Developmental and Comparative Immunology 139 (2023) 104582

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/devcompimm

Avian surfactant protein (SP)-A2 first arose in an early tetrapod before the divergence of amphibians and gradually lost the collagen domain

Srinivasa Reddy Kunchala [a,2,1], Albert van Dijk [b,2], Edwin J.A. Veldhuizen [b,*], Stephen C. Donnellan [c], Henk P. Haagsman [b], Sandra Orgeig [a,**]

[a] Centre for Cancer Diagnostics and Therapeutics, UniSA Cancer Research Institute, UniSA Clinical and Health Sciences, University of South Australia, SA, 5001, Australia
[b] Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands
[c] South Australian Museum, North Terrace, Adelaide, 5000, Australia

ARTICLE INFO

Keywords:
Collectins
Avian SP-A1 & SP-A2
Evolution
Pulmonary innate immunity
Molecular host defence

ABSTRACT

The air-liquid interface of the mammalian lung is lined with pulmonary surfactants, a mixture of specific proteins and lipids that serve a dual purpose-enabling air-breathing and protection against pathogens. In mammals, surfactant proteins A (SP-A) and D (SP -D) are involved in innate defence of the lung. Birds seem to lack the SP-D gene, but possess SP-A2, an additional SP-A-like gene. Here we investigated the evolution of the SP-A and SP-D genes using computational gene prediction, homology, simulation modelling and phylogeny with published avian and other vertebrate genomes. PCR was used to confirm the identity and expression of SP-A analogues in various tissue homogenates of zebra finch and turkey. *In silico* analysis confirmed the absence of SP-D-like genes in all 47 published avian genomes. Zebra finch and turkey SP-A1 and SP-A2 sequences, confirmed by PCR of lung homogenates, were compared with sequenced and *in silico* predicted vertebrate homologs to construct a phylogenetic tree. The collagen domain of avian SP-A1, especially that of zebra finch, was dramatically shorter than that of mammalian SP-A. Amphibian and reptilian genomes also contain avian-like SP-A2 protein sequences with a collagen domain. NCBI Gnomon-predicted avian and alligator SP-A2 proteins all lacked the collagen domain completely. Both avian SP-A1 and SP-A2 sequences form separate clades, which are most closely related to their closest relatives, the alligators. The C-terminal carbohydrate recognition domain (CRD) of zebra finch SP-A1 was structurally almost identical to that of rat SP-A. In fact, the CRD of SP-A is highly conserved among all the vertebrates. Birds retained a truncated version of mammalian type SP-A1 as well as a non-collagenous C-type lectin, designated SP-A2, while losing the large collagenous SP-D lectin, reflecting their evolutionary trajectory towards a unidirectional respiratory system. In the context of zoonotic infections, how these evolutionary changes affect avian pulmonary surface protection is not clear.

1. Introduction

The vertebrate lung serves a dual purpose. It is responsible both for air-breathing and for protecting the host from foreign particles beginning with the first breath (King and Clements 1972; Haagsman and van Golde 1985; Pison et al., 1994; Kingma and Whitsett 2006). This is achieved by lung surfactant, a thin lining at the delicate air-liquid interface made up of 90% lipids and 10% proteins (King et al., 1973; Daniels and Orgeig 2003). Of the four surfactant proteins (SP) classified as SP-A, -B, -C and -D (Possmayer 1988), SP-B & -C are highly hydrophobic proteins that are intricately involved with the surfactant lipids aiding their function to reduce surface tension at the air-liquid interface, thereby supporting air-breathing (Pison et al., 1990; Perez-Gil 2008). The remaining two hydrophilic proteins (SP-A and -D) are peripherally involved in the biophysical functions of surfactant but primarily facilitate pulmonary innate immunity (Possmayer 1990, Holmskov et al.,

* Corresponding author.
** Corresponding author. UniSA Clinical & Health Sciences, University of South Australia, SA, 5001, Australia.
E-mail addresses: e.j.a.veldhuizen@uu.nl (E.J.A. Veldhuizen), sandra.orgeig@unisa.edu.au (S. Orgeig).
[1] Current address: Advanced Respiratory Drug Delivery Solutions Pvt Ltd, 1–88 Annaram, Thungathurthy, Suryapet, Telangana-508280 India.
[2] Authors contributed equally to the manuscript.

https://doi.org/10.1016/j.dci.2022.104582
Received 6 June 2022; Received in revised form 20 October 2022; Accepted 20 October 2022
Available online 25 October 2022

2003; Kingma and Whitsett 2006, Haagsman et al., 2008).

Mammalian SP-A and SP-D show high sequence divergence despite the fact that they are thought to have arisen by gene duplication from a Mannose Binding Lectin (MBL)-like gene (Hughes 2007). Interestingly, SP-A molecules from various mammalian species are highly similar and form one clade on a phylogenetic tree that shares a node with mammalian SP-D (Hamzić, Pinard-van der Laan et al., 2015). The primary sequences of SP-A (Benson et al., 1985; Floros et al., 1986) and SP-D (Persson et al., 1990) show homology with other C-type lectins (Thiel and Reid 1989) in their carbohydrate binding region, but they contain an additional collagen domain that renders them as a separate family known as the collectins (Malhotra et al., 1992; Holmskov et al., 1994). The collectins are characterized by four domains namely, an N-terminal region, a collagen domain, linked by a neck region to the carbohydrate recognition domain (CRD) (Malhotra et al., 1992; Crouch et al., 1994). The collagen regions of three monomer polypeptides align to form a trimer subunit stabilised by disulfide bonds in the neck, CRD and N-terminal domains (Haagsman et al., 1989; Hoppe et al., 1994; Head et al., 2003). The trimerization process leads to higher order oligomer formation through the N-terminal domain (Whitsett et al., 1985; Sanchez-Barbero et al., 2005), which is the active structure in vivo. Several different functions have been assigned to SP-A and SP-D. The former is required for surfactant phospholipid aggregation (Haagsman et al., 1990), enhanced surfactant film formation and the lipid-reorganization process (Palaniyar et al., 2001; Lopez-Rodriguez et al., 2016), while SP-D does actually not interact with pulmonary surfactant phospholipids. Both collectins display calcium-dependent carbohydrate binding properties (Haagsman et al., 1987; Hartshorn et al., 1994), which play a role in neutralization of microbes entering the lung. Microbial (non-self) and host cell (self) surfaces are coated with different sugars to enable various physiological functions. These non-self sugars are recognised by the collectin CRD (Haurum et al., 1993; Casals et al., 2018) which promotes aggregation and opsonisation of the microorganism, to inhibit establishment of infection and spread of foreign particles (Crouch 1998; Haagsman and Diemel 2001, Hartshorn et al., 2010a). In mammals, it appears that SP-D is more important than SP-A in inhibiting pathogens due to both lectin-mediated and *N*-linked sialic acid interactions of the CRD with pathogens (Hartshorn et al., 1997; Reading et al., 1997; Hartshorn et al., 2010b; van Eijk et al., 2012).

Interestingly, many zoonotic respiratory infections such as avian influenza virus (AIV) spill over from the reservoir avian species leading to epidemics and pandemics causing severe infections in humans that makes the molecular protection rendered by these proteins inadequate. Despite the cause of many respiratory epidemics and pandemics that are zoonotic in nature, the majority of the respiratory infection's investigations are focussed on mammalian species. Furthermore, zoonotic respiratory infections such as 2009 swine flu, with increased frequency of recent epidemics as well as future pandemics warrant investigations into the mechanism of respiratory surface protection in reservoir species such as birds.

In avian research, chickens are by far the best studied species and most immunological data is indeed obtained in chicken. Chicken lungs express SP-A that is similar to mammals but is adapted with a putative coiled-coil structure that has replaced the latter part of the collagen domain (Hogenkamp et al., 2006). However, it appears that chickens lack the SP-D gene (Hogenkamp et al., 2006; Hughes 2007), but instead they have an additional SP-A-like gene also known as chicken lung lectin (cLL) (Hogenkamp et al., 2006) or renamed as chicken SP-A2 (cSP-A2). cSP-A2 is expressed as a 3-domain polypeptide with an N-terminal region linked via a neck domain to the CRD. Due to the absence of a collagen domain, an essential characteristic of collectins, this novel lung protein was not included in the collectin family. Nevertheless, initial functional investigations have shown that cSP-A2, like SP-A1 inhibits influenza A virus (IAV) (Hogenkamp et al., 2008). Whether the differences among the SP genes are specific to the chicken or are representative of the avian lineage is a question that we address in this study. Specifically, here we investigated the presence, tissue expression, homology models and evolution of SP-A1 and SP-A2 of zebra finch (*Taeniopygia guttata*) and turkey (*Meleagris gallopavo*), which were the two genomes available in public databases at the time this study began. We used computational gene prediction tools from published genomes, PCR, simulation models and phylogenetic analyses to expand our understanding of avian lung C-type lectin SP-A1 and SP-A2 evolution.

2. Materials and methods

2.1. Database search and primer design

Zebra finch and turkey were selected due to their importance as model organisms and the early availability of their genomes. The zebra finch is a small passerine bird native to central Australia and a model species for ecology, neuroscience and evolution (Warren et al., 2010). The pet birds have been shown to carry bacteria such as *Mycobacterium genavense* (Manarolla et al., 2009) that may affect human health. Turkeys are an agronomically important species with high susceptibility to respiratory influenza virus and *Escherichia coli* infections (Weebadda et al., 2001; Pillai et al., 2010). The zebra finch genome (*Taeniopygia guttata, assembly:* taeGut3.2.4) deposited in the ENSEMBL and NCBI databases was searched for surfactant protein-A1 (SP-A1), -A2 (SP-A2) and -D (SP-D) with the Basic Local Alignment Search Tool (BLAST)-and BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990) respectively. Chicken SP-A1 (cSP-A1; GenBank accession number: AF411083.1), chicken SP-A2 (cSP-A2; GenBank accession number: NP_001034255.1), human SP-D (hSP-D; GenBank accession number: NM_003019) and porcine SP-D (pSP-D; GenBank accession number: AAF22145.2) amino acid (a.a.) sequences were submitted as the query to find ortholog gene sequences in the zebra finch genome. The exon/intron boundaries of BLAT search tool output genome sequences were not clear, hence the Hidden Markov Model (HMM) gene prediction tool (HMMgene (v.1.1), http://www.cbs.dtu.dk/services/HMMgene/) was used to define the exon start and end to obtain complete predicted gene sequences. Primers (Supplementary Table S1) were designed with Primer 3 software (http://bioinfo.ut.ee/primer3-0.4.0/) to both the coding and the 5′ and 3' untranslated regions (UTR) of the HMM-predicted SP gene sequences in order to amplify these genes by PCR from zebra finch lung tissue (see below) to confirm the complete gene sequence. Two primer pair sets (Supplementary Table S1) each were designed for zfSP-A1 and SP-A2 gene sequence amplification. Similar gene search criteria were applied to the turkey genome (*Meleagris gallopavo*, *assembly:* UMD2) to find turkey SP-A1 (tuSP-A1), SP-A2 (tuSP-A2) and SP-D. Primer pairs (Supplementary Table S1) were selected to amplify tuSP-A1 and tuSP-A2 by PCR. Additionally, avian, reptilian, mammalian, fish (zebrafish, lungfish and coelacanth) and other vertebrate genomes published to date were NCBI BLASTP and tBLASTn searched (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul et al., 1990) by selecting the respective species or class in the organism column of the protein BLAST web page to find SP-A1, SP-A2 and SP-D like gene sequences for phylogenetic analysis.

2.2. RNA isolation for cDNA synthesis and polymerase chain reaction (PCR)

Adult healthy female zebra finches (n = 2; 12.4–12.5g), obtained from a colleague who was a breeder, and female turkey chicks (n = 2; 400–430g), obtained from the Gawler Poultry & Veg Markets (Evanston South, South Australia), were immediately (i.e. they were not housed) killed by intraperitoneal injection of 60–100 mg/kg pentobarbitone sodium. Birds were sacrificed according to animal ethics permit (241/10), approved by the Institute of Medical and Veterinary Science (IMVS), South Australia Animal Ethics Committee according to the Australian code for the care and use of animals for scientific purposes.

