Two Rat Surfactant Protein A Isoforms Arise by a Novel Mechanism That Includes Alternative Translation Initiation†

Mamatha Damodarasamy, Mei Zhang, Krista Dienger, and Francis X. McCormack*

Division of Pulmonary/Critical Care Medicine, Department of Internal Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267

Received February 28, 2000; Revised Manuscript Received May 18, 2000

ABSTRACT: A single gene for rat surfactant protein A (SP-A) encodes two isoforms that are distinguished by an isoleucine−lysine−cysteine (IKC) N-terminal extension (SP-A and IKC-SP-A). Available evidence suggests that the variants are generated by alternative signal peptidase cleavage of the nascent polypeptide at a primary site (Cys^{-1}−Asn^{1}) and a secondary site (Gly^{-4}−Ile^{-3}). In this study, we used site-directed mutagenesis and heterologous expression in vitro and in insect cells to the examine mechanisms that may lead to alternative signal peptidase cleavage including alternative translation initiation at two in-frame AUGs (Met^{-30} and Met^{-20}), a suboptimal context for hydrolysis at the primary cleavage site, or cotranslational protein modifications that expose an otherwise cryptic secondary cleavage site. In vitro translation of a rat cDNA for SP-A resulted in both 28 and 29 kDa primary translation products on SDS−PAGE analysis, while translation of cDNAs encoding Met−30Ala and Met−20Ala mutations resulted in only the single 28 and 29 kDa molecular mass species, respectively. These data are consistent with translation initiation at both Met^{-30} and Met^{-20} during in vitro synthesis of SP-A. The Met−30Ala mutation reduced expression of the longer isoform in insect cells, indicating that the Met^{-30} site also contributes to eucaryotic protein expression. Forcing translation initiation at Met^{-30} by optimizing the Kozak consensus sequence surrounding that codon or by mutating the Met^{-20} codon resulted in preferential expression of the longer SP-A isoform but reduced overall expression of the protein almost 10-fold. Both isoforms were generated to some degree whether translation was initiated at the codon for Met^{-30} or Met^{-20}, indicating that the site of translation initiation is not the sole determinant of isoform generation and suggesting that either the context of the primary cleavage site is suboptimal or that cotranslational modifications affect cleavage. Preventing N-terminal glycosylation at Asn^{1} did not affect the site of signal peptidase cleavage. Disruption of interchain disulfide formation at Cys^{-1} by substitution with serine markedly enhanced cleavage at the Gly^{-4}−Ile^{-3} bond, but substitution with alanine enhanced cleavage at the Cys^{-1}−Asn^{1} bond. We conclude that rat SP-A isoforms arise by a novel mechanism that includes both alternative translation initiation at two in-frame AUGs and a suboptimal context for signal peptidase hydrolysis at the primary cleavage site.

Pulmonary surfactant protein A is an oligomeric glycoprotein, composed of an N-terminal collagen-like domain (CLD)[1] and a C-terminal carbohydrate recognition domain (CRD), which has been reported to play important roles in host defense against inhaled pathogens and in the maintenance of surfactant integrity (for review see ref *1*). These functions are mediated by Ca^{2+}-dependent ligand binding interactions between the CRD and the carbohydrates which decorate the surface of microorganisms or the phospholipid components of surfactant. Ligand binding affinity is greatly enhanced by the multimeric organization of the protein, which allows for simultaneous interaction with several sites on multivalent ligands or membranes. The N-terminal domains of SP-A contain the structural elements which are required for oligomeric assembly, including a collagen-like region of Gly-X-Y repeats that form trimers by folding into triple helices, and a "neck" region bridging the CLD and CRD that most likely forms bundled triple α-helical coiled coils. Six trimers are laterally associated through the first half of the CLD, forming a bouquet-like structure with 18 CRDs arrayed on a rigid, collagenous stalk (*2*). The molecule is stabilized by interchain disulfide linkage near the N-terminus, at Cys^{6} and at a newly identified Cys^{-1} within a minor isoform of SP-A (*5*). This slightly elongated variant contains three amino acids (Ile-Lys-Cys or IKC) within an N-terminal extension that was formerly assigned to the C-terminal end of the signal peptide. The interchain disulfide bond at Cys^{-1} is clearly required for the covalent association of six or more subunits of SP-A that is apparent on nonreducing SDS−PAGE and is likely to be important for SP-A functions which are critically dependent on full octadecameric assembly (*5*).

