPLPYCWLCRALIKRIQAMIP-----KGALAVAVAQVCRVVPLV-AGGICQCLAERYS
hs_SAPB_P07602   ----GDVCQDCIQMVTDIQTAVRTNSTFVQALVEHVKEECDRLGPG-MADICKNYISQYS
hs_SAPA_P07602   ----SLPCDICKDVVTAAGDMLKDNAT-EEEILVYLEKTCDWLPKPNMSASCKEIVDSYL
hs_SAPC_P07602   ---SDVYCEVCEFLVKEVTKLIDNNKT-EKEILDAFDKMCSKLPKS-LSEECQEVVDTYG
hs_SAPD_P07602   ---DGGFCEVCKKLVGYLDRNLEKNST-KQEILAALEKGCSFLPDP-YQKQCDQFVAEYE
mm_SAPD_Q61207   ---NGGFCEVCKKLVLYLEHNLEKNST-KEEILAALEKGCSFLPDP-YQKQCDDFVAEYE
ss_NK-l_1NKL_pd  ----GYFCESCRKIIQKLEDMVGPQPN-EDTVTQAASQVCDKLKI--LRGLCKKIMRSFL
hs_GN-l_P22749   ---LGRDYRTCLTIVQKLKKMV-DKPT-QRSVSNAATRVCRTGRSR-WRDVCRNFMRRYQ
                          *  ::      :         :     . *          *      :
                                  71    77
bt_SP-B_P15781   VILLDTLLGRML-PQLVCGLVLRCSS--
hs_SP-B_P07988   VILLDTLLGRML-PQLVCRLVLRCSM--
hs_SAPB_P07602   EIAIQMMMHM-Q-PKEICALVGFCDEV-
hs_SAPA_P07602   PVILDIIKGEMSRPGEVCSALNLCESLQ
hs_SAPC_P07602   SSILSILLEEVS-PELVCSMLHLCSG--
hs_SAPD_P07602   PVLIEILVEVMD-PSFVCLKIGACPSAH
mm_SAPD_Q61207   PLLLEILVEVMD-PGFVCSKIGVCPSAY
ss_NK-l_1NKL_pd  RRISWDILTGKK-PQAICVDIKICKE--
hs_GN-l_P22749   SRVTQGLVAGET-AQQICEDLRLCIPST
                       :      .  :*  :  *
```

Figure S1. Sequence alignment between SP-B proteins and saposins (SapA-D) and saposin-like molecules (NK-l, NK-lysin; GN-l, GN-lysin). bt, B. taurus; hs, H. sapiens; mm, M. musculus; ss, S. scrofa. Conserved cysteine residues are involved in intra-monomer disulfide bridges (residues 8 and 77; 11 and 71; 35 and 46), while the cysteine residue of SP-B involved in the disulfide bridge between monomers (residue 48) is highlighted in red. The numbering refers to *bt* SP-B.

Experimental structures are available for: saposins A, B, C, and D from *H. sapiens*; saposin D from *M. musculus*; NK-lysin from *S. scrofa*; and GN-lysin from *H. sapiens*. These structures have defined two leaflets within these proteins. In the different solved structures, the leaflets exist either in an open, V-shaped configuration,

or a closed conformation. Our simulations used configurations of SP-B (Fig. S2) built on the following templates: open, human saponsin C (PDB ID 2Z9A); closed, human saposin D (PDB ID 2R3B).