Tissues from respiratory and gastrointestinal tracts and other major organs were collected within 30 min of sacrifice. They were snap frozen in liquid nitrogen and stored at −80 °C. Tissues were processed on dry ice to isolate total cellular RNA. RNA from each tissue was extracted with a RNeasy® Mini Kit 50, (Qiagen GmbH, Hilden) using a Precellys®24 homogenizer (Precellys 24, Bertin Technologies, Montigny-le-Bretonneux, France). RNA quality was assessed with a NanoDrop 2000 (Thermo Scientific, Wilmington, MA). Isolated RNA (~1 μg) served as a template to synthesize cDNA using an iScript cDNA synthesis kit (Bio-Rad labs, Hercules, CA). Polymerase chain reaction (PCR) was performed with cDNA and primers (Supplementary Table S1) and GoTaq Green Master Mix (Promega corp. Madison, Wisconsin, USA) according to the manufacturer's instructions. In brief, each PCR reaction (25 μl) was performed for 3 min at 95 °C, followed by 37 cycles (30 s at 95 °C, 30 s annealing temperature according to Supplementary Tables S1 and 1 min at 72 °C) and final extension for 7 min at 72 °C. The PCR products were separated on an agarose (1%) gel. PCR samples with clear single bands of expected size were sequenced (Sanger sequencing service, Australian Genomic Research Facility (AGRF), Adelaide) according to Unpurified PCR products Dual (PD^{+}) protocol described on the service web site (http://www.agrf.org.au/services/sanger-sequencing). In the case of samples displaying multiple bands, the band of the expected size was dissected and gel extracted with a PureLink™ Quick Gel Extraction and a PCR Purification Combo Kit (Invitrogen, Carls, CA, USA) according to the manufacturer's protocol and sequenced by AGRF Sanger purified PCR product sequencing method.

2.3. Tissue expression analysis

PCR analyses of tissue homogenates were performed to confirm the relative expression of zf and tuSP-A1 and SP-A2 mRNA in various tissues (parabronchial lung, trachea, heart, kidney, liver and gastrointestinal organs - oesophagus, crop, proventriculus and large intestine). PCR was performed as described above and product was loaded onto an agarose gel (1%) with equal amounts of sample to show the relative expression.

2.4. Computational structure prediction for avian SP-A1

The zfSP-A1 protein sequence was similar to other known mammalian SP-A sequences, suggesting a similar structure and function. Hence, the rat SP-A neck and CRD region (NCRD) crystal structure, which was available in the protein data bank (PDB, http://www.rcsb.org/pdb/home/home.do), was used to build a homology model. The zfSP-A1 protein sequence was submitted to Schrödinger prime software (Schrödinger Release 2014-1: Prime, version 3.5, Schrödinger, LLC, New York, NY, 2014) to predict a homology model based on the PDB crystal structure 3PAK of the NCRD of rat SP-A (Shang et al., 2011).

2.5. Protein phylogenetic analysis

SP amino acid sequences from zebra finch and turkey were aligned with sequences from a more extensive range of taxa including species from the major avian orders (Supplementary Table S2), as well as accessible reptiles, mammals and other vertebrates, using characterised proteins and published genomes based on NCBI Gnomon predicted amino acid sequences. SP's in FASTA format were downloaded from published NCBI genome databases (http://www.ncbi.nlm. nih.gov/genome/browse/) and were aligned with the Clustal Omega (http://www.ebi.ac. uk/Tools/msa/) (Sievers et al., 2011) software tool to analyse homology and to construct phylogenetic trees. Phylogenetic trees were constructed using Maximum Likelihood (ML) methods implemented in MEGA 11 software package (Tamura et al., 2021) using the Jones-Taylor-Thornton model of amino acid substitution and uniform rates of substitution among sites. The tree was rooted with the zebra fish MBL like protein and nodal support was assessed with bootstrapping.

3. Results

3.1. Database searches for vertebrate surfactant proteins

In silico analysis of zebra finch, turkey, other published avian, reptilian (Squamata, Testudines and Crocodilia), amphibian and fish genomes using chicken SP-A1 and SP-A2 as the query yielded multiple exons with C-type lectin properties. NCBI tBLATn and BLASTP searches with human, mouse or pig SP-D protein sequences in zebra finch (zf) and turkey (tu) genomes or any of the other published bird genomes (47 species) did not reveal SP-D-like gene sequences. Next, we examined the organization of the SP-A locus in zebra finch, chicken and turkey and compared these with the SP-A/SP-D locus in pig and human (Fig. 1). In the mammalian and avian species these loci could be identified due to the presence of up- and downstream conserved flanking genes PPIF, ZCCH24, MAT1A, DYDC1 and TSPAN14. The clear absence of additional C-type lectin related genes in between the locus markers in these avian species supports the hypothesis that SP-D homologs are absent in birds. Despite the absence of SP-D-like sequences in avian genomes, all species in the amphibian clade and in all three reptilian clades showed the presence of predicted SP-A type (with EPN motif), SP-A2 type (with EPS motif) as well as SP-D. Interestingly, both the BLASTP and tBLASTn homology searches to find fish SP-A or SP-D gene sequences in the NCBI fish genome databases returned collagenous C-type lectin sequences with varying amino acid lengths of 235–262 in 7 species, and 331 amino acids in one species, the piranha. However, this search failed to identify collagenous lectin orthologs, with the exception of MBL-like orthologs in fish. These Gnomon predicted sequences are termed as SP-D-like in NCBI (for example: NCBI Reference Sequence: XP_017542004.1 or EMBL Id: A0A3B4DA97) (Table 1), but the majority are significantly shorter than other vertebrate SP-D sequences (371–375). The collagenous lectins within the range of 235–262 amino acids termed as SP-D-like are most likely MBL orthologs in fish. In addition, each of the eight sequences have either one or two breaks in the collagen domain. Finally, they have different antigen binding motifs (QPD/N or EPN/K/D and WN/L/DD).

The PCR sequencing results of predicted SP nucleotide sequences from zfSP-A1 and SP-A2 (Supplementary Figs. S1 and S2) and tuSP-A1 and SP-A2 (Supplementary Figs. S3 and S4) showed high conservation of the CRD and the retention of all structural amino acids involved in the CRD fold and the key amino acid motifs - either EPN (aSP-A1) or EPS (aSP-A2), containing amino acids Glutamic acid (E), Proline (P) and either Asparagine (N) or Serine (S) and WND, containing Tryptophan (W), Asparagine (N) and Aspartic acid (D), which are both critical for calcium ion dependent carbohydrate recognition. The multiple sequence alignment showed the conservation of cSP-A type EPN in 47 species (Supplementary Fig. 5) and cSP-A2 type EPS motif in predicted sequences of 43 avian species (Supplementary Fig. S6).

3.2. In silico identification and PCR sequence confirmation of zebra finch SP-A1 & SP-A2

A BLAT search for SP-A1 revealed C-type lectin-like exons that had sequence similarity to the gene sequences that code for known SP-A1 domains including the N-terminal region, collagen domain and the CRD (Supplementary Fig. S7) (Hoppe and Reid 1994; McCormack 1998). The zfSP-A1 predicted nucleotide sequence comprised three exons. Translation (http://web.expasy.org/translate) of the HMM-predicted complete gene sequence revealed recognisable collectin domains, with the first exon encoding the signal peptide, the N-terminal region and 12 amino acids of the collagen domain. Exon 2 encoded the final 3 amino acids of the collagen domain followed by the complete neck region. The third exon encoded the complete CRD with conserved EPN and WND motifs. PCR amplification to sequence and confirm the complete gene for zfSP-A1 using predicted gene sequence primers confirmed sequences with changes in 4 amino acid positions (Supplementary Fig. S1). The changes were single amino acids in the collagen

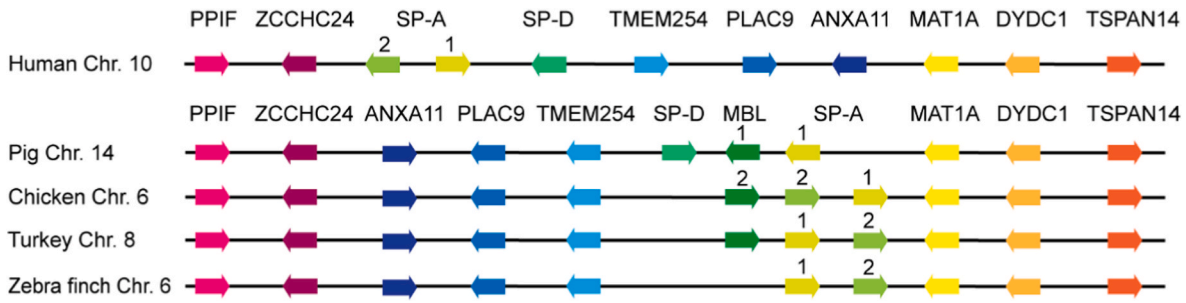

Fig. 1. Chromosome loci analyses of zf, tu, chicken, pig and human SPs. Comparison of the human and pig SP-A(/SP-D) cluster with that of the SP loci of three avian species showed the loss of SP-D in birds which further supports the complete loss of SP-D in these species.
In the examined mammalian and avian species these loci identified by the presence of up- and downstream conserved flanking genes PPIF, ZCCH24, MAT1A, DYDC1 and TSPAN14. The clear absence of additional C-type lectin related genes in between the locus markers in these avian species supports the hypothesis that SP-D homologs are absent in birds. The numbers 1&2 above the arrows represent SP-A1 and SP-A2 respectively. The number 2 above chicken and turkey MBL green arrow represent gene designated MBL2 (ENSGALG00010011044) within their SP-A locus of their respective genome databases.

domain, neck region and 2 amino acids close to the start of the CRD (highlighted in bold Supplementary Fig. S1). The PCR-confirmed zfSP-A1 gene sequence was submitted to GenBank (accession number: XM_003207786).

The cSP-A2 amino acid sequence was used as a query to find the zfSP-A2 gene sequence, yielding three exons. Translation of the zfSP-A2 gene sequence predicted a C-type lectin sequence that lacked the collagen domain. Exon 1 coded for the signal peptide and the N-terminal region; exon 2 coded for 25 amino acids of the neck domain; exon 3 translated to 5 amino acids of the neck domain and the CRD with EPS and WND motifs. However, no collagen domain was found. Intriguingly, there was a pseudogene located on chromosome unknown (ch Un) in the database. This gene appears to have lost the N-terminal region, but the C-terminal region was identical to SP-A2 on chromosome 6 (ch 6). zfSP-A2 amplification using primers selected from the computationally predicted gene sequence followed by sequencing of the overlapping PCR products revealed that the amplified gene was 159 nucleotides shorter than the predicted sequence by the HMM gene prediction tool. Translation of the sequenced PCR product confirmed that the gene comprised three exons (Supplementary Fig. S2), with exon 1 encoding the signal peptide and the N-terminal region, exon 2 encoding the first 25 amino acids of the neck domain and exon 3 encoding the final five amino acids of the neck domain and the CRD. The CRD sequence showed conservation of the two antigen recognition motifs, EPS and WND. The PCR-confirmed zfSP-A2 cDNA sequence was submitted to GenBank (accession number: XM_019618014).

3.3. *In silico identification and PCR sequence confirmation of Turkey SP-A1 & SP-A2*

Similar to the zfSP-A1 search the turkey genome database queried with the cSP-A1 protein sequence yielded five exons. Exon/intron boundaries were analysed by the HMM gene prediction tool to translate into the complete predicted protein sequence which revealed distinct collectin domains. Similarly, using cSP-A2 as a query to find the tuSP-A2 gene sequence resulted in four exons encoding the signal peptide, N-terminal region, neck domain and CRD. PCR amplification confirmed predicted SP-A1 and SP-A2, and their 5′ and 3′ UTR's by sequencing. Sequencing results for tuSP-A1 (Supplementary Fig. S3) and tuSP-A2 (Supplementary Fig. S4) were aligned (Clustal Omega tool) with predicted tuSP-A1 and SP-A2 and showed both gene sequences were 100% identical to the predicted sequences. The PCR-confirmed tuSP-A1 and SP-A2 cDNA sequences were submitted to GenBank (accession numbers: XM_030275876 and XM_030275877 respectively). Furthermore, the UTR regions of both zf (Supplementary Fig. S2) and tu (Supplementary Figs. S3 and S4) SP-A1 and SP-A2 were shorter than those for mammalian SPs.