† This research was supported by Grant HL 61612, the Medical Research Service of the Department of Veteran Affairs, and NIH Grant P30ES-0609606.

* To whom correspondence should be addressed. Tel: 513-558-4831. Fax: 513-558-4858. E-mail: frank.mccormack@uc.edu.

[1] Abbreviations: SP-A, surfactant protein A; CRD, carbohydrate recognition domain; CLD, collagen-like domain; IKC, isoleucine−lysine−cysteine; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; $SP\text{-}A^{hyp}$, hydroxyproline-deficient recombinant SP-A; PTH, phenylthiohydantoin; wt, wild type.

10.1021/bi000448c CCC: $19.00
Published on Web 07/27/2000

Table 1: Mutagenic Oligonucleotides

Met−30Ala	5′-AGGCTCTGTGCGTGGCAGAAG-3′ [a]
Met−20Ala	5′-ATAGTAGCCGCGTCACTGTGT-3′ [a]
Kozak A	5′-GCCACCATGTGGCAGAAGCCACTG-3′ [a]
Kozak B	5′-GCCACCATGGGGCAGAAGCCACTG-3′ [a]
Cys−1Ala	5′-GGTATCAAGGCCAATGTGACA-3′ [b]
Asn1Thr	5′-ATCAAGTGCACTGTGACAGAC-3′ [b]
Thr3Ala	5′-TGCAATGTGGCAGACGTTTGT-3′ [b]

[a] The underline denotes native or mutated AUG initiator codons. [b]The underline denotes the codon for the substituted amino acid.

We had previously reported that relative expression of the two SP-A isoforms was sensitive to the context of amino acid sequences downstream from the leader peptide in several mutant recombinant proteins that were examined (*4*). This is a well-characterized property of signal peptidases (*6*) and suggested that the microheterogeneity at the N-terminus of SP-A was most likely due to alternative cleavage by signal peptidase at primary (Cys^{-1}—Asn^{1}) and secondary (Gly^{-4}—Ile^{-3}) cleavage sites (*4*). Factors that may affect signal peptidase specificity of SP-A include alternative translation initiation at two closely spaced codons for methionine within the leader peptide (Met^{-30} and Met^{-20}) or a suboptimal or ambiguous context for signal peptidase cleavage. During the characterization of a mutant recombinant SP-A containing a Cys−1Ser mutation, we found that the amino acid substitution had a dramatic effect on the isoform composition of the protein despite the fact that Ser is a common residue at the −1 position relative to the signal peptidase cleavage site (*3*, *4*). This result suggested the possibility that cotranslational disulfide bond formation, or perhaps other cotranslational protein modifications, may affect signal peptidase cleavage specificity. We were unable to find a precedent for this possibility among known secreted proteins, but the unusual structural features of an N-glycosylated Asn at the +1 position and a disulfide-forming Cys at the −1 position of rat (and mouse) SP-A provided an opportunity to examine the relationship of these protein modifications to signal peptidase cleavage.

To more fully characterize the mechanisms involved in alternative signal peptidase cleavage during synthesis of SP-A, we assessed the effects of targeted mutations in the leader sequence and adjacent N-terminal domains of SP-A on the isoform composition of the protein.

EXPERIMENTAL PROCEDURES

Production of Mutant Recombinant Proteins. Mutant cDNAs were produced from a 1.6 kb rat SP-A cDNA template (*7*) by overlapping extension PCR (*8*), as described (*9*). The mutagenic oligonucleotides that were used to introduce the amino acid substitutions and deletions in this study are outlined in Table 1. The mutated cDNAs were ligated into the *Eco*RI site of the PVL 1392 transfer vector (Invitrogen), and the correct orientations were confirmed by restriction analysis with *Kpn*I. Nucleotide sequencing of the entire coding region for all PCR amplified sequences was performed to confirm the intended deletion or substitution and the absence of spurious mutations (*10*). Recombinant baculoviruses containing the mutant cDNAs were produced by homologous recombination in *Spodoptera frugiperda* (Sf-9) cells following cotransfection with linear viral DNA and the PVL 1392/mutant SP-A constructs (Baculogold, Pharmingen), as described (*9*). Fresh monolayers of 10^7 *Trichoplusia ni* (*T. ni*) cells were infected with plaque-purified recombinant viruses at a multiplicity of infection (MOI) of 10 and then incubated with serum-free Excel 400 media (JRH Biosciences) supplemented with antibiotics for 72 h. Recombinant SP-A was purified from the culture media by adsorption to mannose−Sepharose 6B columns in the presence of 1−10 mM Ca^{2+} and elution with 2 mM EDTA (*11*). The purified recombinant SP-A was dialyzed against 5 mM Tris (pH 7.4) or passed through a G-25 column (PD-10, Pharmacia) to remove EDTA and then stored at −20 °C.

