JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

Cite this article: Vance SJ, McDonald RE, Cooper A, Smith BO, Kennedy MW. 2013 The structure of latherin, a surfactant allergen protein from horse sweat and saliva. J R Soc Interface 10: 20130453.
http://dx.doi.org/10.1098/rsif.2013.0453

Received: 17 May 2013
Accepted: 29 May 2013

Subject Areas:
biochemistry, biomaterials, biophysics

Keywords:
latherin, horse, sweat, surfactant protein, PLUNC proteins

Authors for correspondence:
Brian O. Smith
e-mail: brian.smith@glasgow.ac.uk
Malcolm W. Kennedy
e-mail: malcolm.kennedy@glasgow.ac.uk

[†]Present address: Department of Biochemistry, School of Biological Sciences, University of Cambridge, Cambridge CB2 1GA, UK.
[‡]Present address: Life Sciences Lead, Strategic Trade, UK Trade and Investment, 1 Victoria Street, London SW1H 0ET, UK.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2013.0453 or via http://rsif.royalsocietypublishing.org.

Royal Society Publishing
Informing the science of the future

The structure of latherin, a surfactant allergen protein from horse sweat and saliva

Steven J. Vance[1,†], Rhona E. McDonald[2,‡], Alan Cooper[1], Brian O. Smith[3] and Malcolm W. Kennedy[2,3]

[1]School of Chemistry, [2]Institute of Biodiversity, Animal Health and Comparative Medicine, and [3]Institute of Molecular, Cell and Systems Biology, University of Glasgow, Glasgow G12 8QQ, UK

Latherin is a highly surface-active allergen protein found in the sweat and saliva of horses and other equids. Its surfactant activity is intrinsic to the protein in its native form, and is manifest without associated lipids or glycosylation. Latherin probably functions as a wetting agent in evaporative cooling in horses, but it may also assist in mastication of fibrous food as well as inhibition of microbial biofilms. It is a member of the PLUNC family of proteins abundant in the oral cavity and saliva of mammals, one of which has also been shown to be a surfactant and capable of disrupting microbial biofilms. How these proteins work as surfactants while remaining soluble and cell membrane-compatible is not known. Nor have their structures previously been reported. We have used protein nuclear magnetic resonance spectroscopy to determine the conformation and dynamics of latherin in aqueous solution. The protein is a monomer in solution with a slightly curved cylindrical structure exhibiting a 'super-roll' motif comprising a four-stranded anti-parallel β-sheet and two opposing α-helices which twist along the long axis of the cylinder. One end of the molecule has prominent, flexible loops that contain a number of apolar amino acid side chains. This, together with previous biophysical observations, leads us to a plausible mechanism for surfactant activity in which the molecule is first localized to the non-polar interface via these loops, and then unfolds and flattens to expose its hydrophobic interior to the air or non-polar surface. Intrinsically surface-active proteins are relatively rare in nature, and this is the first structure of such a protein from mammals to be reported. Both its conformation and proposed method of action are different from other, non-mammalian surfactant proteins investigated so far.

1. Introduction

Surfactants occur widely in nature, typically involving small molecules such as bile acids, or the glycolipids and phospholipids that, in complex with small proteins, comprise the pulmonary surfactants of mammalian lungs. Proteins themselves rarely exhibit intrinsic surfactant activity except when misfolded or denatured, as commonly seen in laboratory preparations, in food products or in some fire retardant foams. Protein-based surfactants are of interest because they can be more efficient on a molar basis than small molecule surfactants yet still be compatible with cell membranes [1–4]. Examples of surfactant proteins that exhibit strong surface activity in their native state and in the absence of associated lipids or glycosylation include the hydrophobins of fungi [5], a protein found in the foam nests of certain amphibians (ranaspumin-2, RSN-2; [6]), and the subject of this report, the latherin protein of horses [3].

Latherin was originally described in the 1980s as an intrinsically surface-active, non-glycosylated protein that is abundant in horse sweat, and is the likely cause of the frothing seen in vigorously exercising animals [7]. It is

© 2013 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/3.0/, which permits unrestricted use, provided the original author and source are credited.

produced and stored in granules in the sweat glands of skin, and is also synthesized in the salivary glands [3]. Its function is thought to be to wet the surface of the waterproofed hair shafts of horses to allow rapid movement and spreading of perspired water over the surface of the pelt for evaporative cooling [3,7]. This idea is reinforced by the demonstration that latherin will coat hydrophobic surfaces [3]. Horses, humans and patas monkeys (*Erythrocebus patas*) are the only mammals known to sweat copiously for thermoregulation, though the compositions of their sweat fluids differ significantly—humans have high salt, low protein sweat, whereas horses have high protein, low salt sweat [8–11]; the composition of patas monkey sweat has not been reported although their eccrine sweat glands are physiologically and morphologically similar to those of humans [11]. For equids, the combined surfactant and surface coating activities of latherin may be a special adaptation associated with their role as large, endurance-running flight animals that have a particular need to shed heat [8,9], a process otherwise impeded by a dense, hairy pelt. Latherin's presence in equine saliva, where it may also cause foaming, is more puzzling, but its wetting properties could assist mastication of, and salivary enzyme penetration into, the dry, coarse, fibrous diet for which equids are specialized. Also, its surfactant activity might directly control the establishment and growth of microbial biofilms on tooth or mucosal surfaces [12].

Latherin is the target of IgE antibodies in some, but not all people who are allergic to horses [3], and its primary structure contains peptide sequences of two previously classified horse dander allergens (Equ c 4 and Equ c 5; [13]), which were presumably cleavage fragments of latherin. A latherin-like allergen protein has also been characterized from the tongue epithelium and salivary glands of cats (Fel d 8; [14]). Whether latherin and its relatives are intrinsically allergenic, or become targets of allergic responses in certain individuals responding coordinately to other allergenic stimuli, remains to be seen. Solving the structure of a family of allergens for which no previous structural information is available will potentially contribute to the continuing search for a relationship between allergenicity and protein structures, despite the seeming unreliability of such predictions [15,16].

Latherin is unusually rich in non-polar amino acids (predominantly leucines), and its amino acid sequence allies it to the PLUNCs (palate, lung, nasal epithelium clones), a large and enigmatic family of proteins of unknown structure present in the oral, nasal and upper respiratory tracts of mammals [17]. The biological function of these proteins is poorly understood, though they have been postulated to be involved in innate immunity at mucosal surfaces [18]. None of the PLUNCs have been shown to have bacteriolytic or bacteriostatic activities, but one from humans (short (S)PLUNC1; new systematic name BPIFA1) exhibits a leucine content of similar order to latherin [3,19], is similarly surface-active, and has anti-microbial biofilm activity [12]. PLUNCs have also been implicated in defence responses to mycoplasma infection, allergic inflammation, as well as in homeostasis of the upper airway and in protection of the middle ear, although their mechanisms of action remain to be defined [20–22]. This postulated connection between PLUNCs and innate immunity is stimulated by their amino acid sequence similarities to larger, two-domain proteins that are directly involved in anti-bacterial activities: lipopolysaccharide-binding protein and bactericidal/permeability-increasing (BPI) protein [23]. These two proteins are similar to cholesteryl ester transfer protein (CETP) and phospholipid transfer protein [24–26], so there is a precedent within the larger protein family for interaction with hydrophobic entities, but not necessarily involvement in immune defence.

Understanding how intrinsically active surfactant proteins work requires a multidisciplinary approach, an essential part of which must be the determination of their macromolecular structures in bulk solution and how they may change at the interface with air or other surfaces with which they associate. We report here the structure of horse latherin, as determined by high-resolution nuclear magnetic resonance spectroscopy (NMR) in solution, and postulate how the structure may explain its surfactant activity. The structure of latherin differs significantly from those observed in other surfactant protein systems (the hydrophobins and ranaspumins). Latherin's mechanism of surfactant action might be similar to the ranaspumins and at least one member of the PLUNC family, but dissimilar to that of the hydrophobins.

2. Material and methods

2.1. Protein preparation

Recombinant protein was expressed from a synthetic latherin gene where codon usage was optimized for expression in *Escherichia coli* (GeneArt, Invitrogen). The gene was incorporated into expression vector pET-32 (Novagen) to produce protein with enterokinase-cleavable, N-terminal His_6 and thioredoxin fusion tags. Expression was carried out in *E. coli* strain Tuner(DE3) (Novagen). Latherin was isolated from the soluble cell lysate by Ni-affinity chromatography, enterokinase cleavage, subtractive Ni-affinity chromatography and size-exclusion chromatography to yield pure protein (>95% by SDS-PAGE). Isotopically enriched latherin (^{15}N, ^{13}C) was prepared using M9 minimal medium incorporating $^{15}NH_4Cl$ and $^{13}C_6$-D-glucose as the sole nitrogen and carbon sources. For collection of residual dipolar couplings (RDCs), a ^{15}N-latherin sample was partially aligned by addition of filamentous phage Pf1 (Profos AG, Regensberg, Germany) at a final phage concentration of 5.0 mg ml^{-1} (10 Hz ^2H splitting).

2.2. NMR data collection and assignment of spectra

NMR resonance assignment of ^{15}N, ^{13}C labelled latherin is described in detail elsewhere [27]. All spectra were recorded at 310 K in 20 mM sodium phosphate, 50 mM NaCl, 1 mM NaN_3, pH 7.5 on a 14.1 T Bruker AVANCE spectrometer equipped with a Cryoprobe. Standard triple resonance experiments were supplemented with methyl-specific TOCSY experiments [28] to aid assignment of the high number of leucine residues. Spectra were processed using AZARA (Wayne Boucher, Department of Biochemistry, University of Cambridge, http://www.bio.cam.ac.uk/azara) and analysed using CcpNmr Analysis v. 2 [29].