3.4. *Comparative sequence analysis of avian SP-A1 and SP-A2 sequences*

Clustal Omega alignment of SP-A1 from more than 20 avian species (Fig. 2) showed that the highest variation in the collagen domain was due to the presence of the putative coil structure within the domesticated species. Non-domesticated avian species either had a reduced collagen domain (Fig. 2) with between two and 13 G-X-Y repeats (Table 1) or an unknown region, for example in the Chimney swift, Downy Woodpecker and Kea parrot, but no known putative coil structure. The multiple sequence alignment was extended to demonstrate the significant variation in collagen domains among mammals, amphibians, turtles, snakes and lizards (Squamata) and birds. Specifically, the human SP-A1.1 (mammal) and painted turtle SP-A1 and SP-A2 sequences were selected and aligned, due to their similar size. The alignment showed the conservation of the turtle SP-A1 collagen domain with 24 G-X-Y repeats with a break after 12 repeats then ending with GLP (Table 1). However, a similar alignment of human SP-A1.1 and painted turtle SP-A2 (Supplementary Fig. S8) showed that both collagen domains are the same size ending with a GLP triplet and multiple breaks (2 + 2+3 + 5+2) (Table 1) in turtle SP-A2. Amphibian SP-A2 demonstrated a large collagen domain with either no interruption in the Western clawed frog (26 G-X-Y triplet repeats repeats) or one interruption in the African clawed frog (13 + 11 repeats) (Table 1). Among the reptilian SP-A2 homologs the SP-A2 sequence of turtles (Testudines), snakes and lizards (Squamata) retained varying lengths (2-14) of G-X-Y triplet repeats in their collagen domain (Table 1), but this domain was completely absent from birds and their sister group, the alligators (Crocodilia). The CRD domain of SP-A2 homologs from all these sauropsid species contained the conserved EPS and WND motifs.

The multiple sequence alignment demonstrates the degree of sequence similarity between avian, frog and human SPs which enabled us to determine the percentage sequence identity between different species and within different areas of the protein (Fig. 3). The zfSP-A1 total protein sequence was 52% identical to cSP-A1, 44% to human and 49% to frog. The SP-A1 sequence identity between these diverse vertebrates increased towards the C-terminal end (Fig. 3). The turkey SP-A1 sequence was highly similar to that of the chicken with 91% identity (Fig. 3) for total protein and the sequence identity increased towards the C-terminal end with 94% for neck and CRD region (NCRD) and 96% for CRD. As with cSP-A1, tuSP-A1 had a very low identity with hSP-A, only 39% identity for the total protein, 51% for NCRD and 54% for CRD. Identity with the frog sequence was higher with 49% and 59% identity to frog total SP-A1 and the NCRD, respectively. The sequence of zfSP-A2 was similar to that of tuSP-A2 and cSP-A2 with 61% identity for the total protein, again with increasing sequence identity towards the C-terminal end with 69% identity for the NCRD and 74% for the CRD (Fig. 3). TuSP-A2 was very similar to cSP-A2 with 92% CRD

Table 1
Summary of major patterns of difference in the collagen domain between the SP-A1 and SP-A2 proteins in mammals, birds, reptiles, frogs and fish derived from BLASTP search of NCBI genome database.

Species	Collagen G-X-Y repeats	Collagen domain function in lungs
Mammalian SP-D	59 with no interruption	Lipid recycling (Orgeig et al., 2010), antigen uptake (Kishore and Reid 2001) by macrophages and anti-inflammatory activity (Olde Nordkamp et al., 2014)
Mammalian SP-A	23-24 with one break after 13th repeat	Surfactant lipid association, formation of tubular myelin (Haagsman et al., 1991); protects surfactant from protein inhibition (Ikegami et al., 2001; Hogenkamp et al., 2006); promotes phagocytosis of antigen (Tenner et al., 1989) chemotaxis in response to antigen (Kishore and Reid 2001)
Chimpanzee and human SP-A1.2	"	Highly similar to mammalian SP-A1.1 allele with only 4 amino acid differences in collagen domain and functions like SP-A1; The formation of the SP-A1.2 allele in these more derived species may be due to duplication to meet both physiological demand and innate defence in large lungs
Bird SP-A1	2-13 (multiple breaks)	Unknown, unpredictable due to gaps in collagen G-X-Y repeats
Bird SP-A2	0	No collagen domain
Alligator SP-A1-like	26 (2 + 15+1 + 8)	Unknown, but based on sequence similarity it is likely to function like mammalian SP-A
Alligator SP-A2-like similar to avian SP-A2	0	No collagen domain
Python SP-A1-like	24-26 (12 + 12 or 2 + 12+12)	Unknown, but based on sequence similarity it is likely to function like mammalian SP-A
Python SP-A2-like similar to avian SP-A2	20 (2 + 10)	Unknown, unpredictable due to gaps in collagen G-X-Y repeats
Lizard SP-A1-like	23 (12 + 11)	Unknown, but based on sequence similarity it is likely to function like mammalian SP-A
Lizard SP-A2-like similar to avian SP-A2	10 no breaks	Unknown, no known functional data
Turtle SP-A1-like	24 (12 + 12) with one break after 12th repeat in both turtles	Unknown, but based on sequence similarity it is likely to function like mammalian SP-A
Painted turtle SP-A2-like similar to avian SP-A2	14 with multiple breaks 2 + 2+3 + 5+2	Unknown, unpredictable due to gaps in collagen G-X-Y repeats
Green sea turtle SP-A2-like similar to avian SP-A2	6 with 2 + 2+2 intervals	Unknown, unpredictable due to gaps in collagen G-X-Y repeats
Frog SP-A1-like	2	Unknown, due to small size
Western clawed frog SP-A2-like similar to avian SP-A2	26 with no interruption	Unknown, based on sequence similarity it is likely to function like mammalian SP-A
African clawed frog SP-A2-like similar to avian SP-A2	24 (13 + 11)	Unknown, based on sequence similarity it is likely to function like mammalian SP-A
Frog SP-D	19 no break	Unknown
Fish MBL-like	21 (12 + 9)	Unknown, but based on sequence similarity it is likely to function like mammalian MBL. No SP sequences were found in 3 published fish genomes
Piranha D-like	47 (12 + 18)	Unknown

conservation (Fig. 3). Finally, cSP-A2 was 51%, 53% and 58% similar to predicted frog SP-A2 total protein, NCRD and CRD, respectively. The sequence comparison also showed identity between predicted frog SP-A1 and SP-A2 with 50% for total protein, 56% for NCRD and 62% in CRD.

3.5. Tissue expression of zf and tu SP-A1 & −2 mRNA

PCR analyses of tissue homogenates showed the relative expression of zf and tu SP-A1 and SP-A2 mRNA in various tissues (parabronchial lung, trachea, heart, kidney, liver and gastrointestinal organs - oesophagus, crop, proventriculus and large intestine) (Fig. 4). SP-A1 and SP-A2 mRNA expression were both higher in lung and trachea than in the other major organs (Fig. 4).

3.6. Computational structure prediction for SP-A1

The predicted avian SP-A1 homology models (Fig. 5) were built on the rat SP-A template (3PAK) (Shang et al., 2011). The superimposed structures showed high similarity to rat SP-A. The CRD of rat SP-A comprises three α-helices and 11 β-sheets to form a T shape (Head et al., 2003), although there is not always a general agreement on assignment of these short secondary structures within CRD's (Zelensky and Gready 2005, Veldhuizen et al., 2011). The zfSP-A1 and rat SP-A NCRD sequences have the same length (148 a.a. long). Their alignment showed 47% sequence identity and 67% of amino acids had a positive score in the substitution matrix, defined as a substitution that preserves the physico-chemical property of the amino acid. Hence, zfSP-A1 contained amino acids required for alpha helices and beta sheets to form a similar distinctive T-shape ribbon structure (Fig. 5A). More specifically the rat C terminus end residue, i.e. Phe-228 (F), forms a salt bridge and an H-bond with side chains of the neck residues His-96 (H) and Gln-100 (Q) to form a unique T shape (Head et al., 2003). The three equivalent zfSP-A1 amino acids, F-191, H-59 and Q-63, were conserved at the same relative positions suggesting the ability to form a similar T-shape. Furthermore, metal ion (calcium) binding in primary and auxiliary sites gives the rat SP-A 198–203 (GQGKEK) loop increased stability which brings the E−202 side chains closer to the ligand binding site which binds to phospholipid and carbohydrate. Similarly predicted zf SP-A1 had a conserved loop between 160 and 166 (GKGDEK) with E−165 which may be able to bind ligands similar to those of rat SP-A due to the location of these a.a. at the exact distance from the C-terminal end. tuSP-A1 formed a similar structure (Fig. 5B) to zfSP-A1 and the conservation of loop (GKGTEK) suggests similar functions.

3.7. Phylogenetic analysis

In the ML tree constructed from amino acid sequences and rooted with the zebra fish MBL-like protein (Fig. 6), avian SP-A1 sequences, characterised by EPN and WND motifs and the presence of a shortened collagen domain formed a sub-clade as one of six sub-clades comprising the amniote SP-A1 clade. The amniote SP-A2 sequences, characterised by EPS and WND motifs and the presence of shortened collagen domain among reptilian groups, i.e. snakes, lizards and turtles, form a clade that comprises the avian SP-A2 sub-clade, characterised by EPS and WND motifs and the absence of a collagen domain and three reptilian sub-clades. The SP-A1 and SP-A2 clades are each other's closest relatives. The amphibian SP-A1 and SP-A2 sequences formed two successively more divergent sister clades to the amniote SP-A1 and SP-A2 clades. Crocodile SP-A2 is counterintuitively placed in what appears to be an unlikely place in the tree, i.e., with Chinese alligator SP-A1 within the amniote SP-A1 clade. The sequence labelled Crocodile SP-A1 is located in the vertebrate SP-D clade. Confirmation of the NCBI Gnomon predicted amino acid sequences and protein classification for the crocodile is needed. The zebra finch and turkey SP-A1 sequences grouped with their evolutionary relatives, i.e., Passeriformes species white-throated

Fig. 2. Collagen domain variation in SP-A1 of avian species. The Clustal Omega multiple sequence alignment of avian SP-A1 showed the presence of a putative coiled coil sequence (in the dotted box) in the domesticated species (chicken, turkey, duck), and an undefined region in Chimney swift, Downy woodpecker and Kea parrot. Due to the presence of the putative coiled-coil sequence in the domesticated species and the undefined domain in the later three species, the avian SP-A1 collagen domain (characterised by G-X-Y triplet, black box) displayed large gaps and reduced conservation relative to the N-terminal and the start of the CRD domains. In this alignment the highlighted 'G' amino acids show the pattern of collagen repeats (G-X-Y) in avian species. The VLA/S/T sequence is the start of the CRD domain (blue box) which showed a more conserved alignment compared with the collagen domain.

sparrow and collared fly catcher; and Galliformes (chicken) and Anseriformes (duck) relatives respectively. The PCR sequenced zfSP-A2 grouped with the other Passeriformes species, i.e., the white-throated sparrow and collared fly catcher as expected from their taxonomic affinities.

The vertebrate SP-D sequences, characterised by EPN and WND motifs in the CRD linked to a large collagen domain (comprising up to 59 repeats of GXY), form a clade with three frog SP-D sequences being the closest relatives of the amniote sub-clade. The SP-D clade is the sister to the SP-A1 and SP-A2 clades.

Bootstrap support for relationships within each of the major SP clades were generally 70% or greater, a value generally regarded as strong support (Hillis and Bull 1993). Nodes towards the base of the major clades and for relationships in general across the base of the tree were not supported strongly.

4. Discussion

Computational gene prediction tools confirmed the presence and predicted the sequence of zf - and tuSP-A1 and SP-A2. The sequences and normal adult lung tissue expression of these proteins were confirmed by PCR. Phylogenetic analysis of 24 avian, as well as representative mammalian, reptilian and amphibian homologs demonstrated that birds have lost SP-D, retained a shorter SP-A1 with a reduced collagen domain and retained SP-A2 that is likely derived from frog type SP-A2 with an EPS motif. Avian SP-A2 was characterised by the presence of the EPS motif in the CRD and the complete loss of the collagen domain. Similarly, alligator SP-A2 also completely lacked the collagen domain (Table 1; Fig. 7). Furthermore, amphibians and non-crocodilian reptiles possessed an avian-like SP-A2, i.e., with the EPS motif, but with variable length collagen domains (Table 1; Fig. 7). Despite the changes or loss of the collagen domain, the CRD domains were highly conserved among all the tetrapods, and bird SPs retained the key amino acid motifs that are

SP-A1

	Frog			Anole			Sea turtle			Alligator			Zebra finch			Chicken			Turkey			Pig			Human		
	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD
Zebra finch	49	53	54	52	56	57	47	63	66	60	62	65	100	100	100	52	68	71	54	71	74	45	49	50	44	48	52
Chicken	50	50	50	45	59	60	50	64	66	58	62	64	53	69	72	100	100	100	91	94	96	50	52	53	38	50	51
Turkey	49	49	59	44	58	60	52	65	69	59	63	66	54	71	74	91	94	96	100	100	100	51	52	55	39	51	54

SP-A2

	Frog			Anole			Sea turtle			Alligator			Zebra finch			Chicken			Turkey		
	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD	Total	NCRD	CRD
Zebra finch	51	51	52	47	59	61	37	47	46	37	61	62	100	100	100	61	69	74	63	61	75
Chicken	51	53	58	51	61	64	36	45	46	58	60	63	61	70	72	100	100	100	86	89	91
Turkey	49	52	58	49	61	65	39	44	47	59	62	64	63	61	75	86	88	92	100	100	100

Fig. 3. Surfactant protein amino acid similarities among avian, human, reptilian and amphibian homologs (%). The total protein includes 4 domains, marked as N-terminal, Collagen domain, Neck region and Carbohydrate recognition domain (CRD). The NCRD refers to the neck and CRD domains in combination. The lowest homology is shown in dark blue, and the increasing intermediate homology is indicated by a color progression from light blue, through decreasing shades of green to yellow, increasing shades of orange and finally to red indicating the highest homology.

