Is lung surfactant protein a lectin – collagen hybrid?

SIR—Pulmonary surfactant lining the surface of alveoli is a complex mixture of lipids, carbohydrates and proteins. This lipoprotein complex is able to decrease surface tension at the air–liquid interface and thus helps maintain alveolar patency at low lung volumes (see ref. 1 for review). Recently the cDNA sequence of the major protein of canine and human pulmonary surfactant has been published; the deduced protein sequence was found to be rather unusual in that it predicts that the protein has a collagen-like region within its amino-terminal half and a non-collagenous globular domain within its

carboxy-terminal half^{2,3}. Although these features are reminiscent of the structures of collagen type IV and C1q^{4,5}, the globular domain of the surfactants seems to be unrelated to those of the other collagenous proteins.

In the gene of the human pulmonary surfactant-associated protein⁶ an individual exon codes for the carboxy-terminal globular domain, raising the possibility that exon-shuffling was responsible for the introduction of this exon into the gene of a collagen-like ancestor. This possibility prompted a comparison, reported here, with modules that have been used in the assembly of various vertebrate proteins^{7–9}.

Sequence comparison revealed that the

non-collagenous domain of pulmonary surfactant resembles a module that is present in hepatic lectins of various vertebrate species^{10–12}, the humoral lectin of the flesh-fly, *Sarcophaga peregrina*¹³, and rat cartilage proteoglycan⁹. The homology of the globular domain of canine and human pulmonary surfactant-associated protein with this module is underlined by the fact that of the 17 residues that are invariant among the various members of this module-family, 16 are conserved (identical or chemically similar) in both canine and human surfactant protein, as seen in the figure.

It should also be noted that in the alignment of the two pulmonary proteins the gaps coincide with the gaps dictated by the alignment of the various lectins and proteoglycan. In the region containing the four conserved cysteines (Cys²⁵–Cys¹²⁰) the homology with the various lectins and proteoglycans is 28–34 per cent. The homology of human and canine surfactant protein with animal lectins and proteoglycan is also supported by the high alignment scores (4.66–8.81 standard deviation units) determined with the ALIGN program¹⁴ (kindly carried out by Dr K.B.M. Reid, Oxford).

The similarity of the globular domain of pulmonary surfactant protein to a module common to various animal lectins raises the possibility that this protein may be involved in interactions with carbohydrate constituents of the lung surfactant complex.

LÁSZLÓ PATTHY

*Institute of Enzymology,
Biological Research Centre,
Hungarian Academy of Sciences,
Budapest, PO Box 7,
Hungary H-1502*

- Goerke, J. *Biochim. biophys. Acta* **344**, 241–261 (1974).
- Benson, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6379–6383 (1985).
- Floros, J. *et al. J. biol. Chem.* **261**, 9029–9033 (1986).
- Oberbaumer, I. *et al. Eur. J. Biochem.* **147**, 217–224 (1985).
- Reid, K.B.M. *Biochem. J.* **231**, 729–735 (1985).
- White, R.T. *et al. Nature* **317**, 361–363 (1985).
- Patthy, L. *Cell* **41**, 657–663 (1985).
- Südhof, T.C. *et al. Science* **228**, 815–822 (1985).
- Doerge, K. *et al. J. biol. Chem.* **261**, 8108–8111 (1985).
- Drickamer, K. *J. biol. Chem.* **256**, 5827–5839 (1981).
- Drickamer, K. *et al. J. biol. Chem.* **259**, 770–778 (1984).
- Spies, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6465–6469 (1985).
- Takahashi, H. *et al. J. biol. Chem.* **260**, 12228–12233 (1985).
- Dayhoff, M.O. *et al. Meth. Enzym.* **91**, 524–545 (1983).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