Protein Analysis. Recombinant SP-A levels in tissue culture media were determined with a rabbit polyclonal IgG against rat SP-A using a sandwich ELISA (*12*). The lower limit of sensitivity of the assay was 0.20 ng/mL, and the linear range extended from 0.16 to 10.0 ng/mL. Routine protein concentrations were determined with the bicinchoninic protein assay kit (BCA) (Pierce) using bovine serum albumin as the standard. Purified SP-A proteins were separated by 8−16% SDS−PAGE and stained with Coomassie blue (*13*).

In Vitro Transcription/Translation. The cDNAs for wild-type rat SP-A, Kozak A, Met−20Ala, and Met−30Ala were ligated into the multiple cloning site of pGEM 4Z (Promega). In vitro transcription and translation were performed sequentially with a cell-free and microsome-free, rabbit reticulocyte lysate based system (PROTEINscript II, Ambion). The ^{35}S-labeled translation products were then size-fractionated on 16% SDS−PAGE gels and placed on film.

Protein Microsequencing. The N-terminal amino acid sequences of native and recombinant SP-As were determined by modified Edman degradation on an automated gas-phase microsequencer (Applied Biosystems 470A) (*14*, *15*). The elution of phenylthiohydantoin (PTH) derivatized amino acids was monitored with a Model 120A PTH analyzer. The lower limit of detection for this method is about 1 pmol of protein. The relative abundance of the SP-A isoforms was estimated by comparison of the yield of representative amino acids from the primary and secondary sequences in each run. Residues in the second or third cycle were usually selected for quantitation, to avoid the higher background in the first cycle and the declining yields due to technical limitations in subsequent cycles.

RESULTS

Description of Mutant cDNAs Produced. To better understand the mechanism of rat SP-A isoform formation, we produced a set of mutant SP-As with ablated translation initiation sites in the leader peptide or substitutions in the posttranslationally modified N-terminal amino acids of the mature protein which may affect signal peptidase cleavage (Figure 1). The mutant cDNAs which were developed encoded (1) alanine substitutions for methionine at positions −20 (Met−20Ala) or −30 (Met−30Ala) to force translation initiation at Met^{-30} and Met^{-20}, respectively, (2) a near optimal Kozak consensus sequence for translation initiation surrounding the codon for Met^{-30} (Kozak A), (3) an optimal Kozak consensus sequence at the Met^{-30} at the cost of a tryptophan to glycine substitution at the −29 position (Kozak B), (4) a Cys−1Ala substitution to prevent interchain

		M^{-30}	W	*Q	K	P	L	G	I	V	A	M^{-20}
GTC	TGT	ATG	TGG	CAG	AAG	CCA	CTG	GGG	ATA	GTA	GCG	ATG
		S	L	C	S	L	A	F	T	L	F	L
		TCA	CTG	TGT	TCT	TTG	GCG	TTC	ACC	CTC	TTC	TTG
		T	V	V	A	G^{-4} ↓	I^{-3}	K	C^{-1} ↓	N^{1}	V	T
		ACT	GTT	GTC	GCT	GGT	ATC	AAG	TGC	AAT	GTG	ACA
		D	V	C^{6}	A							
		GAC	GTT	TGT	GCT							

FIGURE 1: Leader peptide and N-terminal sequences of rat SP-A. Superscripted numbers denote amino acid sequence positions relative to the N-terminus of the major rat SP-A isoform (Asn^1). Cleavage sites for signal peptidase at Gly^{-4}−Ile^{-3} and Cys^{-1}−Asn^1 are indicated with arrows. Translation initiation codons Met^{-30} and Met^{-20} are in boldface. Previously unreported Gln^{-28} is marked with an asterisk.