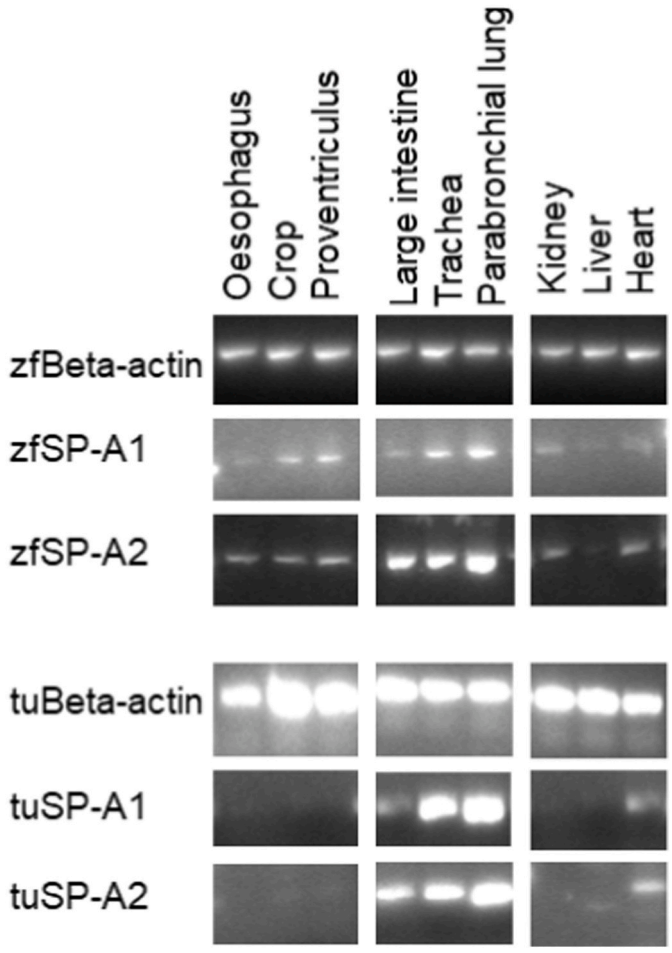

Fig. 4. Relative tissue expression of SP-A1 & SP-A2 mRNA in parabronchial lung, trachea, heart, kidney, liver and gastrointestinal organs of zebra finch and turkey. Bands that represent beta actin were positive loading controls for zf and tu tissues respectively. The SPs higher expression in lung and trachea suggest that these proteins are specialised to serve in lung protection.

required for antigen recognition (i.e. EPN in SP-A1; EPS in SP-A2; and WND in both).

Given the ancient origin of the pulmonary surfactant system, which is thought to predate the evolution of lungs and swim bladders (Daniels et al., 1998; Daniels and Orgeig 2003), and the fact that ancient actinopterygian and sarcopterygian as well as more modern teleost fish possess a surfactant system (Daniels et al., 2004), we undertook comprehensive searches for SP-like genes in NCBI fish genomes. However, these searches failed to identify clear SP orthologs apart from collagenous lectin MBL-like sequences. Specifically, a frog and avian SP-A2 based gene search to find SP-A2 in the NCBI fish databases returned collagenous C-type lectins with an amino acid length range of 235–262. The variability in fish C-type lectin sequence length may be due to either the annotation status of the genome sequences available for the BLAST search or the low sequence coverage. These collagenous C-type lectins in fish (~7 species) are within amino acid length range of 235–262 (Table 1) except one in the piranha SP-D-like (EMBL Id: A0A3B4DA97 or NCBI Reference Sequence: XP_017542004.1) that was 331 amino acid long. Of these 7 collagenous C-type lectins the majority is similar in length to zebra fish MBL with one or two breaks in the collagen domain and varying antigen binding motifs. The longer piranha sequence of 331 amino acids demonstrated two breaks in the collagen domain. This sequence appears to be similar to the well-characterised mammalian SP-D of ~375 amino acids length with a contiguous collagen domain that gives rise to the multimeric cruciform structure (Crouch et al., 1994; Holmskov 2000). However, the two breaks in the piranha SP-D-like collagen domain do not support this protein to classify as SP-D. Furthermore, the predicted fish collagenous C-type lectins adopted altered motifs such as QPD/N or EPN/K/D compared to the MBL type EPN motif in amphibian, all reptilian (Squamata, Testudines and Crocodilia), avian and mammalian sequences. The EPN and WND motif containing collectins in fish are about 250 amino acids long and are similar to zebra fish MBL. Hence the evidence thus far has been able to reveal the presence of MBL as the only collagenous fish lectin. The non-detection of collagenous lectins or SPs in fish may be due to the annotation status of the NCBI fish genome databases.

4.1. Function of domains in avian SP-As

The collectin collagen domain is an important component to promote the coiled-coil trimer structure (Karinch and Floros 1995, Palaniyar et al., 2001; Head et al., 2003), the interaction with macrophage receptors (Tenner et al., 1989; Malhotra et al., 1992; Geertsma et al., 1994; Yang et al., 2015) and phospholipid liposome aggregation (Wright et al., 1987). SP-D monomers contain a large uninterrupted collagen domain that contributes to the formation of a multimeric cruciform structure (Drickamer and Taylor 1993; Kingma and Whitsett 2006). The collagen domains of SP-A, MBL and C1q analogues consist of an intermediate number of G-X-Y repeats with one break, which supports the formation of the so-called bouquet shaped multimers (Voss et al., 1988; McCormack 1998). In human collectin protein collagen domains the number of

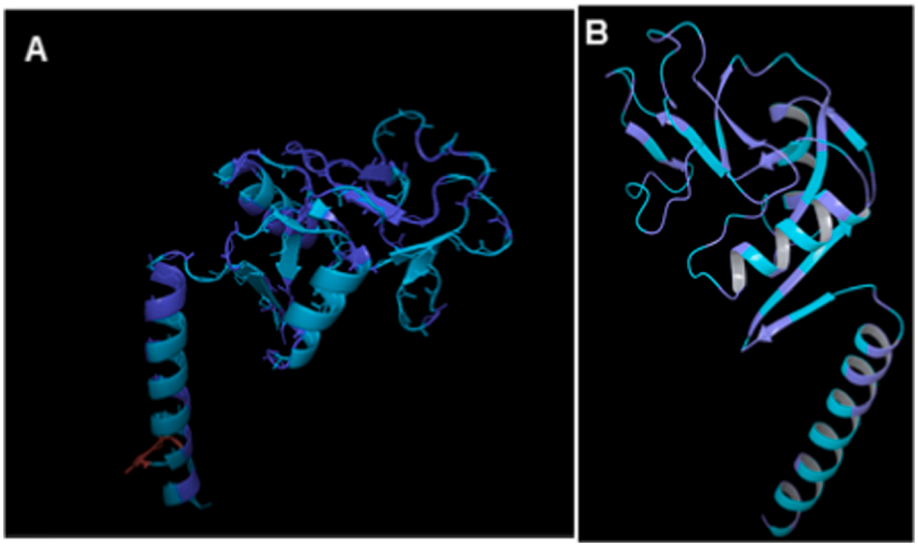

Fig. 5. Ribbon diagram of predicted zf (A) and tu (B) SP-A1 homology models. superimposed on the rat SP-A NCRD crystallographic structure (3PAK). The superimposition of zfSP-A1 (purple, Fig. A), tuSP-A1 (purple, Fig. B) showed complete alignment with rat SP-A (turquoise). Predicted avian protein structures showed 3 α-helices and 11 short β-sheets to form quaternary structures that are highly similar to that of rat SP-A.

G-X-Y repeats ranges from 19 in mammalian MBL (Dommett et al., 2006) to 24 in hSP-A (Haagsman and Diemel 2001) and 59 in hSP-D (Crouch et al., 1991). Considerably fewer repeats are found in avian surfactant proteins, i.e. 3 in cSP-A1 (Hogenkamp et al., 2006), 4 in turkey SP-A1 (Supplementary Fig. S3) and 5 in zebra finch SP-A1 (Supplementary Fig. S1). Predicted SP-A1 sequences of other avian species revealed similarly truncated collagen domains (Fig. 2). The impact of these truncated collagen domains on the function of avian SP-A proteins is unknown, but is expected to be profound, as it may affect their oligomeric structure and immune function.

In cSP-A1, the collagen domain is joined to an unknown putative coiled-coil structural domain (Hogenkamp et al., 2006). We found that this pattern is characteristic for the SP-A1 sequences of the three domesticated avian species investigated, i.e. chicken, turkey (both Galliformes) and duck (Anseriformes). In other mammalian or avian SP-A analogues or other collectin family members, this putative coiled-coil domain is absent and its role in protein structure and/or function is unknown (Hogenkamp et al., 2006).

The neck region is a structurally important domain as it connects the N-terminal region and collagen domain to the globular CRD structure in all collectins (Drickamer 1988; Hoppe and Reid 1994) and initiates the trimerization of the collagen domain and CRD (Crouch et al., 1994; Hakansson et al., 1999) (Supplementary Fig. S7). Intermolecular bonds between neck region histidine and glutamine residues with a CRD phenylalanine residue supports the formation of a "T" shape of the NCRD in rSP-A (Head et al., 2003). These three amino acids are well conserved in zf, tuSP-A1 and other predicted avian and reptilian SP-A proteins, suggesting that their CRDs may adopt similar configurations.

The CRD is the functional unit of collectins with affinity for the various sugars and lipids present on the cell surface of pathogens as well as on abnormal host cells (Haurum et al., 1993; Head et al., 2003; Seaton et al., 2010). As seen in mammalian CRD domains, both zf- and tuSP-A1 CRD contain four conserved cysteine residues that serve to maintain the globular structure of the CRD (Voss et al., 1988; Rust et al., 1991; van de Wetering et al., 2004). Likewise, EPN and WND amino acid motifs, known to be involved in carbohydrate and lipid recognition (Drickamer 1992, Veldhuizen et al., 2011; Rynkiewicz et al., 2017), are well-conserved in avian species. Also, the size of the avian SP-A1 CRD of zebra finch, turkey, chicken and other avian species (~125aa) is similar to that of mammalian orthologs (115–130 aa) (Drickamer 1988, Weis et al., 1991; van de Wetering et al., 2004). Taken together, this overall resemblance emphasizes the functional importance of this domain throughout vertebrate evolution.

4.2. *Biological relevance of SP-As in birds*

Mammalian SP-As exhibit a biophysical function in which the CRD is required for tubular myelin formation which in turn enhances the formation of the phospholipid surface film (Haagsman et al., 1990; Korfhagen et al., 1996; Ikegami et al., 2001; Wang et al., 2010; Lopez-Rodriguez et al., 2016) by binding to DPPC (Kuroki and Akino 1991) and cholesterol (Yu et al., 1999). Other mammalian SP-A functions include surfactant layer stabilization, protection from serum proteins, lipid recycling, regulating the surfactant turnover metabolism and promoting surfactant uptake by type II cells and macrophages (van Golde et al., 1988; Cockshutt et al., 1990; Schurch et al., 1992). In bird lungs, atrial type II pneumocytes secrete surfactants presumably to spread to the air capillaries in order to maintain airflow along the airways (i.e. series of bronchi), patency of the air capillaries to some extent and prevention of edema through formation of a continuous thin surface film that reduces the surface tension to some extent at the air-liquid interface and facilitates breathing (Bernhard et al., 2001; Bernhard et al., 2004). Birds may not need tubular myelin and large aggregates of surfactant phospholipids, but they still need secreted surfactant to reduce surface tension to some extent at the air-liquid interface. While film formation is mainly promoted by SP-B, the additional association with multimeric SP-A enhances the rate of film formation and its turnover metabolism (Pison et al., 1990; Lopez-Rodriguez et al., 2016). The high similarity of the NCRDs of zebra finch, turkey and chicken SP-A1 suggests that this domain is under greater negative or purifying selection, which is not surprising as this domain is directly involved in surfactant biophysical and host defence activities. Increasing sequence identity towards the C-terminal end with highest identity in the CRD was also found in avian SP-A2 sequences. Likewise, frog SP-A1 and SP-A2 are most similar at their C-termini. These findings suggest that broad purifying selection in the NCRD and CRD functional domains is a common phenomenon among the Tetrapods.