2.3. Structure calculation

Nuclear Overhauser effect (NOE) restraints were created from three-dimensional ^{15}N-NOESY-HSQC and ^{13}C-edited ^1H,^1H spectra each with 100 ms mixing time. Distance restraints were derived from NOESY crosspeaks with the initial mapping from normalized intensity to distance following a $1/r^6$ relationship. NOE distance restraints were incorporated in restrained molecular dynamics calculations using the ambiguous distance restraints formalism [30]. Estimates of the average contribution of the dipolar coupling to J_{NH} (and the associated error) were obtained by collecting two independent IPAP-[^{15}N]-HSQC datasets from both isotropic and anisotropic samples. The magnitudes of the axial

and rhombic components of the alignment tensor were estimated using the method described by Clore *et al*. [31]. Eighty-eight D_{NH} restraints were incorporated into the structure calculations via the SANI potential [32] in square-well mode. Thirty hydrogen bond restraints were included for amide protons where signals were still observed in a ^{15}N-HSQC recorded 20 min after dilution of a ^{15}N-latherin sample into 90% (v/v) D_2O. Hydrogen-bond acceptors were identified by inspection of the NOE-refined structures, supported by NOE data. Restraints for the conserved disulfide bond were introduced once juxtaposition of the cysteine residues was observed in structure calculations.

Structures were calculated from randomized initial atomic coordinates using CNS [33] with the PARALLHDG-5.3 force field with PROLSQ non-bonded energy terms [34]. Initial structures were subsequently refined by iteratively filtering the ambiguous distance restraints against the calculated structures to discard duplicate restraints and assignments contributing less than one to five per cent to the total NOE intensity. RDC and hydrogen bond restraints were then introduced. ϕ, φ dihedral angle restraints for areas of regular secondary structure, produced using DANGLE (Cheung, University of Cambridge, http://dangle.sourceforge.net/; [35]) were included within initial stages of structure calculation to aid convergence and then omitted from final cooling steps. After eight rounds of NOE disambiguation using ARIA v. 2.3 [36], the 20 lowest energy models from a final round of 100 calculated structures were refined in explicit water. These 20 models were then used to create the representative ensemble of structures. The quality of these structures was analysed using PROCHECK [37] and their coordinates deposited in the Protein Data Bank (www.wwpdb.org) under accession code 3ZPM.

2.4. ^{15}N Relaxation measurements

N-relaxation rates, R1 and R2 were assessed using the method of Kay and co-workers [38–40] at a field strength of 600 MHz. Relaxation delays for assessment of R1 were 1200, 1600, 2100 and 2600 ms while those for R2 were 17, 34, 68, 102 and 136 ms. The first and third experiments in each series were repeated in order to estimate the inherent error in calculation of crosspeak intensities. Relaxation times T1 and T2 were calculated using nonlinear least-squares fitting. Collection of ^{15}N-HSQC-heteronuclear NOE experiments with and without saturation allowed extraction of $[^1H]^{15}N$ NOE values. Both saturation and reference experiments were repeated for the purpose of error estimation. The rotational correlation time, τ_m, for each amide residue was calculated using the method described by Kay *et al*. [39]. The rotational diffusion tensor of the latherin molecule was then calculated via the quadric representation approach proposed by Bruschweiler *et al*. [41] and Lee *et al*. [42] using the *quadric_diffusion* program (Palmer III, www.palmer.hs.columbia.edu/software.html). The model-free formalism as described by Lipari & Szabo [43,44] was used to determine the amplitudes and timescales of intramolecular motions of the latherin backbone from the three relaxation parameters. This analysis was carried out using the *FAST ModelFree* program (Loria, Yale University, http://xbeams.chem.yale.edu/~loria/software.php; [45]).

2.5. Hydrogen–deuterium exchange

Hydrogen–deuterium exchange rates were calculated by rapid dilution of a ^{15}N-latherin sample to 90% (v/v) D_2O. ^{15}N-HSQC spectra were collected at 20 min intervals for 3 h and then at 1 h intervals for a further 5 h (see examples in the electronic supplementary material, figure S1). Residues that displayed exchange on the timescale of the experiment were assigned as undergoing medium exchange. Residues that had reduced to 10 per cent of the original intensity (relative to a reference) within the 20 min required to record the first spectrum were assigned as undergoing fast exchange. Residues that displayed no change of intensity (relative to a reference) within the 8 h were assigned as undergoing slow exchange. Hydrogen–deuterium exchange rates (R_{H-D}) for medium exchanging residues were calculated by fitting peak intensity (I) against time (t) with $I_t/I_0 = \exp(-R_{H-D}.t) + c$.

Table 1. Experimental restraints and statistics of the calculated structures. Average statistics were calculated from the 20 water-refined structures in the latherin ensemble. The number of violations is shown as the average and standard deviation per structure.

NOE distance restraints	
NOE restraints	6503
ambiguous	2210
unambiguous	4293
intraresidue	1985
interresidue	2308
sequential ($i - j = 1$)	985
medium-range ($i - j < 5$)	518
long-range ($i - j > 5$)	805
violations per structure >0.5 Å	1.15
violations per structure >0.3 Å	8.60
distance restraint RMSD	0.036 Å
other restraints	
RDCs	88
RDC Q factor	0.127
hydrogen bonds	34
dihedral angle restraints	369
disulfide bond	1

3. Results

3.1. Latherin has a BPIF 'super-roll' structure in solution

Purified recombinant latherin, which has previously been found to exhibit properties in solution similar to the natural material [3,7], exhibited sufficiently sharp, well-resolved NMR spectra suitable for high-resolution structure determination after appropriate isotopic (^{13}C, ^{15}N) enrichment. Potential difficulties due to spectral overlap arising from the high proportion of leucine residues were overcome as described in §2 and [27]. A total of 6922 NOE-derived distance restraints were used to calculate the structure of latherin, of which 4293 were unambiguous or manually assigned, with 2210 ambiguous restraints in the final refinement (table 1). These were supplemented by 88 RDC restraints and 34 hydrogen bond restraints. The structure calculations converged well to give good agreement with the experimental data and a tightly defined ensemble of structures (see the electronic supplementary material, tables S1 and S2; figure 1*a*).

Latherin is monomeric in solution and exhibits an almost cylindrical structure about 65 Å long by 25 Å wide (figure 1*a*,*b*). These dimensions are in good agreement with the Stokes radius estimates for natural latherin yielding an axial ratio of approximately 3 : 1 [7]. Latherin comprises a four-stranded β-sheet against which two long anti-parallel helical regions pack, following the groove in the concave face of the β-sheet. The N- and C-termini are found at one

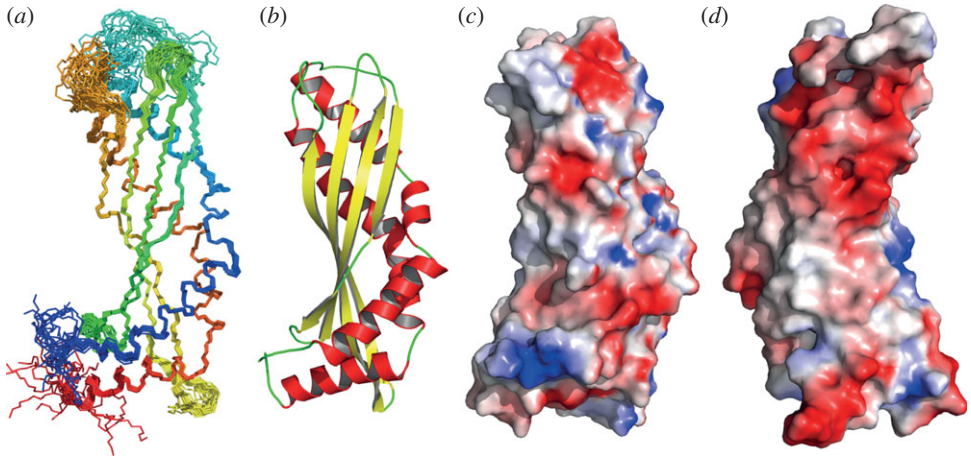


Figure 1. The solution structure of latherin. (*a*) The ensemble of the 20 latherin models (superimposed) that best fit the experimental data, shown in peptide backbone representation, shaded from blue (N-terminus) to red (C-terminus). (*b*) Ribbon model of the representative structure of latherin in solution illustrating secondary structure elements; α-helices are coloured red and β-strands yellow. (*c*) Surface contact potential (blue, positive; red, negative) of latherin mapped on the solvent accessible surface of the protein in the same orientation as (*a*), and (*d*) rotated 180°. Images and contact potential generated using PyMOL [46].