	1	10	20	30	40	
cPSAP	V I S L H E S I L V V G R K V I S S N G Q S I N F N D I Q E L C A G A G G Q I A A P M S P E E N					
hPSAP	A I S L Q G S I M I T V G L K V F S S N G Q S I T F D A I Q E A C A R A G G R I A V P R N P E E N					
rASGPR 1	R T C C P I N H W V E Y E G S C Y W F S S V K P W T E A D K Y C Q L E N A H L V V V T S W E E Q					
rASGPR 2				E N A H L L V I N S R E E Q		
hASGPR 1	R T C C P V N W V E H E R S C Y W F S R S G K A W A D A D N Y C R L E D A H L V V V T S W E E Q					
hASGPR 2	R T C C P V N W V E H Q G S C Y W F S H S G K A W A E A E K Y C Q L E N A H L V V I N S W E E Q					
PG	Q E Q C E E G W I K F Q G H C Y R H F P U R E I W V D A E R R C R E Q Q S H L S S I V T P E E Q					
chL	C G A Q S R D W E Y F E G R C Y Y F S L R M S W H K A K A E C E M H S H L I I I D S Y A K Q					
SpL	A A V P D L Q K A L D G R T Y L I E T E L K Y N W H Q A W H E C A R H D Q D L V T I E S A D K N					
Consensus			W	A	C	L

	50	60	70		
cPSAP	E A V A S I V K K Y N T Y A	Y L G L V E S P S G D	F Q Y M D G A P V N		
hPSAP	E A I A S F Y K K Y N T Y A	Y V G L T E G P S P G D	F R Y S D G T P V N		
rASGPR 1	R F V U Q H M G P L N T	W I G L T D Q N G P	W K W V D G T D Y E T G		
rASGPR 2	F F V V K H R G A F H I	W I G L T D K D G S	W K W V D G T E Y R S N		
hASGPR 1	K F V Q H H I G P V N T	W M G L H D Q D G P	W K W V D G T D Y E T G		
hASGPR 2	K F I V Q H I N P F N T	W I G L T D S G G S	W K W V D G T D Y R H N		
PG	E F V N K N A D D Y Q	W I G L N D R T I E G D	F R W S D G H S L Q		
chL	N F V M F R T R N E R F	W I G L T D E N Q E G E	W W V D G T D T R S S		
SpL	N A I T D L V K R V V G K K S I N L	W L G G N D E Y S S G R D Y G R P F F W S P T G Q A F S			
Consensus		W	G	D	W

	80	90	100	110	120					
cPSAP	Y I N W Y P G E P R G R G K	E Q C V E M Y T D G	Q W N N K N C L Q Y R L A I C E F							
hPSAP	Y I N W Y R G E P A G R G K	E Q C V E M Y T D G	Q W N D R N C L Y S R L I C E F							
rASGPR 1	F K N W R P G P D D W Y G I G L G G G E F C A H F T D G		H W N D V F C R P Y R W V C E T							
rASGPR 2	F K I W A F T F P D N C Q G H E F G G S F D C A E I L S D G		L W N D R F C Q D V H R W A C E R							
hASGPR 1	F K N W R P F O P D W Y G I G L G G G E F C A H F T D G		R W N D V F C Q R P Y R W V C E T							
hASGPR 2	N R N W A V I O P D N W H G H E L G G S E D C V E V Q P D G		R W N D F C L Q V Y R W V C E K							
PG	F E K W R P N D P N F F A I G	F D C V W I W H E R G	E W N D V F C N Y Q L P F T C K K							
chL	F F F K F G E P N R G F N	F D C A H V W I S G	Q W N D V C T Y E C Y Y V C E K							
SpL	F A Y W S E N D P N Y K H Q	E H C H I W D T K P L Y	Q W N D N C D N V K M G Y I C E P							
Consensus		W	P	F	C	W	N	D	C	C

Alignment of the sequences of canine and human pulmonary surfactant-associated proteins (cPSAP, hPSAP) with the sequences of rat and human hepatic asialoglycoprotein receptor (rASGPR, hASGPR), rat cartilage proteoglycan (PG), chicken hepatic lectin (chL) and the humoral lectin of *Sarcophaga peregrina* (SpL). The sequence of hPSAP shown in the alignment corresponds to the translated region of the fifth exon of the hPSAP gene. The residues shown in the last lines of the alignment are the consensus residues invariant in animal lectins and proteoglycan. The residues identical with those of cPSAP and hPSAP are boxed. The numbering of residues refers to the homology unit of asialoglycoprotein receptors and starts at the first tryptophan conserved in vertebrate lectins and proteoglycan. The sequences correspond to: amino acids 107–231 of the 231 residue mature cPSAP; amino acids 104–228 of the 228 residue mature hPSAP; segment 149–277 of the 283 residue rASGPR 1; segment 151–279 of the 291 residue hASGPR 1; amino acids 56–181 of the carboxyl-terminal 269 residues of PG; amino acids 78–203 of the 207 residues of chL and amino acids 22–159 of the 283 residue SpL.