Chicken SP-A1 and SP-A2 gene pulmonary expression in response to viral and bacterial infection (Hogenkamp 2007, Reemers et al., 2010) were investigated. However, the crystal or 3D structures of avian SP-A1 and SP-A2, their carbohydrate binding properties and their interaction with respiratory pathogens are unknown. The predicted homology

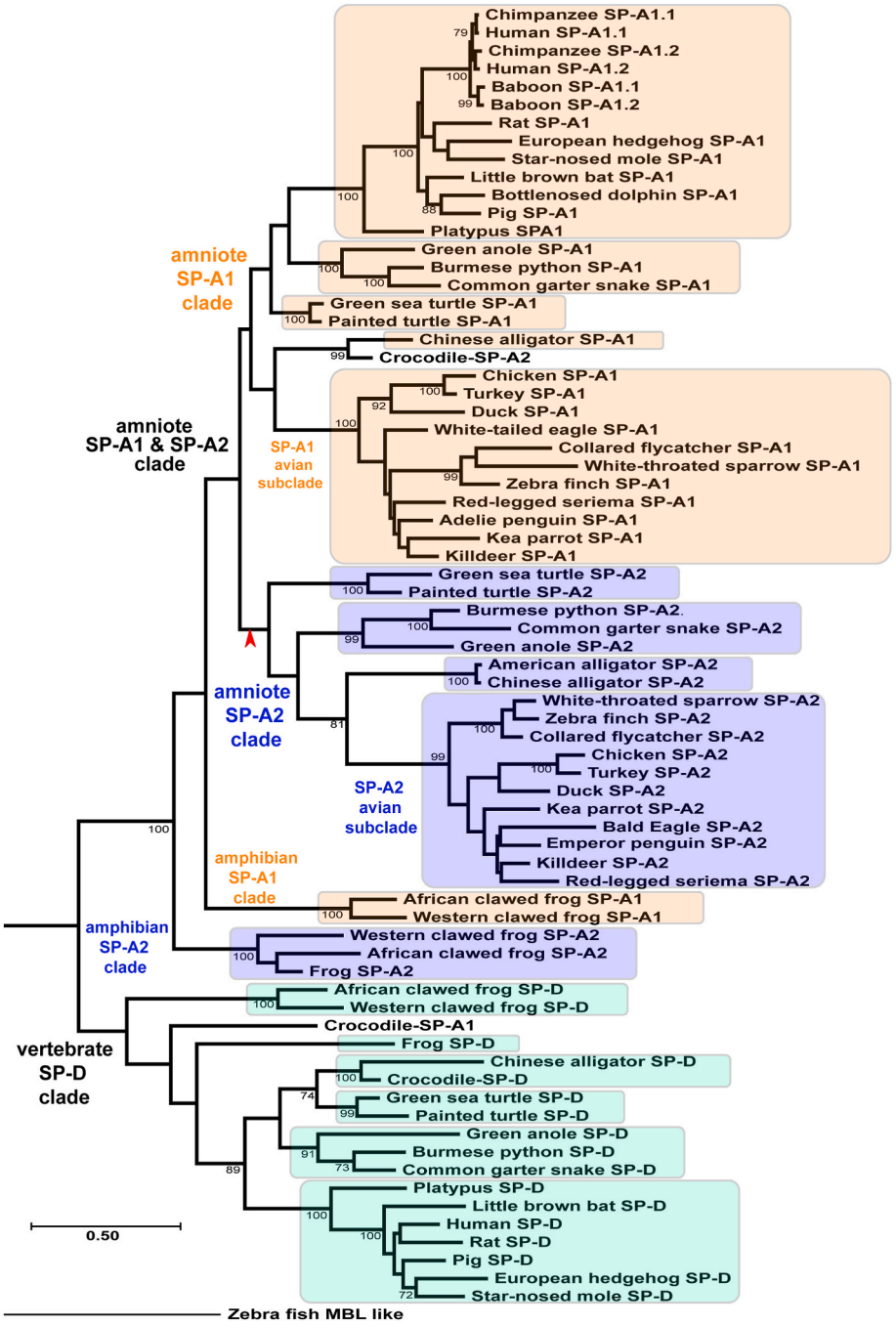

Fig. 6. Vertebrate SP-A protein phylogenetic tree. Phylogenetic tree wase constructed using Maximum Likelihood (ML) methods implemented in MEGA 11 software package (Tamura et al., 2021) using the Jones-Taylor-Thornton model of amino acid substitution and uniform rates of substitution among sites. The tree was rooted with the zebra fish MBL like protein and nodal support was assessed with bootstrapping. Bootstrap support for relationships within each of the major SP clades were generally 70% or greater, a value generally regarded as strong support (Hillis and Bull 1993). Nodes towards the base of the major clades and for relationships in general across the base of the tree were not supported strongly. This may be due to short nature of the sequences and the multi domain nature of these unique proteins where there is size variation due to either addition, expansion/reduction or loss of domains. However, the tree showed avian SP-A1 clades separated from reptilian, mammalian and amphibian SP-A1, SP-A2 and SP-D clades.

models of zf- and tuSP-A1 (Fig. 5) and relative tissue expression of SP-A1 and SP-A2 mRNA in zf and tu (Fig. 4) showed the highest expression in healthy bird trachea and lungs and supports the notion that both proteins are adapted to perform surfactant biophysical and/or pulmonary-specific immune functions. Preliminary functional analyses of cSP-A2 suggest a mild inhibition of haemagglutination by human IAV isolates (Hogenkamp et al., 2008). Interestingly, experimental influenza infection of chickens resulted in up or down regulation of SPs mRNA expression that was site-specific (i.e. trachea and lung) as well as age-dependent (Reemers et al., 2010). The high sequence identity between zfSP-A1, tuSP-A1 and cSP-A1 suggests that zfSP-A1 and tuSP-A1 may have a similar role in AIV respiratory infection, although the functionality against AIV still needs to be demonstrated for both proteins.

4.3. *Evolution of SP-A like C-type lectins to serve surfactant function*

Earlier phylogenetic studies compared chicken SP-A1 and SP-A2 with mammalian SP-A, SP-D, MBL, CL-L1 and CL-K1 and showed that cSP-A1 and cSP-A2 were more closely related to mammalian SP-A than to SP-D (Hogenkamp et al., 2006) and that mammalian and avian SP-A arose from a common ancestor. Another phylogenetic analysis of SP-A, SP-D and MBL from mammals and chicken revealed that mammalian SP-A1 formed a single clade that was more closely related to the chicken clade of SP-A (Hughes 2007) than to the mammalian SP-D branch, while SP-D was related more closely to MBL than to SP-A. Others found that clades of all SP-As were joined to the clade of other collectins and SP-D (Hamzić, Pinard-van der Laan et al., 2015). Furthermore, the cSP-A2 type was thought to have evolved specifically in birds after the separation of Aves and Crocodilia (Hamzić, Pinard-van der Laan et al., 2015).

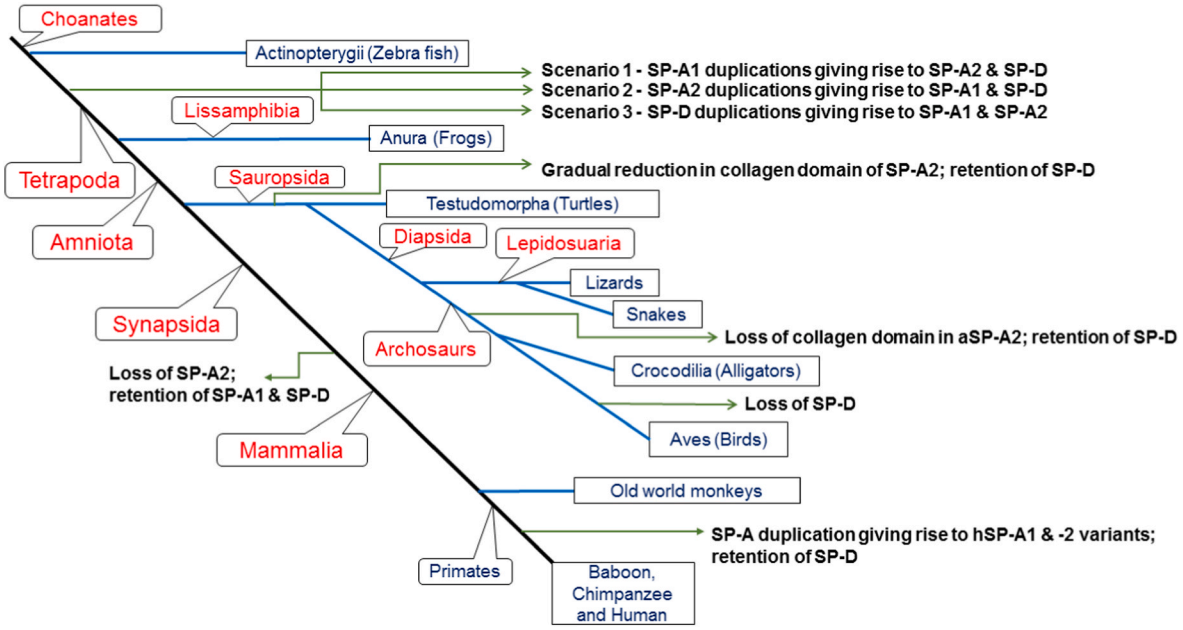

Fig. 7. Proposed models for vertebrate evolution of SP-A and SP-D. A lack of evidence of the existence of SP-A and/or SP-D in fish, limits our ability to determine which SP-A protein is ancestral. Hence we provide three potential scenarios:
1. SP-A1 (the ancestor) was duplicated in an ancestor of the Tetrapoda leading to frog type SP-A2, followed by a second duplication event creating SP-D.
2. SP-A2 (the ancestor) was duplicated in an ancestor of the Tetrapoda leading to frog type SP-A1, followed by a second duplication event creating SP-D.
3. SP-D (the ancestor) was duplicated in an ancestor of Tetrapoda leading to frog type SP-A1 or SP-A2.
SP-A2 was lost in mammals and subjected to changes in the Sauropsida leading to a gradual reduction and finally a complete loss of the collagen domain in the archosaurs. Additionally, the loss of the collagen domain in birds correlated with the loss of SP-D. A third duplication of SP-A1 in a primate ancestore gave rise to hSP-A1.1 and hSP-A1.2.

With the availability of a more representative set of SP sequences from the vertebrates, in particular from the major amniote lineages, we re-evaluated the phylogenetic relationships among SP-A1, SP-A2 and SP-D orthologs (Fig. 6). The absence of SP-D orthologs amongst birds is consistent with the loss of SP-D in the avian lineage after the separation of crocodilian and avian lineages. We speculate this may be due to the loss of need for SP-D function in the small avian rigid parabronchial lung. In the mammalian lung, surfactant homeostasis and lipid metabolism are primarily governed by SP-D (Botas et al., 1998; Kishore and Reid 2001).

This is supported by the observation that a disturbed surfactant homeostasis characteristically occurs in SP-D knockout mice (Botas et al., 1998). As the mammalian bronchoalveolar lung undergoes large and rapid cyclic volume changes, there is a need for matched (dynamic) surfactant film formation, spreading and re-spreading after each breath. To maintain surfactant homeostasis, this requires adaptations in the rate of lipid metabolism and turnover of surfactant components to facilitate each breath. On the other hand, the avian respiratory system is a highly differentiated system (composed of parabronchi and cranial and caudal air sacs) which may demand a different surfactant metabolism. The rigid compact parabronchial lungs of birds do not change volume significantly (Jones et al., 1985) they lack alveoli and are aerated in a single direction by a bellows-like action of air sacs that pump inspired air via a relatively rigid structure of parabronchi (Bernhard et al., 2001; Bernhard et al., 2004; Woodward and Maina 2008). As a result, the avian parabronchial lung may not need enhanced surfactant film formation, extreme surface tension reduction (Daniels et al., 1998), quick spreading, re-spreading after each breath, and hence a high rate of lipid metabolism and a rapid turnover of surfactant components to facilitate each breath. Thus, the loss of the SP-D gene in birds may have been facilitated by a loss-of-function mutation. Despite these evolutionary changes in avian lung structure and function, the pulmonary immune protection from foreign pathogens must however be maintained. With the evolution of aSP-A1 (with an EPN motif), and the additional aSP-A2, as well as other unknown adaptations, it is possible that the innate immune function of SP-D became redundant at the avian pulmonary surface. Hence, there may have been two loss of function evolutionary drivers, which led to the loss of aSP-D.