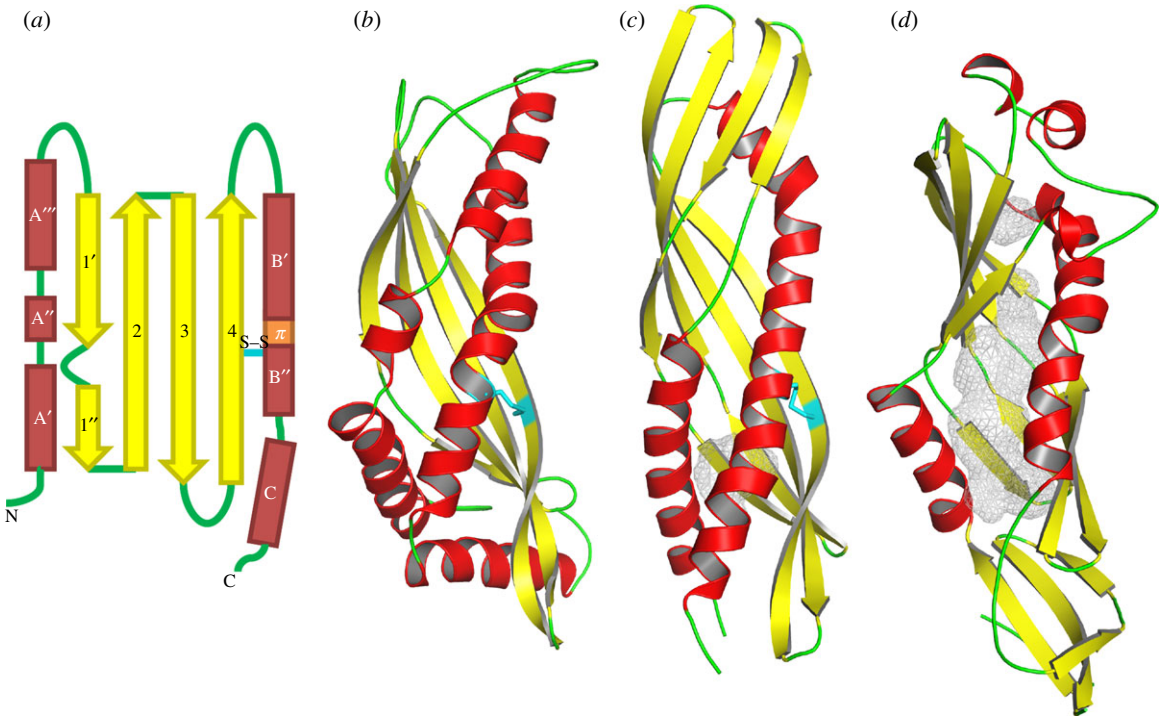


Figure 2. Topology of latherin and comparison with BPI. (*a*) Topology model of latherin. α-Helices are represented by red rectangles; β-strands by yellow arrows; non-regular secondary structure as green lines. The intramolecular disulfide bond, Cys133–Cys175, is shown as a cyan line labelled 'S–S'. The short section of π-helix is coloured orange, and the β-bulge by a curved green line between strands 1′ and 1″. (*b*) Cartoon representation of latherin compared with, (*c*) and (*d*), the N- and C-terminal domains, respectively, of BPI protein (PDB code 1BP1; [47]). The grey mesh in (*d*) encloses the internal cavities in the BPI C-terminal domain that are accessible to a 1.4 Å radius probe. Images created using PyMOL [46].

end of the cylinder (the 'terminal end'), where a third short helical region (helix α_C) is angled across the diameter of the molecule. The other end of the molecule (the 'loop end') is distinguished by three extensive flexible loops that are notable in their relatively high content of exposed apolar side chains, in particular leucines.

The N-terminal helix, labelled α_A, stretches from residues 7–47 with two breaks in the regular secondary structure at residues 22–23 and 31–33. Helix α_A can therefore be subdivided into three sections: α'_A (7–21), α''_A (24 – 30) and α'''_A (34 – 47) (see topology diagram in figure 2*a*). Helix α_B (152–203) is also interrupted such that α'_B (152 – 169) and α''_B (175 – 185) are separated by a short region of π-helix as indicated by the signature i to $i+5$ hydrogen bonding between the amides of residues 171–172 and the carbonyl groups of 166–167, respectively. Sections of π-helix can destabilize a helix and are often associated with functional sites [48–51]. Helix α_C comprises residues 188–203. The four strands that make up the anti-parallel β-sheet are β_1 (61–77), β_2 (83–97), β_3 (104–120) and β_4 (126–142). β_1 is interrupted by a β-bulge at a point where a proline (Pro89) in strand 2 introduces an irregularity in the packing of the two strands and does not present a hydrogen bonding partner to strand β1. It can therefore be divided into two sections β'_1 (61 – 70) and β''_1 (74 – 77). The disulfide bond (Cys133–Cys176) connects β_4 to α''_B.

The interior of the protein is notable for its paucity of polar side chains. The single exception is Thr110, which is closely surrounded by apolar aliphatic amino acids. Its position there, however, may be stabilized by hydrogen bonding between its hydroxyl to the main chain carbonyl of Leu136.

As expected from amino acid sequence comparisons [17,23], database examinations identify the fold adopted by latherin as a BPI domain-like fold (SCOP; [52], code 55393) or as a super-roll (CATH; [53], code 3.15). A search for other proteins with related folds using DALI [54] identified several structures with better than marginal match scores (z score >5). In addition to the structures of BPI and CETP, the search identified several similar structures from outwith the cognate BPI superfamily of proteins, including Der p 7, a dust mite allergen [55], juvenile hormone-binding protein (JHBP; [56]) and takeout protein 1 (Top1; [57]), all of which are involved in binding hydrophobic ligands, Aha1, which is apparently functionally unrelated, being an intracellular co-chaperone of the molecular chaperone, Hsp90 [58], and Yceb, which is an as yet uncharacterized lipoprotein from *E. coli* (protein structure database (PDB) code 3L6I). β-strands in latherin are shorter than those of the other proteins with a similar fold, with the result that the β-sheet at the 'loop end' does not twist as far around the helices. The single disulfide bond, which links the final strand of the β-sheet to the C-terminal helix in latherin, is a feature found in an analogous position in the N-terminal domains of BPI and CETP, as well as JHBP. Latherin is unique among the members of the superfamily in having two helical regions juxtaposed for the entire length of the concave face of the β-sheet.

3.2. Latherin is not obviously amphiphilic but has surface exposed hydrophobic residues at one end

The structure of latherin displays little evidence of any amphiphilicity that might have been anticipated by comparison with the distinct patches of polar and apolar side chains seen on the surface of hydrophobins [59–61]. By contrast, the exterior of latherin in the bulk phase, shows no such surface patches, being almost exclusively decorated with the side chains of hydrophilic residues and predominantly anionic due to the higher proportion of aspartic and glutamic acids over arginines, histidines and lysines (figure 1*c*,*d*). This is intriguing and initially unexpected, especially in view of the unusually high proportion of leucines in latherin (49 of the 208 residues), a trait common to human SPLUNC1, which also exhibits surfactant activity [12]. In the latherin structure, the leucines are evenly distributed along the length of the structure, being mainly confined to the interior in the ordered regions of the protein (figure 3*a*). But, at the loop end, about one-third of all the leucines are exposed to solvent (cf. figure 3*b* for loop end, and figure 3*c* for termini end; and see the electronic supplementary material, table S3 for numerical comparison). That these loop leucines and other adjacent aliphatic residues do not form an obvious hydrophobic patch is in part because their polar main-chain groups are solvent exposed and also because they are interspersed with polar residues.

We previously detected no interaction between latherin and 8-anilino-1-naphthalenesulfonic acid (ANS), a small fluorescent probe for exposed apolar regions or pockets in a protein [3], which might otherwise be expected of an amphiphilic molecule. Analysis of the latherin structure, however, revealed no evidence for cavities or pockets to which ANS might bind when in bulk solution.

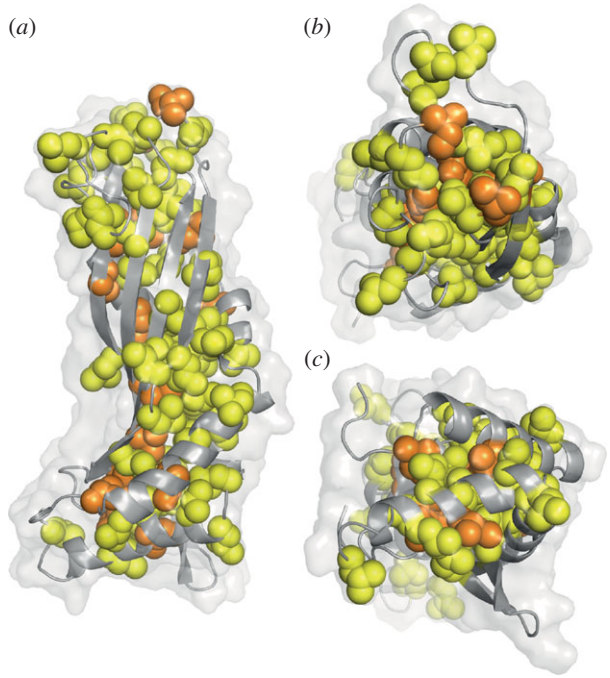


Figure 3. Distribution of leucine and isoleucine residues in the latherin structure. Latherin's main chain is displayed in cartoon representation, with the solvent accessible surface envelope shown in transparency. Leucine side chains are displayed as yellow spheres, and isoleucine side chains as orange spheres. In (*a*), the 'loop' end is at the top and the 'termini' end at the bottom. (*b*) and (*c*) show views from the 'loop' and 'termini' ends, respectively. Leucine and isoleucine residues predominantly line the core of the fold except at the 'loop' end. Image created using PYMOL [46].

3.3. Latherin is well ordered on the picosecond to nanosecond timescale with a few dynamic loops

The lack of obvious surface-exposed hydrophobic regions on latherin, its monomeric state in solution and our previous neutron reflection findings [3], suggest that a radical conformational change is required for latherin to facilitate surface tension reduction and association at an air : water or non-polar interface. This is likely to be reflected in features of the protein in the bulk phase observable as regions exhibiting unusual dynamic properties. The molecular dynamics of latherin in solution was therefore investigated using two methods: firstly, by examining the relaxation dynamics of its backbone amides for evidence of regions with high internal motion; secondly, by monitoring the rates of solvent exchange of labile hydrogens when dissolved in D_2O.