For each system, 20 models were generated. The initial structure for simulation used the model with the lowest discrete optimized protein energy score (Figure S2). The models for SP-B included the conserved intramonomeric disulfide bonds, but not the bridge between monomers, for which guidance from experimental studies was unavailable. The structure of SP-C was modeled with two palmitoyl chains, added at

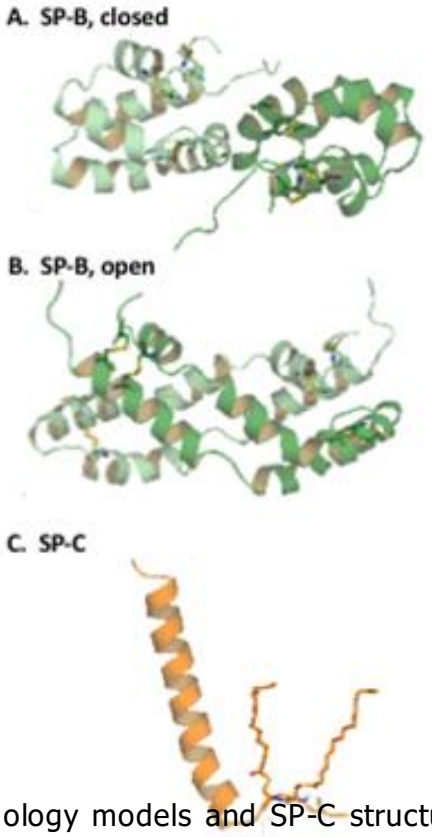

Figure S2. SP-B homology models and SP-C structure used in this study. SP-B was modeled: (A) in the closed conformation using the structure of human SapD (PDBID: 2RB3); (B) in the open conformation using human SapC (PDBID: 2Z9A). Disulfide bridges (shown as sticks) were modeled between residues 8 and 77; 11 and 71; 35 and 46. The structure of pig SP-C (C, PDBID: 1SPF) was modified by adding palmitoyl chains on cysteines 5 and 6.

cysteine 5 and 6 using CHARMM-GUI.

All simulations used either 384 (large systems) or 96 (small systems) central processing units. The systems were built with SP:DOPC in the x-y plane using a mean area per lipid 0.61-0.72 nm^2 and with systems between 7-22 nm laterally and 12-15 nm in the z direction. The following systems were created:

System 1 (large): 2 open SP-B (one protein per monolayer); 10 transmembrane SP-C; 1566 DOPC lipids (783/monolayer); 94,080 TIP3 water molecules; 50 Cl^- ions to neutralize SPs); l_x=l_y = 22-22.6 nm, l_z = ~15 nm. Lipid:protein molar ratio = 157:1.

System 2 (large): 2 closed SP-B (one/monolayer); 10 transmembrane SP-C; 1568 DOPC lipids (783/monolayer); 94,080 TIP3 water molecules; 50 Cl^- ions to neutralize SPs; l_x=l_y = 22-22.6 nm, l_z = ~15 nm. Lipid:protein molar ratio = 157:1.

System 3 (small): pure DOPC (A_L = 0.64 nm^2); 160 DOPC lipids (80/monolayer); 16,000 TIP3 water molecules; l_x=l_y=~7.2 nm, l_z = ~12 nm).

System 4 (small): pure DOPC (A_L = 0.68 nm^2); 160 DOPC lipids (80/monolayer); 16,000 TIP3 water molecules; l_x=l_y=~7.4 nm, l_z = ~12 nm).

System 5 (small): pure DOPC (A_L = 0.70 nm^2); 160 DOPC lipids (80/monolayer); 16,000 TIP3 water molecules; l_x=l_y=~7.6 nm, l_z = ~12 nm).

System 6 (small): 2 SP-C (one/monolayer) located in hydrocarbon; 160 DOPC lipids (80/monolayer); 15,502 TIP3 water molecules; 6 Cl^- ions to neutralize proteins; l_x=l_y=~7.4 nm, l_z =~12 nm. Lipid:protein molar ratio = 80:1.

System 7 (small): 2 transmembrane SP-C; 160 DOPC lipids (80/monolayer); 15,999 TIP3 water molecules; 6 Cl^- ions to neutralize proteins; $l_x=l_y=\sim7.3$ nm, $l_z=\sim12$ nm. Lipid:protein molar ratio = 80:1.

System 8 (small): 2 closed SP-B (one/monolayer); 160 DOPC lipids (80/monolayer); 15,958 TIP3 water molecules; 20 Cl^- ions to neutralize proteins; $l_x=l_y=\sim7.6$ nm, $l_z=\sim12$ nm. Lipid:protein molar ratio = 80:1.

System 9 (small): 2 open SP-B (one/monolayer); 160 DOPC lipids (80/monolayer); 15,854 TIP3 water molecules; 20 Cl^- ions to neutralize proteins; $l_x=l_y=\sim7.6$ nm, $l_z=\sim12$ nm. Lipid:protein molar ratio = 80:1.