Our constructed phylogenetic tree incorporated avian species as well as amphibians and the polyphyletic reptilians among which were several crocodilians, the only living members of the archosaurs (Green et al., 2014). We found variable numbers of collagen repeats among amphibian and reptilian SP-D, SP-A1 and SP-A2 orthologs (Table 1). The SP-A2 collagen domain was conserved in turtles, snakes and lizards, but absent in alligators, the closest living relative of birds (Green et al., 2014). Like birds, alligators possess a unidirectional respiratory system (Schachner et al., 2014) with a similar arrangement of the primary and secondary bronchi (anatomical similarity), branching of bronchi terminating in a sac-like tip, similar topography of the bronchial passages and the concentration of gas-exchange parenchyma in small areas of large lungs (Schachner et al., 2013; Farmer 2015). As both crocodilians and avian possess a unidirectional respiratory system, it has been suggested that this is an ancestral (i.e. plesiomorphic) trait; that is it did not evolve as a result of either endothermy or flight.

Because amphibians and reptiles express all three surfactant collectins it is plausible that an initial duplication event leading to SP-A2 must have occurred in an ancestral tetrapod predating the amphibians. The loss of the SP-A2 collagen domain and deletion of SP-D in birds appears to represent complete adaptations of this group towards a compact rigid parabronchial lung. We postulate that the partial loss in non-crocodilian reptilians of the collagen domain in SP-A2 gradually led to the complete loss of the SP-A2 collagen domain in crocodiles and birds. For example, painted turtle SP-A2 contains an interrupted collagen domain (break in between G-X-Y repeats e.g. 2 + 2+3 + 5+2 (Table 1) that corresponds to the intact collagen domain in human SP-A1.1 and terminates with a GLP triplet (Supplementary Fig. S8). The second duplication of surfactant proteins occurred in an ancestor of the Tetrapoda leading to the formation of SP-D with a gradual expansion of the collagen domain and

changes in other amino acids in the N-terminal and NCRD domains according to the changing needs of the host. The expanding SP-D collagen domain, conserved among all non-avian vertebrates, presumably fulfils the pulmonary innate immune and homeostasis functions of amphibians, reptilians and mammals (Kishore et al., 2006). Furthermore, the presence of SP-A1, A2 and -D in amphibians indicates that these proteins diverged very early. The presence of SP-D in amphibians, all reptilian branches and in mammals suggests that this protein evolved slowly among these groups, due to strong stabilizing selection pressures, likely related to the protein's innate immune function. SP-A2, on the other hand evolved rapidly after the separation of the crocodilian and avian lineages and was lost completely in mammals. Mammalian SP-A1 arose from reptilian SP-A1 and maintained a stable size even in the most derived species, the primates. However, this protein evolved rapidly in the avian lineage, likely to meet the changing respiratory needs of this lineage. The third duplication event gave rise to human SP-A1.1 and SP-A1.2 and occurred in a recent primate ancestor of the chimpanzee and human, likely to meet the changing needs of primates, which represent a more derived group of mammals. The shift from a plant-based food diet to a more meat-enriched diet, the need for endurance to hunt large prey to exhaustion (Hart 2011) as well as the change to communal living under soiled conditions (Curtis 2007; Curtis 2014; Perry 2014, Poirotte et al., 2017) may have placed greater innate immune and biophysical demands on the bronchoalveolar lung. To serve these increased innate immune defence needs it is likely that the SPs were co-opted (exapted) and subjected to greater positive selection pressures leading to functional divergence at the respiratory surface of these two primate species, relative to other mammals. The fact that the SP-A2 paralog was found to be retained in the majority of the investigated avian species indicates that avian SP-A gene duplication and sub-functionalization must have resulted in added value, most likely in an increased protection of the avian respiratory system. It should be noted however that the EPS motif in cSP-A2 was shown to exhibit a lower affinity for mannose-based ligands than SP-D that contains an EPN motif (Hogenkamp et al., 2008). Naturally, these adaptations in avian SPs may render the reservoir species protected against zoonotic pathogens such as AIV while the same pathogens are highly virulent to humans and other mammals. Further investigations are needed to determine how birds compensate for the absence of an SP-D analog (Hogenkamp et al., 2006) that has important immune functions in mammals (Crouch et al., 2006; Orgeig et al., 2010).

In summary, the avian SP-A1 N-terminal region is less conserved than the C-terminal region. Birds have lost SP-D, have a reduced SP-A1 collagen domain and lost the SP-A2 collagen domain. All are presumably the result of functional adaptations to the small rigid parabronchial lung of birds. The short SPs that may act as a first line of defence combined with unidirectional ventilation which presumably is able to flush out foreign particles more efficiently than tidal ventilation, thereby likely contributing to avian parabronchial lung protection from air-associated pathogens. Hence, bird surfactant proteins must have retained key amino acid residues in the CRDs required for antigen recognition to fulfil the host defence role. Further support for the unique evolutionary trajectory of the bird surfactant collectins comes from computational gene sequence analyses of reptilians (Squamata, Testudines and Crocodilia), in which different groups possess either collagenous or non-collagenous avian type SP-A2. Thus, the reptilian surfactant proteins appear to serve as transition proteins between birds and mammals.

4.4. Potential evolutionary scenarios for the origin of SP-A1 and -A2 and SP-D

Because we are currently lacking critical information on SPs in tetrapod ancestors, specifically the fish, we cannot determine whether either frog type SP-A1 or SP-D with its EPN motif or the frog SP-A2 with its EPS motif represent the ancestral SP-A. Hence, in the absence of this critical evidence, we provide three alternate evolutionary scenarios; the first in which SP-A1 serves as the ancestor, the second in which SP-A2 serves as the ancestor and the third scenario in which SP-D with an EPN motif similar to MBL-like proteins serves as the ancestor. Evidence for each of these scenarios comes from the evolutionary history as well as published functional data of these important functional motifs.

4.4.1. SP-A1 as the ancestor

In this scenario we propose two duplication events of the ancestral SP-A gene (Fig. 7) at some time before the origin of the amphibians; one leading to SP-A2 and the other leading to SP-D. SP-A2 subsequently evolved an EPS motif which remained throughout its evolution in amphibians, the polyphyletic reptilians and birds. This adaptation may have led to a reduction in the protein's biophysical functionality as part of a subfunctionalisation mutation process. During its evolution among the Squamata, Testudines and Crocodilia to the birds, SP-A2 gradually underwent a reduction in its collagen domain (Fig. 7), possibly representing further specialisation for an innate immunity role both in the embryo and the adult stage. The only collectin, which adapted an EPS motif, is CL-10 (Hansen et al., 2016) which has specialised roles during embryogenesis and morphogenesis (Munye et al., 2017). A similar function can likely be performed by EPS-containing SP-A2 that is also expressed in the embryo, at least in chickens (Zhang et al., 2016). Throughout its evolution SP-D retained the EPN motif and evolved a larger collagen domain. This likely represents further immune specialisation and suggests significant functional redundancy in all reptilian orders due to the retention in these vertebrate groups of both types of SP-A proteins and SP-D, with both SP-A1 and -D containing an EPN motif. While birds lost the SP-A2 collagen domain, mammals completely lost the frog type SP-A2 protein. However, SP-A1 was retained from frogs to mammals and was reduplicated in the chimpanzee and human to form human type SP-A1.2.

4.4.2. SP-A2 as the ancestor

In this scenario we propose two duplication events in the ancestor of the amphibians (Fig. 7); one leading to SP-A1 and the other leading to SP-D, both of which subsequently evolved an EPN motif. The ancestral SP-A2 had a long uninterrupted collagen domain which was gradually reduced among the reptilian orders and lost in the birds. It also had an EPS motif that was retained throughout its evolution from amphibians to all reptilian orders to the birds. SP-A2 was retained likely due to the requirement for a less stringent innate immune function during the final stages of embryonic development to meet the necessary immune tolerance at that stage and innate immune protection immediately after hatching. SP-A1 retained the EPN motif throughout its evolution from amphibians to all reptilian orders to the birds but in birds the collagen domain was significantly reduced thereby likely reducing its ability to perform biophysical functions. In mammalian SP-A1 the EPN motif evolved to an EPA/R motif (human, cow, dog and rat) possibly to meet the greater biophysical demands placed on the bronchoalveolar lung. Furthermore, there was a 3rd duplication event in a primate ancestor of the chimpanzee and human leading to the human SP-A1.1 and A1.2 type variants incorporating 4 amino acid changes in the collagen domain. Throughout the evolution from amphibians to all reptilian orders and mammals, SP-D retained the EPN motif but evolved a larger collagen domain in the reptilian orders and the mammals. This may represent further immune specialisation and a significant limitation in biophysical function associated with this protein.

4.4.3. SP-D as the ancestor

Interestingly, the presence of predicted SP-D from frogs to mammals except in birds and its consistent linkage map with MBL and SP-A on human chromosome 10 (Hughes 2007; Hamzić, Pinard-van der Laan et al., 2015) suggests that an SP-D or MBL-like C-type lectin may have served as the ancestor to all the extant surfactant protein C-type lectins. This alternative possibility can only be resolved upon identification of the amphibian ancestral collagenous C-type lectin gene sequences.

5. Conclusions

In conclusion, we have shown in the vast majority of bird genomes investigated, the presence of both an avian SPA-1 and an avian SPA-2 gene, indicating their functional importance at the pulmonary surface. In all birds, aSP-A1 was shorter than its mammalian orthologs due to a reduction in the size of the collagen domain. The avian-type SP-A2 gene characterised by an EPS motif is also present in amphibians and all groups of reptilians. In the amphibians and non-crocodilian reptilians there is a variable-sized collagen domain. However, the aSPA-2 gene in birds and their sister group, the crocodilians, lacks the collagen domain entirely, making this non-collagenous protein a uniquely archosaurian characteristic. The two most modern derived vertebrate lineages - birds and mammals, adopted two different strategies of adaptive simplification that are reflected by two separate lineage-specific deletions. Birds lost SP-D but retained SP-A2 with its EPS motif and mammals lost SP-A2. We therefore suggest that in amphibians and reptilian orders the evolution of surfactant protein collectins represents a transition state that incorporates significant functional redundancy across the proteins, but that birds/crocodiles and mammals have refined their arsenal and lost some of that redundancy.

Finally, the loss of the major innate immune agent SP-D, reduction of the collagen domain in SP-A1 as well as complete loss of the collagen domain in SP-A2 of the three species investigated would logically have important implications for the immune system, and the susceptibility towards pathogens. Yet, despite specific differences between species in lectin repertoire and collagen domain composition the compact rigid avian parabronchial lung does not seem to lack protection against common air-associated pathogens. It is tempting to speculate that the immune response role of mammalian SP-Ds may have been taken over by aSP-A2 in birds, which would be an oversimplification and warrants a broader examination of (innate) immune cells and other immune effector molecules. Based on their conserved sequences in all species, it is clear though that the neck and lectin domain are important in the initial recognition of pathogens and that multimerization of lectin domains, recognizing carbohydrate residues on pathogens, appears essential to initiate this immune response. To what extent aSP-A2 may compensate for SP-D in recognizing specific pathogens is unknown but this is also only a small part that defines susceptibility. Overall susceptibility will also be determined by the pathways that clear collectin-bound pathogens through aggregation and mechanical removal and by interaction with immune cells, with pathway preferences that will likely deviate between species.

Ethics approval and consent to participate

Animal studies were undertaken according to the Australian code for the care and use of animals for scientific purposes under animal ethics permit (241/10), approved by the Institute of Medical and Veterinary Science (IMVS), South Australia Animal Ethics Committee.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Funding

Aspects of the work were funded by a Nature Foundation SA, Australia grant to SRK and funding body did not contribute in any way intellectually.