Latherin's backbone dynamics were analysed using the Lipari–Szabo model free approach based on amide ^{15}N relaxation measurements. The data were best fit with an overall correlation time of 11.3 ns and an axially symmetric diffusion tensor with a $D_{||}/D_{\perp}$ ratio of 1.68 to obtain order parameters, local correlation times and exchange broadening terms (figure 4), revealing that the four β-strands show low levels of internal motion, with the exception of the residues preceding (Gln66, Thr68 and Leu70) and within the β bulge (Leu71, Gln72 and Leu73), and those preceding and following the short loop between β_3 and β_4 at the terminal end. The two termini are themselves dynamic, as is the nearby loop (121–125), though the other loop in this region (residues

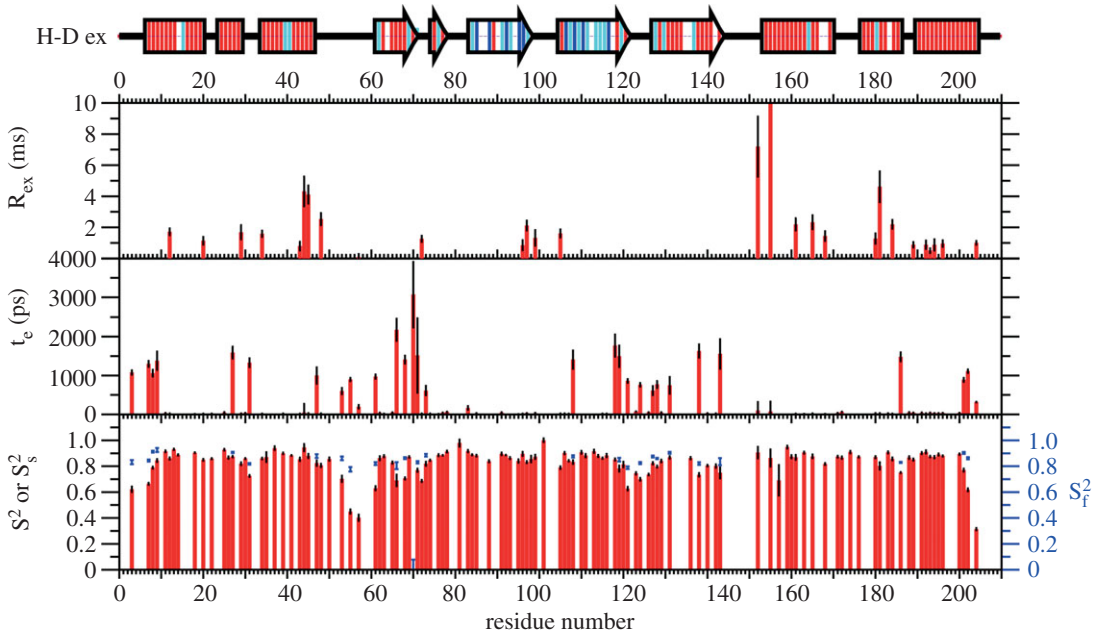


Figure 4. Latherin backbone dynamics. The Lipari–Szabo extended model-free parameters derived from ^{15}N relaxation measurements at 14.1 T are plotted by residue number in the three graphs. Backbone amide hydrogen–deuterium exchange rates indicated on the secondary structure schematic above, with residues undergoing fast exchange (lifetime <20 min) coloured red, medium exchange (20–480 minutes) cyan and slow exchange (>480 min) blue in the secondary structure cartoon.

78–82) shows relatively little internal motion. At the other end of the molecule, the loop between α4 and β1 (48–60) exhibits depressed order parameters indicating a high degree of flexibility that correlates with the poor definition of this part of the structure. The shorter loop between β2 and β3 (98–103), in contrast, displays motion on the millisecond timescale for the residues for which relaxation data can be obtained. The long loop between β_4 and α'_B (143–151) could not be examined directly because of the absence of amide cross-peaks for residues in this region (although Gln143 at one end of the loop was seen to be dynamic), which may itself be evidence of a substantial degree of motion.

Hydrogen–deuterium change experiments also revealed significant differences between various parts of the molecule. Notably, all of the residues showing slow rates of exchange, and most of those showing medium rates, occur in the two central strands β2 and β3, consistent with the relaxation dynamics data. The remainder of the molecule, including the three main helical regions, undergoes fast exchange, with only a few isolated regions at the centre of each α-helical segment exhibiting lower rates of exchange. The hydrogen–deuterium exchange results therefore agree well with the relaxation dynamics analysis, further demonstrating the stability of the two central strands of the β-sheet.

In summary, latherin is a molecule with a more rigid core and two dynamic ends, the loop end in particular having three loops which show high levels of chemical exchange.

4. Discussion

4.1. Latherin's structure and possible conformation change at an interface

The three-dimensional molecular structure of latherin in its aqueous solution phase described here is the first such structure reported for an intrinsically surfactant protein from mammals. This new information promises a broader understanding of how proteins can exhibit strong surfactant properties in their native states without the involvement of other cofactors such as lipids or glycans. As we and others have shown, naturally occurring surfactant proteins such as hydrophobins, ranaspumins and latherin exhibit significant surfactant properties at concentrations several orders of magnitude lower than usually observed with other proteins. Moreover, as reviewed in [4], this surfactant activity is directly related to structure, either because of amphiphilicity (as in the case of hydrophobins) or clam-shell/hinge-opening (as postulated for RSN-2), where the native conformation predisposes such proteins to biologically significant surface interaction. This is clearly distinct from that which is observed in the non-specific interfacial unfolding of other proteins, which usually takes much longer to appear and often requires quite aggressive denaturation treatment depending on the protein. The overall structure and molecular surface properties of the latherin molecule are radically different from those of other surfactant proteins known to date, and suggest yet another means of achieving spontaneous protein surfactant activity.

As we have observed elsewhere [6], the structure of a surfactant protein in bulk solution does not necessarily reflect its disposition at the air : water interface. Indeed, for monomeric proteins in solution, conformational change at the interface would seem to be a requirement in order to reconcile the need for good aqueous solubility in the bulk phase, while presenting a more amphipathic appearance at an interface. Such a radical conformational change also seems to be required for latherin. Our previous neutron reflection data indicate that latherin forms a relatively thin (mono)layer approximately 10 Å deep at the air : water surface, and with an area of approximately 4350 Å^2 per molecule [3]. Interestingly, neutron reflection studies show that non-specific irreversible interfacial layers sometimes observed with other proteins are much thicker in comparison, typically of order 30 Å [62–65]. The latherin cylinder in the bulk phase is about 75 Å long by about 25 Å in diameter, which is incompatible with a 10 Å layer in its fold in the bulk phase. Complete unfolding and flattening of

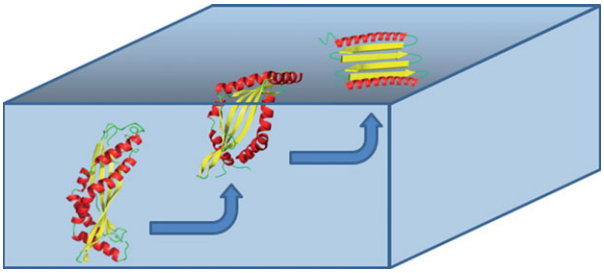


Figure 5. Latherin unfolding at an air : water interface. Speculative model of how latherin may transform from its fold in the bulk phase to an opened-out conformation at an air : water interface, thereby exposing its apolar interior to the air. The model shows three stages, from left to right: latherin in the bulk phase in which recognition of the interface occurs via the relatively hydrophobic loops; initial unzipping of the two α-helices initiated from the 'loop' end and a final open, planar, conformation, retaining secondary structure but with the hydrophobic core exposed at the interface. There will likely be dynamic exchange between the three conformations. A similar process may apply for latherin associating with a hydrophobic solid surface.

a cylinder of these dimensions yields an area of approximately 5890 Å^2, which, while accepting the crudity of this approximation, is compatible with the value obtained from neutron reflection. How latherin initially associates with an interface, and the events that follow, remain unknown, but its structure and dynamics provide both clues and topological constraints. The dynamic, unstructured, apolar side chain-rich loops are the most likely place where the protein could associate, penetrate and anchor to a surface, and the loops would have sufficient flexibility to then splay out with apolar side chains oriented towards the air or a non-polar solid substrate. Subsequent unzippering of the protein cylinder is unlikely to occur between any of the β strands because of the cooperative hydrogen bonding between them. This constraint is reinforced by the hydrogen–deuterium exchange data that identifies the inter-strand H-bonds (in particular, between strands 2 and 3) as the most stable in the molecule. Unstitching between strand 4 and helix B is unlikely given the disulfide bond that connects them approximately midway down their lengths. Assuming minimal change in the secondary structure elements, this leaves the seams between the two helices, or between helix A and strand 1 as likely fault lines. Given that the solvent-excluded interfacial area buried between the two helices is approximately 2000 Å^2, and that between helix A and strand 1 is approximately 1300 Å^2, the latter case appears to be the more favourable. Conversely, assuming that unfolding initiates from the apolar loops, as proposed above, the ability of helix A and strand 1 to reorient in an independent manner is likely to be inhibited by the loop that connects the two features at this end of the protein. By contrast, there is no such topological constraint on the relative orientation of the two helices. A possible unfolding sequence involving an opening between the helices is illustrated in figure 5.

4.2. Latherin operates as a surfactant differently from hydrophobins

The mechanism by which latherin operates as a surfactant is clearly different to that of the hydrophobins. These are rigid, amphiphilic molecules with a distinct surface patch of apolar amino acid side chains on the face of each molecule [5,60,61]. In solution, they form oligomers in which the hydrophobic patches are isolated from solvent water, allowing the proteins to remain in solution as dimers or tetramers [5]. At an air : water interface, hydrophobins orientate with the apolar surfaces projecting into air, whereas the polar regions remain immersed in the water phase. Charge interactions on the flanks of adjacent hydrophobin molecules yields self-associating monolayers [5] without the necessity for any conformational change. Latherin, in contrast, remains monomeric in solution, and appears more akin to the frog foam nest protein, RSN-2 in this respect (though with a quite different fold) and also appears to undergo significant conformational change at an air : water interface [6].

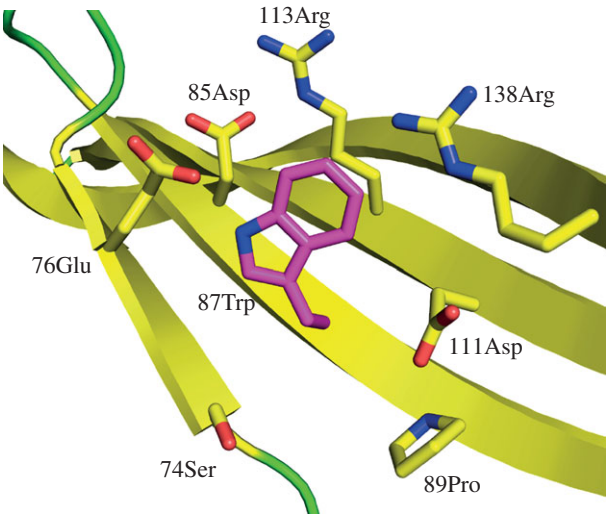


Figure 6. The environment of the solvent exposed tryptophan. Trp87 and surrounding side chains are shown in stick representation, with side-chain oxygens and nitrogens coloured red and blue, respectively. Image created using PyMOL [46].