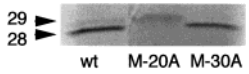

FIGURE 2: In vitro translation of mutant SP-A cDNAs. cDNAs for SP-A encoding $SP\text{-}A^{hyp}$ (wt) and SP-As with alanine substitutions for initiating methionines Met^{-30} and Met^{-20} were transcribed and translated in vitro, separated on 16% SDS−PAGE gels, and placed on film.

disulfide bond formation at the −1 position, and (5) a Thr3Ala substitution to prevent N-glycosylation at the N-terminus. Previously characterized mutant cDNAs used in this study included those that encoded a Ser substitution for an interchain disulfide forming cysteine at Cys^{-1} [Cys−1Ser (*4*)], a deletion of the collagen-like domain (ΔGly^8−Pro^{80}) (*3*), and an Asn1Thr substitution that prevents glycosylation at the N-terminus (*9*). All mutant cDNAs were sequenced to verify the intended nucleotide changes and exclude spurious mutations. This analysis unexpectedly revealed that the leader sequence for rat SP-A contains an additional Gln (CAG) codon at the −28 position that was not reported in the original sequence (*7*).

In Vitro Transcription and Translation of SP-As with Translation Initiation Codon Mutations. In vitro transcription and translation were performed to determine if both Met^{-30} and Met^{-20} are utilized as translation initiation sites in the synthesis of rat SP-A, and the results are shown in Figure 2. The cDNAs for rat SP-A, Met−30Ala, and Met−20Ala were cloned into pGEM4Z and transcribed using the SP6 promoter. After translation in the presence of [^{35}S]Met, the mixture was separated on 16% SDS−PAGE gels and placed on film. Translation of the rat SP-A cDNA resulted in two bands at apparent molecular masses of approximately 28 and 29 kDa. However, translation of the Met−30Ala cDNA produced only a 28 kDa band, and translation of the Met−20Ala cDNA produced only a 29 kDa band. These data indicate that the translation can be initiated from either Met^{-30} or Met^{-20} and are consistent with use of both Met^{-30} and Met^{-20} in the synthesis of rat SP-A in vitro.

Synthesis and Characterization of Mutant Recombinant Proteins. For analysis of signal peptidase cleavage and posttranslational processing, proteins were synthesized in insect cells and purified by mannose−Sepharose affinity chromatography. Wild-type recombinant SP-A that is produced in this system retains the functional properties of the native protein despite reduced levels of hydroxyproline, which is indicated by the superscript "hyp" (*15*). The electrophoretic analysis of recombinant proteins with mutations in translation initiation loci are shown in Figure 3. Under reducing conditions, the $SP\text{-}A^{hyp}$ migrated as a diffuse

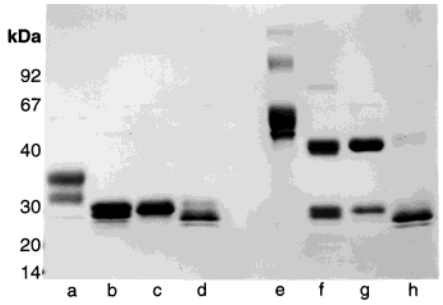

FIGURE 3: Electrophoretic analysis of SP-As containing mutations in translation initiation codons. Protein loads of 5−7 μg of purified SP-A per lane were size fractionated on 16% SDS−PAGE gels under reducing (a−d) and nonreducing (e−h) conditions and stained with Coomassie blue. Shown are rat SP-A (lanes a and e), $SP\text{-}A^{hyp}$ (lanes b and f), Met−20Ala (lanes c and g), and Met−30Ala (lanes d and h).

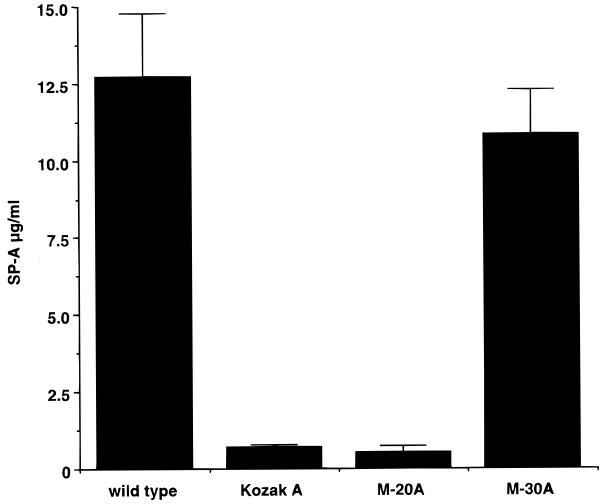

FIGURE 4: Effect of mutations of initiating methionines on the expression level of SP-A in insect cells. Insect cells were infected with recombinant baculoviruses encoding wild-type (wt) recombinant $SP\text{-}A^{hyp}$, SP-A containing the Met−30Ala mutation that forced translation initation at Met^{-20}, and Met−20Ala and Kozak A mutations that force translation initiation at Met^{-30}. The level of SP-A in the culture media was measured after a 72 h incubation.