Authors' contributions

SRK was involved in the design of the experiments, collected, analysed and interpreted the data and first drafted the manuscript and took responsibility for incorporating edits, SO was responsible for conception and design of the study, supervising the execution of the study on a day-by-day basis, undertaking the animal work, interpreting the data and had a major role in writing the manuscript, HH, EJAV and AvD were involved in the conception of the study and design of the experiments and trained SRK in various aspects of collectin biology, SCD was responsible for training SRK in phylogenetic analyses and contributed to data interpretation, All authors read, contributed to and approved the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

We thank Dr. Greg Johnston for his generosity in providing native zebra finches and Dr Malika Kumarasiri for assistance with homology modelling.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2022.104582.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410.
Benson, B., Hawgood, S., Schilling, J., Clements, J., Damm, D., Cordell, B., White, R.T., 1985. Structure of canine pulmonary surfactant apoprotein: cDNA and complete amino acid sequence. Proc. Natl. Acad. Sci. U.S.A. 82 (19), 6379–6383.
Bernhard, W., Gebert, A., Vieten, G., Rau, G.A., Hohlfeld, J.M., Postle, A.D., Freihorst, J., 2001. Pulmonary surfactant in birds: coping with surface tension in a tubular lung. Am. J. Physiol. Regul. Integr. Comp. Physiol. 281 (1), R327–R337.
Bernhard, W., Haslam, P.L., Floros, J., 2004. From birds to humans: new concepts on airways relative to alveolar surfactant. Am. J. Respir. Cell Mol. Biol. 30 (1), 6–11.
Botas, C., Poulain, F., Akiyama, J., Brown, C., Allen, L., Goerke, J., Clements, J., Carlson, E., Gillespie, A.M., Epstein, C., 1998. Altered surfactant homeostasis and alveolar type II cell morphology in mice lacking surfactant protein D. Proc. Natl. Acad. Sci. USA 95 (20), 11869–11874.
Casals, C., Campanero-Rhodes, M.A., García-Fojeda, B., Solís, D., 2018. The role of collectins and galectins in lung innate immune defense. Front. Immunol. 9.
Cockshutt, A.M., Weitz, J., Possmayer, F., 1990. Pulmonary surfactant-associated protein A enhances the surface activity of lipid extract surfactant and reverses inhibition by blood proteins in vitro. Biochemistry 29 (36), 8424–8429.
Crouch, E., Persson, A., Chang, D., Heuser, J., 1994. Molecular structure of pulmonary surfactant protein D (SP-D). J. Biol. Chem. 269 (25), 17311–17319.
Crouch, E., Rust, K., Persson, A., Mariencheck, W., Moxley, M., Longmore, W., 1991. Primary translation products of pulmonary surfactant protein D. Am. J. Physiol. Lung Cell Mol. Physiol. 260 (4), L247–L253.
Crouch, E.C., 1998. Collectins and pulmonary host defense. Am. J. Respir. Cell Mol. Biol. 19 (2), 177–201.
Crouch, E.C., Smith, K., McDonald, B., Briner, D., Linders, B., McDonald, J., Holmskov, U., Head, J., Hartshorn, K., 2006. Species differences in the carbohydrate binding preferences of surfactant protein D. Am. J. Respir. Cell Mol. Biol. 35 (1), 84–94.
Curtis, V.A., 2007. Dirt, disgust and disease: a natural history of hygiene. J. Epidemiol. Community 61 (8), 660–664.
Curtis, V.A., 2014. Infection-avoidance behaviour in humans and other animals. Trends Immunol. 35 (10), 457–464.
Daniels, C.B., Lopatko, O.V., Orgeig, S., 1998a. Evolution of surface activity related functions of vertebrate pulmonary surfactant. Clin. Exp. Pharmacol. Physiol. 25 (9), 716–721.
Daniels, C.B., Orgeig, S., 2003. Pulmonary surfactant: the key to the evolution of air breathing. NIPS (News Physiol. Sci.) 18, 151–157.
Daniels, C.B., Orgeig, S., Sullivan, L.C., Ling, N., Bennett, M.B., Schürch, S., Luis Val, A., Brauner, C.J., 2004. The origin and evolution of the surfactant system in fish: insights into the evolution of lungs and swim bladders. Physiol. Biochem. Zool. 77 (5), 732–749.

Daniels, C.B., Orgeig, S., Wood, P.G., Sullivan, L.C., Lopatko, O.V., Smits, A.W., 1998b. The changing state of surfactant lipids: new insights from ancient animals. Am. Zool. 38 (2), 305–320.
Dommett, R.M., Klein, N., Turner, M.W., 2006. Mannose-binding lectin in innate immunity: past, present and future. Tissue Antigens 68 (3), 193–209.
Drickamer, K., 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. J. Biol. Chem. 263 (20), 9557–9560.
Drickamer, K., 1992. Engineering galactose-binding activity into a C-type mannose-binding protein. Nature 360 (6400), 183.
Drickamer, K., Taylor, M.E., 1993. Biology of animal lectins. Annu. Rev. Cell Biol. 9 (1), 237–264.
Farmer, C.G., 2015. The evolution of unidirectional pulmonary airflow. Physiology 30 (4), 260–272.
Floros, J., Steinbrink, R., Jacobs, K., Phelps, D., Kriz, R., Recny, M., Sultzman, L., Jones, S., Taeusch, H.W., Frank, H.A., 1986. Isolation and characterization of cDNA clones for the 35-kDa pulmonary surfactant-associated protein. J. Biol. Chem. 261 (19), 9029–9033.
Geertsma, M.F., Nibbering, P.H., Haagsman, H.P., Daha, M.R., van Furth, R., 1994. Binding of surfactant protein A to C1q receptors mediates phagocytosis of Staphylococcus aureus by monocytes. Am. J. Physiol. Lung Cell Mol. Physiol. 267 (5 Pt 1), L578–L584.
Green, R.E., Braun, E.L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M. W., John, J.A.S., Capella-Gutiérrez, S., Castoe, T.A., 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science 346 (6215), 1254449.
Haagsman, H., van Golde, L., 1985. Lung surfactant and pulmonary toxicology. Lung 163 (1), 275–303.
Haagsman, H.P., Batenburg, J.J., Clercx, C., Geelen, M.J.H., van Golde, L.M.G., 1990. Surfactant Lipids and Proteins in the Perinatal and Adult Lung. Endocrine and Biochemical Development of the Fetus and Neonate. Springer US, Boston, MA, pp. 231–250. J. M. Cuezva, A. M. Pascual-Leone and M. S. Patel.
Haagsman, H.P., Diemel, R.V., 2001. Surfactant-associated proteins: functions and structural variation. Comp. Biochem. Physiol. Mol. Integr. Physiol. 129 (1), 91–108.
Haagsman P., Henk, Elfring H., R, Van Buel L.M., B, Voorhout F., W, 1991. The lung lectin surfactant protein A aggregates phospholipid vesicles via a novel mechanism. Biochem. J. 275 (1), 273–276. https://doi.org/10.1042/bj2750273.
Haagsman, H.P., Hawgood, S., Sargeant, T., Buckley, D., White, R.T., Drickamer, K., Benson, B.J., 1987. The major lung surfactant protein, SP 28-36, Is a calcium-dependent, carbohydrate-binding protein. J. Biol. Chem. 262 (29), 13877–13880.
Haagsman, H.P., Hogenkamp, A., van Eijk, M., Veldhuizen, E.J.A., 2008. Surfactant collectins and innate immunity. Neonatology 93 (4), 288–294.
Haagsman, H.P., White, R.T., Schilling, J., Lau, K., Benson, B.J., Golden, J., Hawgood, S., Clements, J.A., 1989. Studies of the structure of lung surfactant protein SP-A. Am. J. Physiol. Lung Cell Mol. Physiol. 257 (6), L421–L429.
Hakansson, K., Lim, N.K., Hoppe, H.J., Reid, K.B.M., 1999. Crystal structure of the trimeric alpha-helical coiled-coil and the three lectin domains of human lung surfactant protein D. Structure 7 (3), 255–264.
Hamzić, E., Pinard-van der Laan, M.-H., Bed'Hom, B., Juul-Madsen, H.R., 2015. Annotation and genetic diversity of the chicken collagenous lectins. Mol. Immunol. 65 (2), 277–286.
Hansen, S.W., Ohtani, K., Roy, N., Wakamiya, N., 2016. The collectins CL-L1, CL-K1 and CL-P1, and their roles in complement and innate immunity. Immunobiology 221 (10), 1058–1067.
Hart, B.L., 2011. Behavioural defences in animals against pathogens and parasites: parallels with the pillars of medicine in humans. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 366 (1583), 3406–3417.
Hartshorn, K.L., Crouch, E.C., White, M.R., Eggleton, P., Tauber, A.I., Chang, D., Sastry, K., 1994. Evidence for a protective role of pulmonary surfactant protein D (SP-D) against influenza A viruses. J. Clin. Invest. 94 (1), 311–319.
Hartshorn, K.L., White, M.R., Shepherd, V., Reid, K., Jensenius, J.C., Crouch, E.C., 1997. Mechanisms of anti-influenza activity of surfactant proteins A and D: comparison with serum collectins. Am. J. Physiol. Lung Cell Mol. Physiol. 273 (6), L1156–L1166.
Hartshorn, K.L., White, M.R., Smith, K., Sorensen, G., Kuroki, Y., Holmskov, U., Head, J., Crouch, E.C., 2010a. Increasing antiviral activity of surfactant protein d trimers by introducing residues from bovine serum collectins: dissociation of mannan-binding and antiviral activity. Scand. J. Immunol. 72 (1), 22–30.
Hartshorn, K.L., White, M.R., Tecle, T., Sorensen, G., Holmskov, U., Crouch, E.C., 2010b. Viral aggregating and opsonizing activity in collectin trimers. Am. J. Physiol. Lung Cell Mol. Physiol. 298 (1), L79–L88.
Haurum, J., Thiel, S., Haagsman, H., Laursen, S., Larsen, B., Jensenius, J., 1993. Studies on the carbohydrate-binding characteristics of human pulmonary surfactant-associated protein A and comparison with two other collectins: mannan-binding protein and conglutinin. Biochem. J. 293 (3), 873–878.
Head, J.F., Mealy, T.R., McCormack, F.X., Seaton, B.A., 2003. Crystal structure of trimeric carbohydrate recognition and neck domains of surfactant protein A. J. Biol. Chem. 278 (44), 43254–43260.
Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.
Hogenkamp, A., 2007. Expression Levels of Chicken (Col)lectin mRNA during Infection with Influenza A Virus or Infectious Bronchitis Virus Superinfected with Escherichia coli PhD. Utrecht University.
Hogenkamp, A., Isohadouten, N., Reemers, S.S.N., Romjin, R.A., Hemrika, W., White, M. R., Tefsen, B., Vervelde, L., van Eijk, M., Veldhuizen, E.J.A., Haagsman, H.P., 2008. Chicken lung lectin is a functional C-type lectin and inhibits haemagglutination by influenza A virus. Vet. Microbiol. 130 (1–2), 37–46.
Hogenkamp, A., van Eijk, M., van Djik, A., van Asten, A., Veldhuizen, E.J.A., Haagsman, H.P., 2006. Characterization and expression sites of newly identified chicken collectins. Mol. Immunol. 43 (10), 1604–1616.
Holmskov, U., Malhotra, R., Sim, R.B., Jensenius, J.C., 1994. Collectins: collagenous C-type lectins of the innate immune defense system. Immunol. Today 15 (2), 67–74.
Holmskov, U., Thiel, S., Jensenius, J.C., 2003. Collectins and ficolins: humoral lectins of the innate immune defense. Annu. Rev. Immunol. 21, 547–578.
Holmskov, U.L., 2000. Collectins and collectin receptors in innate immunity. APMIS (Acta Pathol. Microbiol. Immunol. Scand.) Suppl. 108 (S100), 7–59.
Hoppe, H.J., Barlow, P.N., Reid, K.B., 1994. A parallel three stranded alpha-helical bundle at the nucleation site of collagen triple-helix formation. FEBS (Fed. Eur. Biochem. Soc.) Lett. 344 (2–3), 191–195.
Hoppe, H.J., Reid, K.B.M., 1994. Collectins-soluble proteins containing collagenous regions and lectin domains-and their roles in innate immunity. Protein Sci. 3 (8), 1143–1158.
Hughes, A.L., 2007. Evolution of the lung surfactant proteins in birds and mammals. Immunogenetics 59 (7), 565–572.
Ikegami, M., Elhalwagi, B.M., Palaniyar, N., Dienger, K., Korfhagen, T., Whitsett, J.A., McCormack, F.X., 2001. The collagen-like region of surfactant protein A (SP-A) is required for correction of surfactant structural and functional defects in the SP-A null mouse. J. Biol. Chem. 276 (42), 38542–38548.
Jones, J.H., Effmann, E.L., Schmidt-Nielsen, K., 1985. Lung volume changes during respiration in ducks. Respir. Physiol. 59 (1), 15–25.
Karinch, A.M., Floros, J., 1995. 5' splicing and allelic variants of the human pulmonary surfactant protein A genes. Am. J. Respir. Cell Mol. Biol. 12 (1), 77–88.
King, R.J., Clements, J.A., 1972. Surface active materials from dog lung. II. Composition and physiological correlations. Am. J. Physiol. 223 (3), 715–726.
King, R.J., Klass, D.J., Gikas, E.G., Clements, J.A., 1973. Isolation of apoproteins from canine surface active material. Am. J. Physiol. 224 (4), 788–795.
Kingma, P.S., Whitsett, J.A., 2006. In defense of the lung: surfactant protein A and surfactant protein D. Curr. Opin. Pharmacol. 6 (3), 277–283.
Kishore, U., Greenhough, T.J., Waters, P., Shrive, A.K., Ghai, R., Kamran, M.F., Bernal, A. L., Reid, K.B., Madan, T., Chakraborty, T., 2006. Surfactant proteins SP-A and SP-D: structure, function and receptors. Mol. Immunol. 43 (9), 1293–1315.
Kishore, U., Reid, K.B., 2001. Structures and functions of mammalian collectins. Results Probl. Cell Differ. 33, 225–248.
Korfhagen, T.R., Bruno, M.D., Ross, G.F., Huelsman, K.M., Ikegami, M., Jobe, A.H., Wert, S.E., Stripp, B.R., Morris, R.E., Glasser, S.W., 1996. Altered surfactant function and structure in SP-A gene targeted mice. Proc. Natl. Acad. Sci. USA 93 (18), 9594–9599.
Kuroki, Y., Akino, T., 1991. Pulmonary surfactant protein A (SP-A) specifically binds dipalmitoylphosphatidylcholine. J. Biol. Chem. 266 (5), 3068–3073.
Lopez-Rodriguez, E., Pascual, A., Arroyo, R., Floros, J., Perez-Gil, J., 2016. Human pulmonary surfactant protein SP-A1 provides maximal efficiency of lung interfacial films. Biophys. J. 111 (3), 524–536.
Malhotra, R., Haurum, J., Thiel, S., Sim, R.B., 1992. Interaction of C1q receptor with lung surfactant protein A. Eur. J. Immunol. 22 (6), 1437–1445.
Manarolla, G., Liandris, E., Pisoni, G., Sassera, D., Grilli, G., Gallazzi, D., Sironi, G., Moroni, P., Piccinini, R., Rampin, T., 2009. Avian mycobacteriosis in companion birds: 20-year survey. Vet. Microbiol. 133 (4), 323–327.
McCormack, F.X., 1998. Structure, processing and properties of surfactant protein A. Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1408 (2–3), 109–131.
Munye, M.M., Diaz-Font, A., Ocaka, L., Henriksen, M.L., Lees, M., Brady, A., Jenkins, D., Morton, J., Hansen, S.W., Bacchelli, C., 2017. COLEC10 is mutated in 3MC patients and regulates early craniofacial development. PLoS Genet. 13 (3), e1006679.
Olde Nordkamp J. M., Marloes, van Eijk, Martin, Urbanus T., Rolf, Bont, Louis, Haagsman P., Henk, Meyaard, Linde, 2014. Leukocyte-associated Ig-like receptor-1 is a novel inhibitory receptor for surfactant protein D. J. Leukoc. Biol. 96 (1), 105–111. https://doi.org/10.1189/jlb.3AB0213-092RR.
Orgeig, S., Hiemstra, P.S., Veldhuizen, E.J.A., Casals, C., Clark, H.W., Haczku, A., Knudsen, L., Possmayer, F., 2010. Recent advances in alveolar biology: evolution and function of alveolar proteins. Respir. Physiol. Neurobiol. 173, S43–S54.
Palaniyar, N., Ikegami, M., Korfhagen, T., Whitsett, J., McCormack, F.X., 2001. Domains of surfactant protein A that affect protein oligomerization, lipid structure and surface tension. Comp. Biochem. Physiol. Mol. Integr. Physiol. 129 (1), 109–127.
Perez-Gil, J., 2008. Structure of pulmonary surfactant membranes and films: the role of proteins and lipid–protein interactions. Biochim. Biophys. Acta Biomembr. 1778 (7–8), 1676–1695.
Perry, G.H., 2014. Parasites and human evolution. Evol. Anthropol. Issues News Rev. 23 (6), 218–228.
Persson, A., Chang, D., Crouch, E., 1990. Surfactant protein D is a divalent cation-dependent carbohydrate-binding protein. J. Biol. Chem. 265 (10), 5755–5760.
Pillai, S.P., Pantin-Jackwood, M., Yassine, H.M., Saif, Y.M., Lee, C.W., 2010. The high susceptibility of turkeys to influenza viruses of different origins implies their importance as potential intermediate hosts. Avian Dis. 54 (1 Suppl. l), 522–526.
Pison, U., Max, M., Neuendank, A., Weissbach, S., Pietschmann, S., 1994. Host defence capacities of pulmonary surfactant: evidence for 'non-surfactant'functions of the surfactant system. Eur. J. Clin. Invest. 24 (9), 586–599.
Pison, U., Shiffer, K., Hawgood, S., Goerke, J., 1990. Effects of the surfactant-associated proteins, SP-A, SP-B and SP-C, on phospholipid surface film formation. Basic Research on Lung Surfactant 25, 271–273.
Poirotte, C., Massol, F., Herbert, A., Willaume, E., Bomo, P.M., Kappeler, P.M., Charpentier, M.J.E., 2017. Mandrills use olfaction to socially avoid parasitized conspecifics. Sci. Adv. 3 (4).
Possmayer, F., 1988. A proposed nomenclature for pulmonary surfactant-associated proteins. Am. Rev. Respir. Dis. 138 (4), 990–998.