4.3. Unusual environment of latherin's single tryptophan side chain

One puzzle presented by our previous work on latherin relates to the fluorescence properties of its single tryptophan residue (Trp87 in the structure; [3]). Latherin in dilute aqueous solution exhibits a relatively red-shifted Trp fluorescence emission spectrum, usually indicative of exposure to solvent water or a charged local protein environment [66,67]. Quenching of fluorescence emission by Trp87 by neutrally charged compounds (succinimide, acrylamide) was efficient, consistent with side-chain exposure to polar solvent water [3]. But, quenching by iodide (I^-), normally a highly efficient quenching agent for exposed Trp residues, was unexpectedly ineffective [3]. The new structure explains this conundrum in that Trp87 is exposed on the exterior of the protein, midway down the concave side of the curved cylinder, with its indole side chain sandwiched between, and encircled by, nearby charged amino acid side chains (Asp85, Asp111, Arg113, Arg138 and Glu76; figure 6). So, Trp87 is in a position to encounter solvent water and be quenched by neutral compounds, but its local environment is sufficiently dominated by negatively charged groups to repel a normally highly efficient but negatively charged quencher.

4.4. Structural similarities and differences between latherin and its relatives

Although relatively uncommon, the latherin super-roll fold has been observed elsewhere, most notably as part of much

larger proteins such as in domains of the BPI and CETPs of humans [24,68]. Both of these proteins have two latherin-like domains, the N-terminal domain of each being similar to latherin, with the C-terminal domains appearing to have diverged in structure (see figure 2 and electronic supplementary material, S2). This similarity is nicely illustrated in a superposition of latherin and BPI's C-terminal domain (not shown). Interestingly, the short section of π-helix in the α'_B helical region of latherin is a feature shared in the C-terminal helices of both domains of BPI and CETP, and π-helices have been proposed to have some role in conformational exchange associated with function in CETP [51].

In contrast to the ligand-binding cavities seen in BPI and CETP, the close packing of large, apolar amino acid side chains in the hydrophobic core of latherin leaves no internal cavity. BPI is, like latherin, a long, slightly curved cylinder, of similar diameter to latherin but about twice as long. It has two latherin-like domains that are fused closely end-to-end, with some of their chains intertwining, forming a single piece, boomerang-shaped molecule. Each of the domains exhibits latherin-like folds, the N-terminal domain particularly so. These proteins interact with lipids, and CETP is known to have a cavity into or through which lipids may move [51]. So, it is conceivable that both BPI and CETP arose from domain duplications of latherin- and PLUNC-like ancestors, followed by specialization and structural alteration of one (BPI) or both (CETP) domains. Their presumptive descent from a common ancestor protein is also indicated by similarities in the intron positions in their encoding genes [17].

Perhaps, the most interesting comparison would be between the structure of latherin and the PLUNCs, which are, like latherin, single domain members of the BPI superfamily. One of these from humans (BPIFA1; SwissProt Q9NP55) is highly expressed in the trachea, progressively less so from proximal (bronchial) to distal (bronchiolar) airways, and, as latherin, has an unusually high content of leucines and exhibits surfactant activity [12,19]. No structure for a PLUNC is yet available, so we attempted here to model BPIFA1 using latherin as a template. Despite the familial affiliation, a simple alignment of the two sequences illustrates how divergent the two are, and latherin exhibits several amino acid position deletions relative to human BPIFA1 (see the electronic supplementary material, figure S3). Nevertheless, a reasonably acceptable model could be created which is similar in overall structure to that predicted previously using the X-ray crystal structure of the much larger BPI (PDB, accession 1BP1) as template [17]. The current model suggests that this PLUNC may share with latherin the long unstructured loop regions at the end of the molecule equivalent to latherin's loop end, together with a similar distribution and concentration of leucine residues (see the electronic supplementary material, figure S4A). With regards to the mechanism of its surface activity, the model of BPIFA1, as with our empirical structure for latherin, shows no sign of patches of hydrophobic or charged amino acid side chains exposed on its surface (see the electronic supplementary material, figure S4B). The corresponding regions of several members of the PLUNC/BPI family have been identified as key motifs in their ability to bind the target lipid [26,69,70]. The proposition that these loops are involved in surface detection in latherin may infer a similar mode of substrate recognition throughout the BPIF superfamily, and we cannot at this stage rule out the possibility that latherin may interact with lipids.

4.5. Other 'super-roll' proteins: common ancestry or convergent evolution?

A wider search for non-mammalian proteins with a similar fold to latherin yielded several structures. These derive from a wide range of eukaryotic taxa, and a pertinent question would be whether they represent true descent from an ancient common ancestor protein, or cases of convergent evolution. These proteins are found in insects (a juvenile hormone-binding protein, and a lipid-binding takeout protein in lepidopterans [57]); arachnids (the Der p 7 allergen of a house dust mite [55]) and in an apicomplexan, the malaria parasite *Plasmodium falciparum* (AHA-1; activator of Hsp90 ATPase from yeast; [58]; see the electronic supplementary material, figure S2 for comparisons). These show distinct structural similarities to the BPI/CETP/PLUNC/latherin (BPIFA) family of mammals, despite highly divergent amino acid sequences, so the question arises as to whether they represent true descendents of an ancestral protein, or a case of convergent evolution. The argument for descent from a common ancestor is most convincing for the mammalian and insect proteins, whose structures most closely resemble that of latherin in that they have intramolecular disulfide bonds in similar positions, and their encoding genes have similar arrangements of introns [17,23]. Der p 7 and AHA-1 resemble latherin the least, so may represent independent evolution of a latherin-like super-roll fold, but there is as yet insufficient phylogenetic and protein bioinformatic information of these proteins in such diverse taxa to be sure that missing links do not exist.

5. Conclusions

The structure of latherin confirms that it is a member of a family of proteins in mammals with similarities in their structures, but a remarkable divergence in their biological functions. Its closest relatives are synthesized in the salivary glands, oral cavity and associated structures, so latherin possibly evolved from a PLUNC-like ancestor as a specialization in equids for processing dry, fibrous dietary materials, and/or to control microbial biofilms on teeth and mucosal surfaces. While continuing to perform such functions, it may then have been recruited to the skin as equids evolved into large-bodied flight animals capable of sustained exercise requiring rapid onset and efficient heat dissipation. Latherin represents the first intrinsically surfactant protein of mammals whose structure is known, but, more, it reveals a potential mechanism of action that has not been demonstrated before for an animal protein in its native state, with the exception of the RSN-2 frog foam nest protein [6], but clearly different from that of other classes of surfactant proteins whose structures are known [5,61]. Consequently, this is of general interest across a broad range of disciplines including not only protein structural biology and biophysics, but also having potential implications in veterinary science, human health and bio- and nano-technologies involving protein–surface interactions.

No animals or animal products were used other than minor ingredients of bacterial culture media.

This study was supported by grants to M.W.K., A.C. and B.O.S. from the Wellcome Trust (GR070994MA). S.J.V. was supported by a studentship from the Biotechnology and Biological Sciences Research Council.