band with an apparent molecular mass of 28−31 kDa (*9*), and the migrations of the Kozak A, Met−20Ala, and Met−30Ala proteins were similar to that of $SP\text{-}A^{hyp}$. Under nonreducing conditions, $SP\text{-}A^{hyp}$ formed a ladder of disulfide-linked oligomers. The mutant proteins also migrated as a range of multimeric species, but the Met−20Ala formed more extensive dimers than $SP\text{-}A^{hyp}$ or the Met−30Ala.

The site of translation initiation had a large effect on the level of secreted protein in the culture media (Figure 4). Infection of 10^7 *T. ni* cells with the recombinant virus encoding wild-type $SP\text{-}A^{hyp}$ at an MOI of 10 resulted in an SP-A concentration of 12.7 ± 2.1 μg/mL at 72 h postinfection. Under the same conditions, infection of the cells with recombinant viruses containing the Met−30Ala cDNA that forced translation initiation at Met^{-20} resulted in an SP-A concentration of 10.8 ± 1.5 μg/mL in the culture media (85% of $SP\text{-}A^{hyp}$ levels). The viruses which contained the Kozak A and Met−20Ala cDNAs that directed translation initiation at Met^{-30} produced SP-A levels of only 0.7 ± 0.1 and 0.5 ± 0.2 μg/mL, respectively, each less than 5% of $SP\text{-}A^{hyp}$ levels. These data indicate that synthesis and/or secretion of SP-A is much more efficient when initiated from Met^{-20} than

Table 2: N-Terminal Sequence Analysis of Recombinant SP-As Containing Mutations in the Translation Initiation Loci

	aa position[a]											
	−3	−2	−1	1	2	3	4	5	6	7	8	%[e]
SP-A^{hyp}												
sequence 1[b,c]				N_{10}	V_{43}	T	D	V	*C*	A		
sequence 2[d]	I_{12}	K	*C*	−	V	T	D	V	*C*			22
SP-A$^{hyp,M-20A}$												
sequence 1				$N_{1.7}$	V_8	T	D	V	*C*	A		
sequence 2	I_{15}	K	*C*	−	V	T	D					65
SP-A$^{hyp,M-30A}$												
sequence 1				N_5	V_{22}	T	D	V	*C*	A		
sequence 2	I_2	K	−	−	−	T						8
SP-A$^{hyp,Kozak\ A}$												
sequence 1				N	$V_{0.8}$	T	D					
sequence 2	$I_{3.9}$	K	*C*	N	V	T	D	V				83
SP-A$^{hyp,Kozak\ B}$												
sequence 1				$N_{2.4}$	$V_{8.2}$	T	D	V	*C*			
sequence 2	I_{26}	K	*C*	−	V	T	D	V	*C*			76
SP-A$^{\Delta G8-P80}$												
sequence 1				N	V_{25}	T	D	V	*C*	A	A	
sequence 2	I_6	K	−	N	V	T	D	V				18

[a] Numbered according to the position in the reported native SP-A sequence. [b] Subscripted numbers represent the molecular yield in picomoles. [c] Italicized cysteines were disulfide-linked on the basis of elution in the position of DiPTH-Cys. [d] (−) = blank cycle. [e] Calculated using the molecular yield (picomoles) of representative amino acids from each sequence [e.g., % sequence 1 isoform = pmol of sequence 1/(pmol of sequence 1 + pmol of sequence 2)].