Table S1. Fixed areas of small simulations

Simulation	Area per lipid ($Å^2$)
DOPC alone	71.7
Closed SP-B	72.8
Open SP-B	72.0
SP-C in hydrocarbon	68.4
Transmembrane SP-C	65.1
Two transmembrane SP-Cs	66.8

Additional Results:

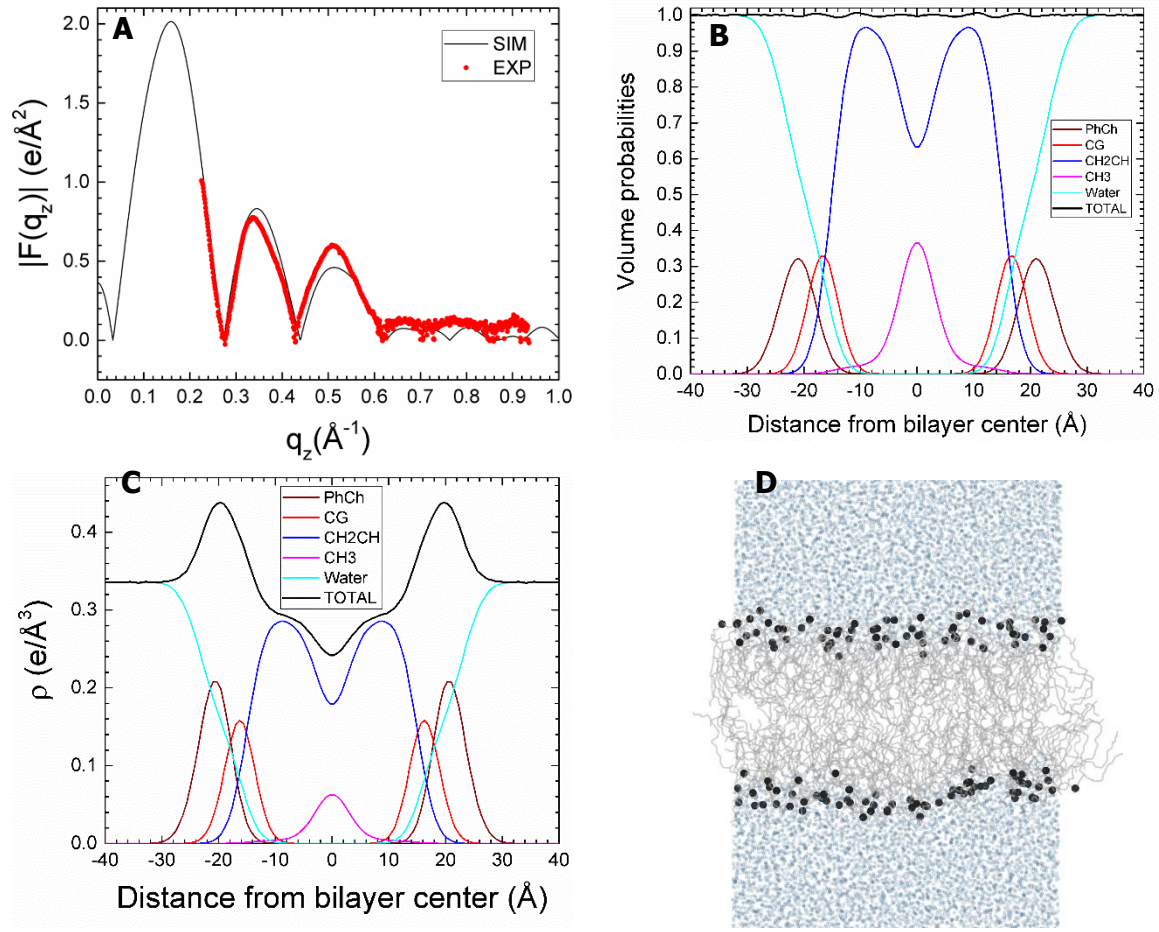

Figure S3. Structure of DOPC, without protein. Results from XDS data collected at 30°C, and from MD simulation constrained to area per lipid (A_L) = 72 Å^2. **A**. Form factors obtained from simulations (black traces) and experimental XDS (red symbols). **B**. Volume probabilities of components, obtained from simulation. **C**. Simulated electron density (ρ) profile for component groups: PhCh, phosphocholine; CG, carbonyl-glycerol; CH2CH, methylene-methine; CH3, terminal methyl. **D**. Snapshot of the system taken near the end of the 300 nsec simulation. The lipids are gray lines, phosphate atoms are black spheres, and water molecules are represented by light blue spheres.

References to SI

1. Bruhn, H. A short guided tour through functional and structural features of saposin-like proteins. *Biochem J* **2005,** *389*, 249-257.