References

1. Cooper A, Kennedy MW, Fleming RI, Wilson EH, Videler H, Wokosin DL, Su TJ, Green RJ, Lu JR. 2005 Adsorption of frog foam nest proteins at the air–water interface. *Biophys. J.* **88**, 2114–2125. (doi:10.1529/biophysj.104.046268)
2. Fleming RI, Mackenzie CD, Cooper A, Kennedy MW. 2009 Foam nest components of the túngara frog: a cocktail of proteins conferring physical and biological resilience. *Proc. R. Soc. B* **276**, 1787–1795. (doi:10.1098/rspb.2008.1939)
3. McDonald RE, Fleming RI, Beeley JG, Bovell DL, Lu JR, Zhao X, Cooper A, Kennedy MW. 2009 Latherin: a surfactant protein of horse sweat and saliva. *PLoS ONE* **4**, e5726. (doi:10.1371/journal.pone.0005726)
4. Cooper A, Kennedy MW. 2010 Biofoams and natural protein surfactants. *Biophys. Chem.* **151**, 96–104. (doi:10.1016/j.bpc.2010.06.006)
5. Linder MB. 2009 Hydrophobins: proteins that self assemble at interfaces. *Curr. Opin. Colloid Interface Sci.* **14**, 356–363. (doi:10.1016/j.cocis.2009.04.001)
6. Mackenzie CD, Smith BO, Meister A, Blume A, Zhao X, Lu JR, Kennedy MW, Cooper A. 2009 Ranaspumin-2: structure and function of a surfactant protein from the foam nests of a tropical frog. *Biophys. J.* **96**, 4984–4992. (doi:10.1016/j.bpj.2009.03.044)
7. Beeley JG, Eason R, Snow DH. 1986 Isolation and characterization of Latherin, a surface-active protein from horse sweat. *Biochem. J.* **235**, 645–650.
8. Jenkinson DM, Elder HY, Bovell DL. 2006 Equine sweating and anhidrosis. I. Equine sweating. *Vet. Dermatol.* **17**, 361–392. (doi:10.1111/j.1365-3164.2006.00545.x)
9. Jenkinson DM, Elder HY, Bovell DL. 2007 Equine sweating and anhidrosis. II. Anhidrosis. *Vet. Dermatol.* **18**, 2–11. (doi:10.1111/j.1365-3164.2007.00571.x)
10. Cohn BA. 1998 The vital role of the skin in human natural history. *Int. J. Dermatol.* **37**, 821–824. (doi:10.1046/j.1365-4362.1998.00575.x)
11. Gisolfi CV, Sato K, Wall PT, Sato F. 1982 *In vivo* and *in vitro* characteristics of eccrine sweating in patas and rhesus monkeys. *J. Appl. Physiol.* **53**, 425–431.
12. Gakhar L, Bartlett JA, Penterman J, Mizrachi D, Singh PK, Mallampalli RK, Ramaswamy S, McCray Jr PB. 2010 PLUNC is a novel airway surfactant protein with anti-biofilm activity. *PLoS ONE* **5**, e9098. (doi:10.1371/journal.pone.0009098)
13. Botros HG, Poncet P, Rabillon J, Fontaine T, Laval JM, David B. 2001 Biochemical characterization and surfactant properties of horse allergens. *Eur. J. Biochem.* **268**, 3126–3136. (doi:10.1046/j.1432-1327.2001.02217.x)
14. Smith W, O'Neil SE, Hales BJ, Chai TLY, Hazell LA, Tanyaratsrisakul S, Piboonpocanum S, Thomas WR. 2011 Two newly identified cat allergens: the von Ebner gland protein Fel d 7 and the latherin-like protein Fel d 8. *Int. Arch. Allergy Immunol.* **156**, 159–170. (doi:10.1159/000322879)
15. Aalberse RC, Stadler BM. 2006 *In silico* predictability of allergenicity: from amino acid sequence via 3-D structure to allergenicity. *Mol. Nutr. Food Res.* **50**, 625–627. (doi:10.1002/mnfr.200500270)
16. Aalberse RC, Crameri R. 2011 IgE-binding epitopes: a reappraisal. *Allergy* **66**, 1261–1274. (doi:10.1111/j.1398-9995.2011.02656.x)
17. Bingle CD, Leclair EE, Havard S, Bingle L, Gillingham P, Craven CJ. 2004 Phylogenetic and evolutionary analysis of the PLUNC gene family. *Protein Sci.* **13**, 422–430. (doi:10.1110/ps.03332704)
18. Bingle CD, Craven CJ. 2002 PLUNC: a novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. *Hum. Mol. Genet.* **11**, 937–943. (doi:10.1093/hmg/11.8.937)
19. Kennedy MW. 2011 Latherin and other biocompatible surfactant proteins. *Biochem. Soc. Trans.* **39**, 1017–1022. (doi:10.1042/BST0391017)
20. Chu HW *et al*. 2007 Function and regulation of SPLUNC1 protein in mycoplasma infection and allergic inflammation. *J. Immunol.* **179**, 3995–4002.
21. Gally F *et al*. 2011 SPLUNC1 Promotes lung innate defense against *Mycoplasma pneumoniae* infection in mice. *Am. J. Pathol.* **178**, 2159–2167. (doi:10.1016/j.ajpath.2011.01.026)
22. McGillivary G, Bakaletz LO. 2010 The multifunctional host defense peptide SPLUNC1 is critical for homeostasis of the mammalian upper airway. *PLoS ONE* **5**, e13224. (doi:10.1371/journal.pone.0013224)
23. Bingle CD, Craven CJ. 2004 Meet the relatives: a family of BPI- and LBP-related proteins. *Trends Immunol.* **25**, 53–55. (doi:10.1016/j.it.2003.11.007)
24. Liu SP, Mistry A, Reynolds JM, Lloyd DB, Griffor MC, Perry DA, Ruggeri RB, Clark RW, Qiu XY. 2012 Crystal structures of cholesteryl ester transfer protein in complex with inhibitors. *J. Biol. Chem.* **287**, 37 321–37 329. (doi:10.1074/jbc.M112.380063)
25. Lagrost L, Desrumaux C, Masson D, Deckert V, Gambert P. 1998 Structure and function of the plasma phospholipid transfer protein. *Curr. Opin. Lipidol.* **9**, 203–209. (doi:10.1097/00041433-199806000-00004)
26. Desrumaux C, Labeur C, Verhee A, Tavernier J, Vandekerckhove J, Rosseneu M, Peelman F. 2001 A hydrophobic cluster at the surface of the human plasma phospholipid transfer protein is critical for activity on high density lipoproteins. *J. Biol. Chem.* **276**, 5908–5915. (doi:10.1074/jbc.M008420200)
27. Vance SJ, McDonald RE, Cooper A, Kennedy MW, Smith BO. 2013 Resonance assignments for latherin, a natural surfactant protein from horse sweat. *Biomol. NMR Assign.* (doi:10.1007/s12104-013-9485-3)
28. Uhrin D, Uhrinova S, Leadbeater C, Nairn J, Price NC, Barlow PN. 2000 3D HCCH3-TOCSY for resonance assignment of methyl-containing side chains in C-13-labeled proteins. *J. Magn. Reson.* **142**, 288–293. (doi:10.1006/jmre.1999.1951)
29. Vranken WF *et al*. 2005 The CCPN Data model for NMR spectroscopy: developement of a software pipeline. *Proteins* **59**, 687–696. (doi:10.1002/prot.20449)
30. Nilges M. 1995 Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660. (doi:10.1006/jmbi.1994.0053)
31. Clore GM, Gronenborn AM, Bax A. 1998 A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural Information. *J. Magn. Reson.* **133**, 216–221. (doi:10.1006/jmre.1998.1419)
32. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A. 1997 Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat. Struct. Mol. Biol.* **4**, 732–738. (doi:10.1038/nsb0997-732)
33. Brunger AT *et al.* 1998 Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921.
34. Linge JP, Nilges M. 1999 Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J. Biomol. NMR* **13**, 51–59. (doi:10.1023/A:1008365802830)
35. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW. 2010 DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Magn. Reson.* **202**, 223–233. (doi:10.1016/j.jmr.2009.11.008)
36. Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M. 2007 ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381–382. (doi:10.1093/bioinformatics/btl589)
37. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. 1996 AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486. (doi:10.1007/BF00228148)
38. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. 1994 A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**, 727–734. (doi:10.1007/BF00404280)
39. Kay LE, Torchia DA, Bax A. 1989 Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979. (doi:10.1021/bi00449a003)
40. Kay LE, Xu GY, Yamazaki T. 1994 Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. Mag. Reson. A* **109**, 129–133. (doi:10.1006/jmra.1994.1145)

41. Bruschweiler R, Liao X, Wright PE. 1995 Long-range motional restrictions in a multidomain zinc-finger protein from anisotropic tumbling. *Science* **268**, 886–889. (doi:10.1126/science.7754375)
42. Lee LK, Rance M, Chazin WJ, Palmer AG. 1997 Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation. *J. Biomol. NMR* **9**, 287–298. (doi:10.1023/A:1018631009583)
43. Lipari G, Szabo A. 1982 Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity. *J. Am. Chem. Soc.* **104**, 4546–4559. (doi:10.1021/ja00381a009)
44. Lipari G, Szabo A. 1982 Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. II. Analysis of experimental results. *J. Am. Chem. Soc.* **104**, 4559–4570. (doi:10.1021/ja00381a010)
45. Cole R, Loria JP. 2003 FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J. Biomol. NMR* **26**, 203–213. (doi:10.1023/A:1023808801134)
46. The PyMOL Molecular Graphics System. 1.3 ed. Schrodinger, LLC. See www.pymol.org.
47. Beamer LJ, Carroll SF, Eisenberg D. 1997 Crystal structure of human BPI and two bound phospholipids at 2.4 angstrom resolution. *Science* **276**, 1861–1864. (doi:10.1126/science.276.5320.1861)
48. Cartailler JP, Luecke H. 2004 Structural and functional characterization of π bulges and other short intrahelical deformations. *Structure* **12**, 133–144. (doi:10.1016/j.str.2003.12.001)
49. Cooley RB, Arp DJ, Karplus PA. 2010 Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. *J. Mol. Biol.* **404**, 232–246. (doi:10.1016/j.jmb.2010.09.034)
50. Weaver TM. 2000 The π-helix translates structure into function. *Protein Sci.* **9**, 201–206. (doi:10.1110/ps.9.1.201)
51. Hall J, Qiu X. 2011 Structural and biophysical insight into cholesteryl ester-transfer protein. *Biochem. Soc. Trans.* **39**, 1000–1005. (doi:10.1042/BST0391000)
52. Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995 SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540. (doi:10.1016/S0022-2836(05)80134-2)
53. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton J, Orengo CA. 2009 The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res.* **37**, D310–D314. (doi:10.1093/nar/gkn877)
54. Holm L, Rosenstrom P. 2010 Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549. (doi:10.1093/nar/gkq366)
55. Mueller GA, Edwards LL, Aloor JJ, Fessler MB, Glesner J, Pomes A, Chapman MD, London RE, Pedersen LC. 2010 The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. *J. Allergy Clin. Immunol.* **125**, 909–917. (doi:10.1016/j.jaci.2009.12.016)
56. Kolodziejczyk R, Bujacz G, Jakob M, Ozyhar A, Jaskolski M, Kochman M. 2008 Insect juvenile hormone binding protein shows ancestral fold present in human lipid-binding proteins. *J. Mol. Biol.* **377**, 870–881. (doi:10.1016/j.jmb.2008.01.026)
57. Hamiaux C, Stanley D, Greenwood DR, Baker EN, Newcomb RD. 2009 Crystal structure of *Epiphyas postvittana* takeout 1 with bound ubiquinone supports a role as ligand carriers for takeout proteins in insects. *J. Biol. Chem.* **284**, 3496–3503. (doi:10.1074/jbc.M807467200)
58. Meyer P *et al*. 2004 Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. *EMBO J.* **23**, 511–519. (doi:10.1038/sj.emboj.7600060)
59. Hakanpaa J, Linder M, Popov A, Schmidt A, Rouvinen J. 2006 Hydrophobin HFBII in detail: ultrahigh-resolution structure at 0.75 angstrom. *Acta Crystallogr. D* **62**, 356–367. (doi:10.1107/S0907444906000862)
60. Hakanpaa J, Paananen A, Askolin S, Nakari-Setala T, Parkkinen T, Penttila M, Linder MB, Rouvinen J. 2004 Atomic resolution structure of the HFBII hydrophobin, a self-assembling amphiphile. *J. Biol. Chem.* **279**, 534–539. (doi:10.1074/jbc.M309650200)
61. Hakanpaa J, Szilvay GR, Kaljunen H, Maksimainen M, Linder M, Rouvinen J. 2006 Two crystal structures of *Trichoderma reesei* hydrophobin HFBI: the structure of a protein amphiphile with and without detergent interaction. *Protein Sci.* **15**, 2129–2140. (doi:10.1110/ps.062326706)
62. Lu JR, Su TJ, Penfold J. 1999 Adsorption of serum albumins at the air/water interface. *Langmuir* **15**, 6975–6983. (doi:10.1021/la990131h)
63. Lu JR, Su TJ, Thomas RK. 1999 Structural conformation of bovine serum albumin layers at the air–water interface studied by neutron reflection. *J. Colloid Interface Sci.* **213**, 426–437. (doi:10.1006/jcis.1999.6157)
64. Lu JR, Su TJ, Thomas RK, Penfold J, Webster J. 1998 Structural conformation of lysozyme layers at the air/water interface studied by neutron reflection. *J. Chem. Soc. Faraday Trans.* **94**, 3279–3287. (doi:10.1039/a805731a)
65. Zhao X, Pan F, Lu JR. 2009 Interfacial assembly of proteins and peptides: recent examples studied by neutron reflection. *J. R. Soc. Interface* **6**, S659–S670. (doi:10.1098/rsif.2009.0168.focus)
66. Eftink MR, Ghiron CA. 1976 Exposure of tryptophanyl residues in proteins: quantitative determination by fluorescence quenching studies. *Biochemistry* **15**, 672–679. (doi:10.1021/bi00648a035)
67. Eftink MR, Ghiron CA. 1984 Indole fluorescence quenching studies on proteins and model systems: use of the efficient quencher succinimide. *Biochemistry* **23**, 3891–3899. (doi:10.1021/bi00312a016)
68. Kleiger G, Beamer LJ, Grothe R, Mallick P, Eisenberg D. 2000 The 1.7 angstrom crystal structure of BPI: a study of how two dissimilar amino acid sequences can adopt the same fold. *J. Mol. Biol.* **299**, 1019–1034. (doi:10.1006/jmbi.2000.3805)
69. Beamer LJ, Carroll SF, Eisenberg D. 1998 The BPI/LBP family of proteins: a structural analysis of conserved regions. *Protein Sci.* **7**, 906–914. (doi:10.1002/pro.5560070408)
70. Lamping N *et al*. 1996 Effects of site-directed mutagenesis of basic residues (Arg 94, Lys 95, Lys 99) of lipopolysaccharide (LPS)-binding protein on binding and transfer of LPS and subsequent immune cell activation. *J. Immunol.* **157**, 4648–4656.