Table 3: N-Terminal Sequence Analysis of Recombinant SP-As Containing Mutations in the Sites of Glycosylation and Interchain Disulfide Bond Formation

	aa position[a]											
	−3	−2	−1	1	2	3	4	5	6	7	8	%[e]
SP-A^{hyp,TM}												
sequence 1[b,c]				N_{18}	V_{17}	T	D	V	*C*	A		
sequence 2[d]	I_2	K	*C*	−	V	T	D					18
SP-A^{hyp,N1T}												
sequence 1				T	V_{65}	T	D	V	*C*	A		
sequence 2	I_{13}	K	−	−	−	T						17
SP-A^{hyp,T3A}												
sequence 1				N_{18}[f]	V_{18}	A	D	V	*C*	A		20
sequence 2	$I_{4.4}$	K	−	N	V	A	D	V	−	A		
SP-A$^{hyp,C-1A}$												
sequence 1				N_{22}	V_{60}	T	D	V	*C*	A	G	
sequence 2	I_8	K	A	−	−	T	D	V	−	A		12
SP-A$^{hyp,C-1S,TM}$												
sequence 1				N_{16}	V_{16}	T	D	V				
sequence 2	I_{98}	K	S	N	V	T						86

[a] −[e] See legend to Table 2. [f] Eluted as 7.2 pmol of Asn and 10.5 pmol of Asp, presumably due to spontaneous chemical deamidation.

from Met^{-30}, even when the Kozak consensus sequence surrounding the codon for Met^{-30} is optimized.

Determinants of SP-A Isoform Formation. To examine the effects of various leader sequence and N-terminal mutations on the expression of SP-A isoforms, we performed N-terminal sequence analysis on native, recombinant, and mutant SP-As. We have previously reported that SP-A isolated from the lungs of rats is composed of 18−22% IKC-SP-A isoform (Table 2). Recombinant SP-A expressed in insect cells is also composed of approximately 20% IKC-SP-A, and even large downstream mutations such as the deletion of the entire collagen-like domain (ΔG8−P80) do not affect the relative abundance of isoforms (*4*). However, restriction of translation initiation to the codon for Met^{-20} by Ala substitution of Met^{-30} decreased the relative expression of the IKC-SP-A isoform to only 8% of total protein. This result indicates that both isoforms can be derived by translation initiation at Met^{-20} and that translation initiation at Met^{-30} also contributes to SP-A expression. Restriction of translation initiation to Met^{-30} was accomplished both by optimizing the Kozak consensus sequence for translation initiation around that codon (to GCCACCaugG) and by Ala substitution of Met^{-20}. Since optimization of the Kozak consensus at the Met^{-30} site to include a guanine at the +4 position could not be accomplished without an amino acid substitution at Trp^{-29}, we produced both a mutant with an optimal Kozak sequence containing a Trp−29Gly substitution (Kozak A) and a second protein with an nearly optimized Kozak sequence except that no nucleotide substitution was made at the +4 position (Kozak B). Both the Kozak A and Kozak B proteins were composed of approximately 80% longer isoform, indicating that signal peptidase cleavage occurs primarily at the Gly^{-4}−Ile^{-3} site in proteins that begin with translation at Met^{-30}. We could not exclude the possibility that leaky scanning resulted in some translation initiation at the Met^{-20} codon in the Kozak mutants. To address this issue, we expressed a mutant protein with an ablated translation initiation site at Met^{-20}. The Met−20Ala SP-A was composed of 65% IKC-SP-A isoform, which is consistent with the finding that the Met−20Ala SP-A forms more extensive dimers than the Met−30Ala SP-A on nonreducing SDS−PAGE analysis (Figure 3). These data indicate that forcing translation initiation at Met^{-30} resulted in preferential expression of the longer isoform and that both isoforms can be generated following translation initiation at Met^{-30}. We conclude that alternative translation initiation contributes to the expression of two rat SP-A isoforms in an in vitro system. However, since both isoforms can be generated by translation initiation at either Met^{-30} or Met^{-20}, there must be other factors that contribute to alternative signal peptidase cleavage specificity for SP-A.

Two other mechanisms that were considered included posttranslational modifications of the nascent peptide that may direct signal peptidase to an otherwise cryptic cleavage site or a primary cleavage site context that is less than optimal for reasons that are not readily apparent. The primary signal peptidase cleavage site is flanked by Cys^{-1}, which participates in interchain disulfide bond formation, and Asn^{1}, which is variably glycosylated (Figure 1). Preventing glycosylation by Asn1Thr substitution resulted in a 17% abundance of the longer SP-A isoform, close to the 22% reported for SP-A^{hyp} (Table 3). Interpretation of this result is complicated by the fact that substitutions at the +1 position can also affect signal peptidase cleavage by altering secondary structure, by introducing steric hindrances, or by changing charge or hydrophobic character near the cleavage site. Disruption of N-terminal glycosylation by Ala3Thr substitution in the third position of the Asn-Val-Thr consensus sequence circumvented this problem and also had little or no effect on the relative abundance of isoforms. A third approach used that did not involve mutagenesis was to synthesize SP-A^{hyp} in the presence of tunicamycin (TM SP-A), an inhibitor of the first step in the formation of the core oligosaccharide. Glycosylation was prevented on the basis of the normalization of the relative abundance of Asn^{1} and Val^{2} on N-terminal sequencing (Table 3). The TM SP-A was composed of about