Possmayer, F., 1990. The role of surfactant-associated proteins. Am. Rev. Respir. Dis. 142, 749–752.
Reading, P.C., Morey, L.S., Crouch, E.C., Anders, E.M., 1997. Collectin-mediated antiviral host defense of the lung: evidence from influenza virus infection of mice. J. Virol. 71 (11), 8204–8212.
Reemers, S.S., Veldhuizen, E.J.A., Fleming, C., van Haarlem, D.A., Haagsman, H., Vervelde, L., 2010. Transcriptional expression levels of chicken collectins are affected by avian influenza A virus inoculation. Vet. Microbiol. 141 (3–4), 379–384.
Rust, K., Grosso, L., Zhang, V., Chang, D., Persson, A., Longmore, W., Cai, G.-Z., Crouch, E., 1991. Human surfactant protein D: SP-D contains a C-type lectin carbohydrate recognition domain. Arch. Biochem. Biophys. 290 (1), 116–126.
Rynkiewicz, M.J., Wu, H., Cafarella, T.R., Nikolaidis, N.M., Head, J.F., Seaton, B.A., McCormack, F.X., 2017. Differential ligand binding specificities of the pulmonary collectins are determined by the conformational freedom of a surface loop. Biochemistry 56 (31), 4095–4105.
Sanchez-Barbero, F., Strassner, J., García-Cañero, R., Steinhilber, W., Casals, C., 2005. Role of the degree of oligomerization in the structure and function of human surfactant protein A. J. Biol. Chem. 280 (9), 7659–7670.
Schachner, E.R., Cieri, R.L., Butler, J.P., Farmer, C.G., 2014. Unidirectional pulmonary airflow patterns in the savannah monitor lizard. Nature 506 (7488), 367–370.
Schachner, E.R., Hutchinson, J.R., Farmer, C., 2013. Pulmonary anatomy in the Nile crocodile and the evolution of unidirectional airflow in Archosauria. PeerJ 1, e60.
Schurch, S., Possmayer, F., Cheng, S., Cockshutt, A.M., 1992. Pulmonary SP-A enhances adsorption and appears to induce surface sorting of lipid extract surfactant. Am. J. Physiol. Lung Cell Mol. Physiol. 263 (2), L210–L218.
Seaton, B.A., Crouch, E.C., McCormack, F.X., Head, J.F., Hartshorn, K.L., Mendelsohn, R., 2010. Structural determinants of pattern recognition by lung collectins. Innate Immun. 16 (3), 143–150.
Shang, F., Rynkiewicz, M.J., McCormack, F.X., Wu, H., Cafarella, T.M., Head, J.F., Seaton, B.A., 2011. Crystallographic complexes of surfactant protein A and carbohydrates reveal ligand-induced conformational change. J. Biol. Chem. 286 (1), 757–765.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7 (1), 539.
Tamura, K., Stecher, G., Kumar, Su., 2021. MEGA11: Molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38 (7), 3022–3027. https://doi.org/10.1093/molbev/msab120. Submitted for publication.
Tenner, A.J., Robinson, S.L., Borchelt, J., Wright, J.R., 1989. Human pulmonary surfactant protein (SP-A), a protein structurally homologous to C1q, can enhance FcR-and CR1-mediated phagocytosis. J. Biol. Chem. 264 (23), 13923–13928.
Thiel, S., Reid, K.B.M., 1989. Structures and functions associated with the group of mammalian lectins containing collagen-like sequences. FEBS (Fed. Eur. Biochem. Soc.) Lett. 250 (1), 78–84.
van de Wetering, J.K., van Golde, L.M.G., Batenburg, J.J., 2004. Collectins - players of the innate immune system. Eur. J. Biochem. 271 (7), 1229–1249.
van Eijk, M., Rynkiewicz, M.J., White, M.R., Hartshorn, K.L., Zou, X., Schulten, K., Luo, D., Crouch, E.C., Cafarella, T.R., Head, J.F., Haagsman, H.P., Seaton, B.A., 2012. A unique sugar-binding site mediates the distinct anti-influenza activity of pig surfactant protein D. J. Biol. Chem. 287 (32), 26666–26677.
van Golde, L., Batenburg, J.J., Robertson, B., 1988. The pulmonary surfactant system: biochemical aspects and functional significance. Physiol. Rev. 68 (2), 374–455.
Veldhuizen, E.J.A., van Eijk, M., Haagsman, H.P., 2011. The carbohydrate recognition domain of collectins. FEBS J. 278 (20), 3930–3941.
Voss, T., Eistetter, H., Schäfer, K.P., Engel, J., 1988. Macromolecular organization of natural and recombinant lung surfactant protein SP 28-36 : structural homology with the complement factor C1q. J. Mol. Biol. 201 (1), 219–227.
Wang, G., Guo, X., DiAngelo, S., Thomas, N.J., Floros, J., 2010. Humanized SFTPA1 and SFTPA2 transgenic mice reveal functional divergence of SP-A1 and SP-A2 formation of tubular myelin in vivo requires both gene products. J. Biol. Chem. 285 (16), 11998–12010.
Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Kunstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., Heger, A., Kong, L., Ponting, C.P., Jarvis, E.D., Mello, C.V., Minx, P., Lovell, P., Velho, T.A.F., Ferris, M., Balakrishnan, C.N., Sinha, S., Blatti, C., London, S.E., Li, Y., Lin, Y.-C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backstrom, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A.R., Howard, J., Volker, M., Skinner, B.M., Griffin, D.K., Ye, L., McLaren, W.M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X.S., Olender, T., Lancet, D., Smit, A.F.A., Hubley, R., Konkel, M.K., Walker, J.A., Batzer, M.A., Gu, W., Pollock, D.D., Chen, L., Cheng, Z., Eichler, E.E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S.V., Yang, S.-P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E.R., Wilson, R.K., 2010. The genome of a songbird. Nature 464 (7289), 757–762.
Weebadda, W.K.C., Hoover, G.J., Hunter, D.B., Hayes, M.A., 2001. Avian air sac and plasma proteins that bind surface polysaccharides of Escherichia coli O2. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 130 (3), 299–312.
Weis, W., Crichlow, G., Murthy, H., Hendrickson, W., Drickamer, K., 1991. Physical characterization and crystallization of the carbohydrate-recognition domain of a mannose-binding protein from rat. J. Biol. Chem. 266 (31), 20678–20686.
Whitsett, J.A., Hull, W., Ross, G., Weaver, T., 1985. Characteristics of human surfactant-associated glycoproteins A. Pediatr. Res. 19 (5), 501–508.
Woodward, J., Maina, J., 2008. Study of the structure of the air and blood capillaries of the gas exchange tissue of the avian lung by serial section three-dimensional reconstruction. J. Microsc. 230 (1), 84–93.
Wright, J.R., Wager, R.E., Hawgood, S., Dobbs, L., Clements, J.A., 1987. Surfactant apoprotein Mr = 26,000-36,000 enhances uptake of liposomes by type II cells. J. Biol. Chem. 262 (6), 2888–2894.
Yang, L., Carrillo, M., Wu, Y.M., DiAngelo, S.L., Silveyra, P., Umstead, T.M., Halstead, E. S., Davies, M.L., Hu, S., Floros, J., 2015. SP-R210 (Myo18A) isoforms as intrinsic modulators of macrophage priming and activation. PLoS One 10 (5), e0126576.
Yu, S.-H., McCormack, F.X., Voelker, D.R., Possmayer, F., 1999. Interactions of pulmonary surfactant protein SP-A with monolayers of dipalmitoylphosphatidylcholine and cholesterol: roles of SP-A domains. JLR (J. Lipid Res.) 40 (5), 920–929.
Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. FEBS J. 272 (24), 6179–6217.
Zhang, W., Cuperus, T., van Dijk, A., Skjødt, K., Hansen, S., Haagsman, H.P., Veldhuizen, E.J.A., 2016. Developmental regulation of chicken surfactant protein A and its localization in lung. Dev. Comp. Immunol. 61, 80–87.