Electronic supplementary material (ESM)

The structure of latherin, a surfactant allergen protein from horse sweat and saliva.

Steven J. Vance[1¶], Rhona E. McDonald[2 §], Alan Cooper[1], Brian O. Smith[3]*, and Malcolm W. Kennedy[2,3]*

[1]School of Chemistry, [2]Institute of Biodiversity, Animal Health & Comparative Medicine, [3]Institute of Molecular, Cell & Systems Biology, University of Glasgow, Glasgow G12 8QQ, Scotland, UK.

Table S1. RMSD's calculated using UWMN software for structured and unstructured regions.

Table S2. A summary of the Ramachandran statistics for the ensemble of latherin structures as determined by PROCHECK_NMR.

Table S3. Distribution of leucines in latherin by secondary structure.

Figure S1. Latherin hydrogen-deuterium exchange data.

Figure S2. Comparison of latherin with other super-roll fold proteins.

Figure S3. Alignment of the amino acid sequence of latherin with human SPLUNC1 (BPIFA1).

Figure S4. Model of human SPLUNC1 (BPIFA1) based on the latherin structure - distribution of leucine and isoleucine residues within, and surface charge distribution.

Table S1. RMSD's calculated using UWMN software for structured and unstructured regions. The structured regions were selected as residues 8-30; 34-52; 62-69; 74-120; 127-142 and 158-201.

	RMSD to the unbiased mean (Å)	
	All	Structured
All heavy atoms	1.65	0.93
Backbone heavy atoms	1.40	0.66
C_α's	1.44	0.68

Table S2. A summary of the Ramachandran statistics for the ensemble of latherin structures as determined by PROCHECK_NMR.

Procheck Group	% residues
Residues in most favoured regions	82.3
Residues in additional allowed regions	15.4
Residues in generously allowed regions	1.7
Residues in disallowed regions	0.6

Table S3. Distribution of leucines in latherin by secondary structure. The values for leucines exposed to the solvent are shown in brackets. Solvent exposed leucines were identified through manual inspection and from PROCHECK analysis of the calculated structures.

Secondary Structure	# of residues	# of leucines	% leucines
α-helix	81	20 (4)	24.7 (4.9)
β-strand	53	14 (1)	26.4 (1.9)
Other	74	15 (5)	20.3 (6.8)
Overall	208	49 (10)	23.6 (4.8)

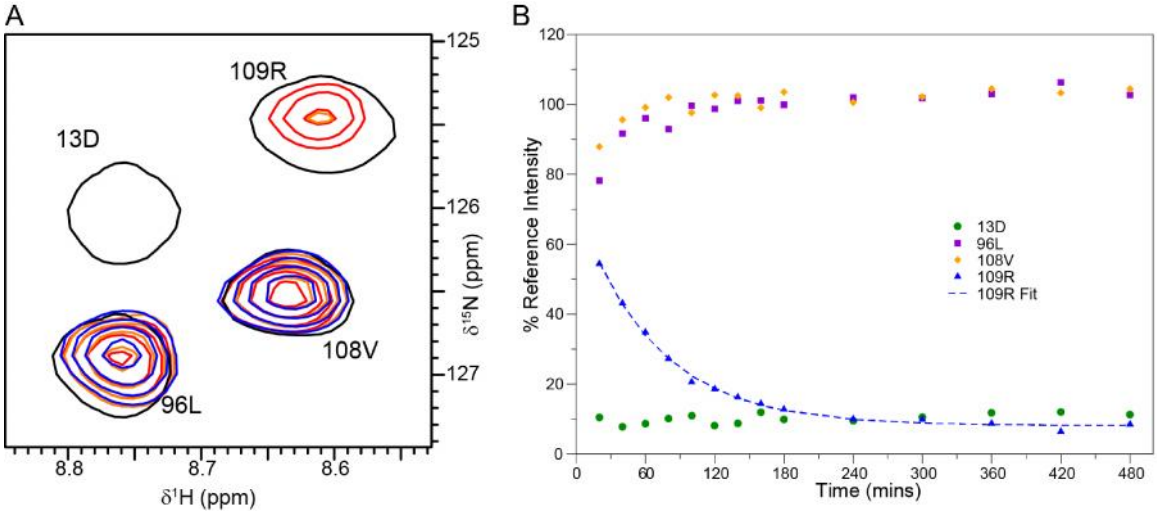


Figure S1. Latherin hydrogen-deuterium exchange data. (A) A region of the ^{15}N HSQC spectra of latherin undergoing hydrogen-deuterium exchange. Outlines of crosspeaks for residues 13D, 96L, 108V and 109R in a reference spectrum are shown in black. Spectra recorded 20 minutes after dilution to 90% D_2O are shown as red contours, after 120 minutes (orange) and after 480 minutes (blue). (B) Normalised crosspeak intensities for the four residues shown in A plotted with respect to time.

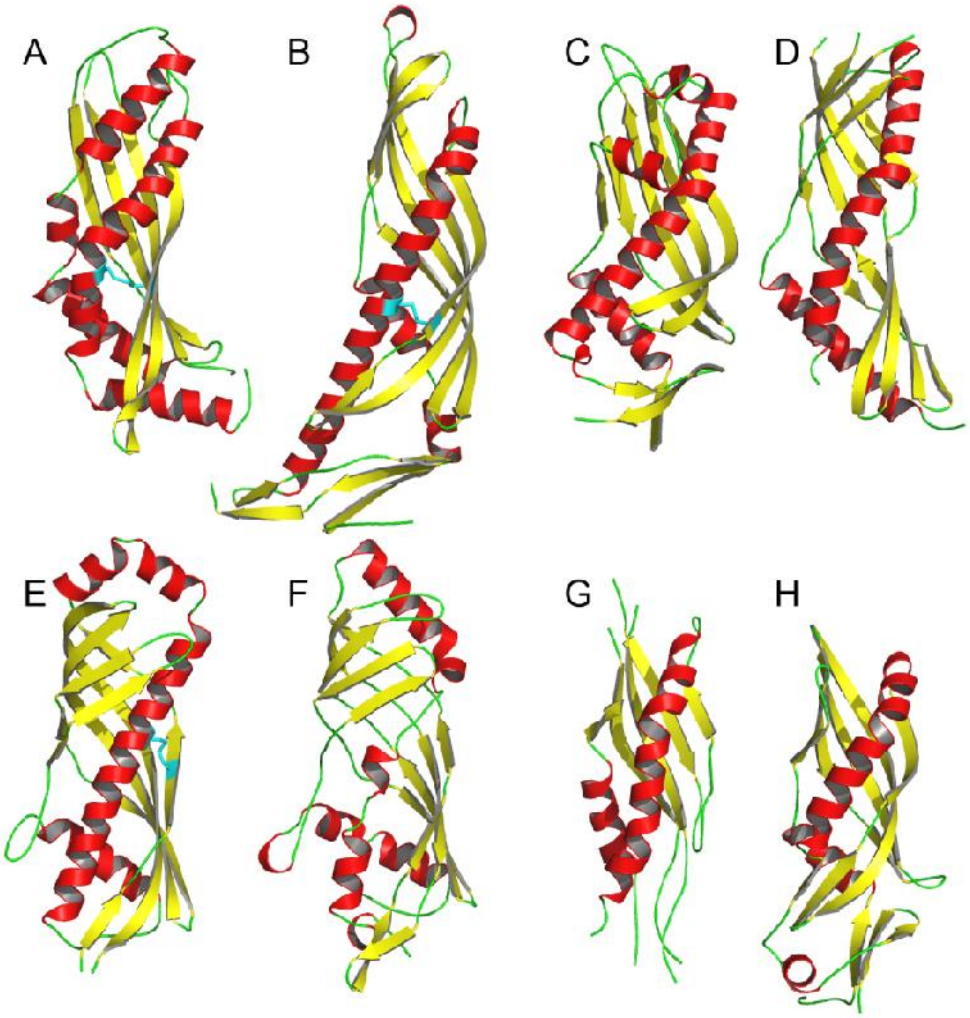


Figure S2. Comparison of latherin with other 'super-roll' fold proteins. Cartoon representations of (A) latherin, CETP N-terminal (B) and C-terminal (C) domains, (D) Der P 7 (allergen of the house dust mite *Dermatophagoides pteronyssinus*), (E) JHBP (insect juvenile hormone-binding protein), (F) Top1 (insect takeout protein), (G) Aha1 (activator of Hsp90 ATPase from yeast) and (H) Yceb (lipoprotein form *E. coli*). Disulphide bonds are shown in cyan. CETP N and C-term from 2OBD. Images created using PyMOL (1) from PDB IDs 3H4Z, 2RCK, 3E8T, 1USV and 3EYR.