18% IKC-SP-A isoform, suggesting that lack of carbohydrate modification at Asn^1 has little if any effect on signal peptidase cleavage at the $Cys^{-1}-Asn^1$ bond (Table 3). To examine the effect of cotranslational disulfide formation on signal peptidase cleavage specificity, we substituted Cys^{-1} with nondisulfide-forming amino acids. We have previously reported that a Cys−1Ser substitution resulted in a marked increase in the expression of the longer isoform, from 22% for $SP\text{-}A^{hyp}$ to 88% for the Cys−1Ser mutant (*5*). Although Ser is frequently found in the −1 position relative to the cleavage site (*6*), it is possible that the Cys−1Ser substitution affected signal peptidase cleavage by a direct (steric, charge, hydrophobicity, secondary structure) effect on the susceptibility of the $Ser^{-1}-Asn^1$ bond to cleavage or by a secondary effect on glycosylation at the Asn^1 position, rather than through disruption of disulfide bond formation. The latter possibility was determined to be unlikely, since prevention of glycosylation of the nascent Cys−1Ser SP-A by addition of TM to the culture media did not affect the abundance of the longer isoform relative to that of the Cys−1Ser (86% vs 88% longer isoform; TM,Cys−1Ser SP-A vs Cys−1Ser SP-A, respectively) (Table 3). Finally, we assessed the effect of a Cys−1Ala substitution on the isoform composition of the protein. The Cys−1Ala SP-A was composed of 12% IKC-SP-A isoform, indicating that disruption of interchain disulfide formation does not enhance the expression of the longer isoform when an optimal amino acid for signal peptidase cleavage is present in the −1 position (Table 3). We concluded that a suboptimal context for signal peptidase cleavage at the primary site, but not cotranslational modifications of amino acids that flank the site, plays an important role in the amino acid sequence heterogeneity at the N-terminus of the mature protein.

DISCUSSION

The purpose of this study was to examine the mechanism of formation of a novel isoform of SP-A which contains a structurally important cysteine. The results indicate that alternative translation initiation at two in-frame AUGs and a suboptimal context or secondary structure for signal peptidase cleavage at the primary $Cys^{-1}-Asn^1$ cleavage site both play a role in the generation of the two SP-A isoforms. Neither cotranslational interchain disulfide bond formation at Cys^{-1} nor N-terminal glycosylation at Asn^1 appear to have an important influence on the site of signal peptidase cleavage.

The common structure of leader peptides includes a positively charged N-terminus, a hydrophobic core region, and a neutral but polar C-terminal region (*6*). Structural features that are thought to promote efficient cleavage include small and neutral amino acids at the −1 and −3 positions with respect to the cleavage site, with Ala, Gly, Ser, Cys, and Thr being most common (in that order), and a β-turn structure 5−6 residues upstream of the cleavage site. By these criteria, both the $Gly^{-4}-Ile^{-3}$ bond and the $Cys^{-1}-Asn^1$ bond within the leader peptide of SP-A qualify as potential signal peptidase cleavage sites (Figure 1). Recently, an empirical, neural network based method for prediction of signal peptidase cleavage sites was developed using leader sequence data from more than 1000 secreted proteins (www.cbs.dtu.dk and ref *16*). Application of this analysis to the rat SP-A sequence from the Met^{-30} of the leader peptide to Gly^8 of the mature protein also indicated that both the $Gly^{-4}-Ile^{-3}$ bond and the $Cys^{-1}-Asn^1$ bond in SP-A were potential cleavage sites, with the former being slightly more probable (Figure 5). The conclusion was the same when the Met^{-20} codon was presumed to be the unique translation initiation site (not shown). Thus, empirical analyses reveal that both of the observed signal peptidase cleavage sites of SP-A have favorable properties for cleavage, but the mechanisms for alternative cleavage and for preferential cleavage at the $Cys^{-1}-Asn^1$ bond have not been previously addressed.