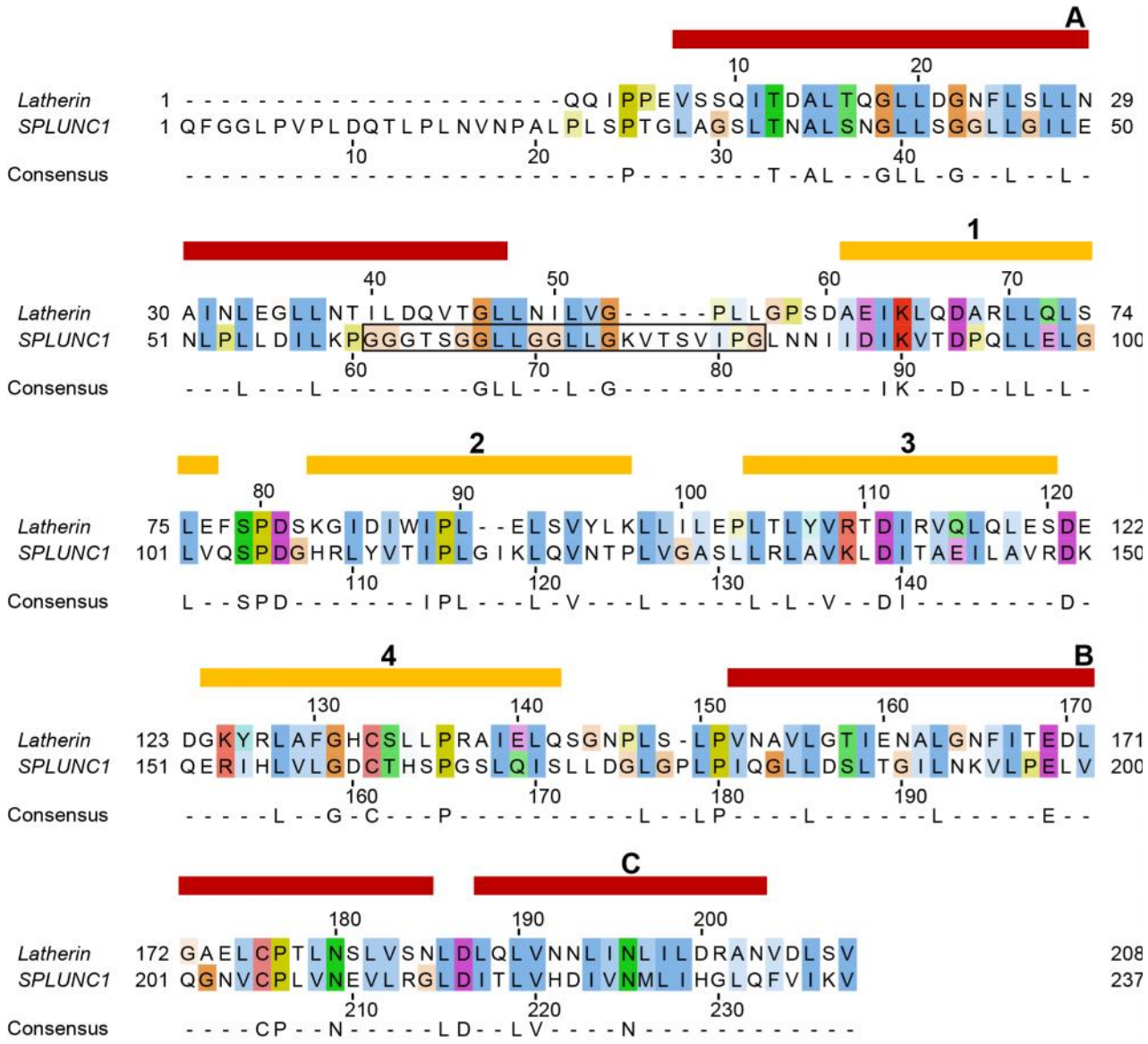
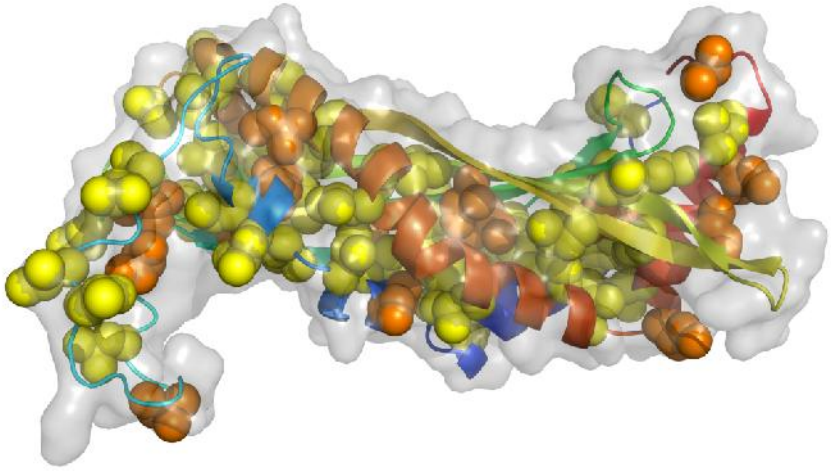
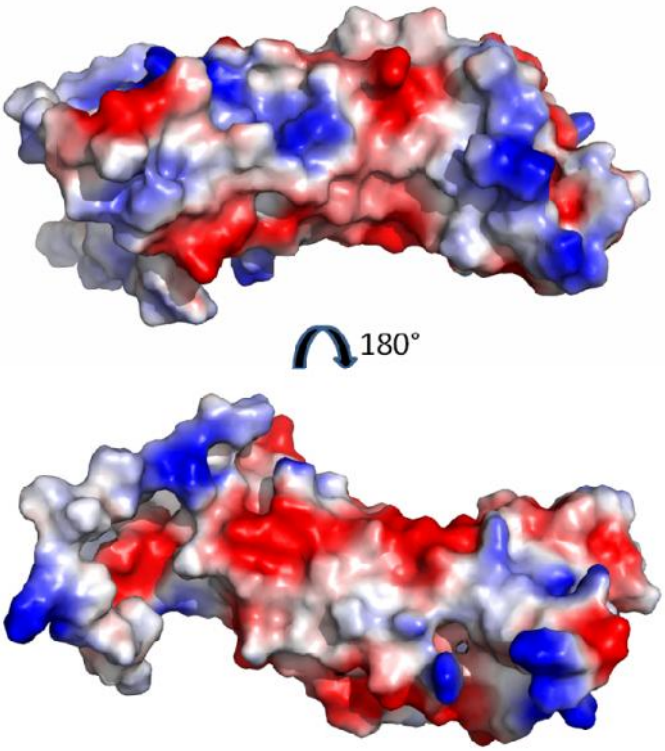


Figure S3. Alignment of the amino acid sequence of latherin with that of human SPLUNC1 (BPIFA1). Latherin's regular secondary structure elements are indicated with coloured bars and labeled red, α-helices; yellow, β-strands. Alignment produced using Jalview. The first 21 amino acids of SPLUNC1 have no counterpart region in the latherin sequence, so were excluded from the modeling procedure (see Figure S4). Multiple alignments with the sequences of other related proteins for which crystal structures are available (JHBP, BPI and CETP) indicated that they were too disparate to be

informatively included in the modelling procedure. The boxed region in the SPLUNC1 sequence is the glycine-rich region referred to in Figure S4.



A



B

Figure S4. Model of human SPLUNC1 (BPIFA1) based on the latherin structure - distribution of leucine and isoleucine residues within, and surface charge distribution. The model was created using MODELLER (2) (http://salilab.org/modeller/)

to create five models based on the sequence alignment in **Errore. L'origine riferimento non è stata trovata.** using the latherin structure as template. In order to create a model as representative of the likely SPLUNC1 structure as possible, the extended N-terminus which is not present in latherin was omitted and the extended loops and glycine rich sequence (boxed in Figure S3) were left unrestrained in the alignment. The models were analysed and the model with the lowest Discrete Optimised Potential Energy (DOPE) score (3) was selected. (A) Leucine side chains are displayed as yellow spheres. Isoleucine side chains as orange spheres. Human PLUNC is displayed in cartoon representation, coloured from blue at the N-terminus to red at the C-terminus. The leucines are distributed evenly throughout the protein apart from a high number found on the loops at the non-termini end of the protein (to the left of the figure). Image created using PyMOL (1). (B) The surface contact potential of the human PLUNC model, shown for two opposite faces of the molecule. Red = anionic, blue = cationic and grey = uncharged. The model suggests no evidence of any large uncharged areas that may correspond to a hydrophobic patch. Image created using PyMOL (1).

References for Electronic supplementary material (ESM)

1. The PyMOL Molecular Graphics System. 1.3 ed: Schrodinger, LLC.
2. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using Modeller. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2006 Oct;Chapter 5:Unit 5 6.
3. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein science : a publication of the Protein Society. [Comparative Study Evaluation Studies Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. 2006 Nov;15(11):2507-24.