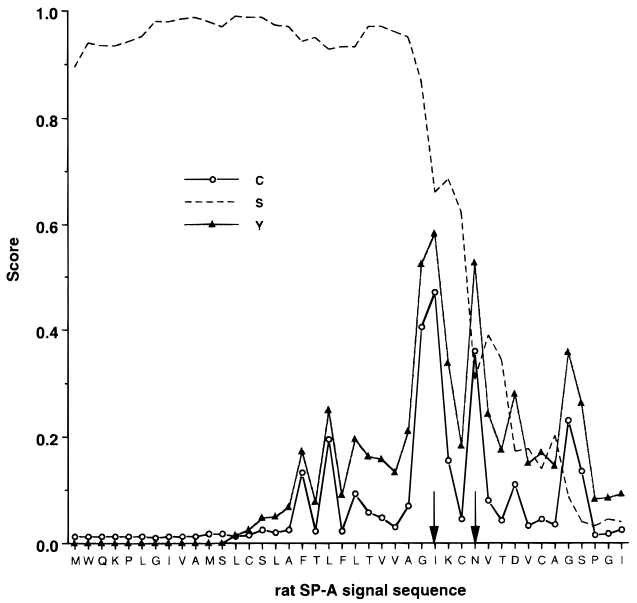

FIGURE 5: Predicted site of signal peptidase cleavage. An empirical, neural-based method developed by Neilsen et al. (*16*) was used to predict the site of signal peptidase cleavage in rat SP-A. The most likely cleavage site corresponds to the point where the S score transitions from high to low and the Y score and C score are maximal. This analysis results in a slightly higher predictive score for signal peptidase cleavage at $Gly^{-4}-Ile^{-3}$ than at $Cys^{-1}-Asn^1$, the opposite of what is observed in nature.

Alternative translation initiation has been increasingly recognized as a mechanism of protein isoform formation (*17*). Several studies have documented that, for eucaryotic mRNAs, the 40S ribosomal subunit enters at the 5′ end of the RNA and scans linearly in the 3′ direction until the first AUG codon is encountered. If the nucleotide sequence surrounding that codon provides an acceptable context for initiation of translation, approximating the optimal sequence for eucaryotes of GCCACCAUGG (*18*), translation initiation is restricted to that site (*19*). If, however, the first AUG that the ribosome encounters lies in a weak context, such as a sequence which does not contain the most critical elements of a purine at position −3 and a guanine at position +4 (relative to the first nucleotide of the AUG codon), translation may be alternatively initiated at the next in-frame downstream AUG (*20*). Two independent proteins may thus be produced from a single RNA by "leaky scanning", as has been reported for histone H4 (*21*) and a low molecular weight form of IL-1 receptor antagonist called icIL-1RaI (*22*). This mechanism for isoform formation is more common for genes encoding regulatory peptides such as cytokines, receptors, protein kinases, transcription factors, and growth factors than for genes encoding structural proteins (*21*).

The first AUG (Met^{-30}) following the transcription initiation site in rat SP-A lies in a particularly weak context. For

this reason, and because less than 10% of eucaryotic leader peptides are greater than or equal to 30 amino acids in length (www.cbs.dtu.dk and ref *16*), only the second AUG (Met^{-20}) was originally predicted to function as a translation initiation site (*7*, *23*). However, in vitro translation of a cDNA for rat SP-A in this study produced two similarly sized but distinct translation products, which comigrated with 28 and 29 kDa translation products that were individually generated with Met−30Ala and Met−20Ala cDNAs, respectively. These data indicate that both Met^{-30} and Met^{-20} translation initiation sites are utilized during in vitro synthesis of rat SP-A. Further evidence that Met^{-30} does indeed play a role in protein translation was provided by the finding that synthesis of the Met−30Ala protein in insect cells produced less of the IKC-SP-A isoform than is present in the wild-type protein. In addition to alternative translation initiation, the observation that both isoforms are produced by cDNAs which restrict translation initiation to Met^{-20} (or to Met^{-30}) indicates that an additional mechanism(s) must be operative to fully explain the N-terminal heterogeneity of SP-A.

SP-A is unusual in that there are two co- or posttranslationally modified structural amino acids flanking the primary signal peptidase cleavage site: an N-linked carbohydrate attachment site at the +1 position and an interchain disulfide bond forming Cys at the −1 position. Because the *N*-glycosyltransferase that catalyzes the first step in N-glycosylation is intimately associated with the signal peptidase (*24*), we considered the possibility that cotranslational carbohydrate modification of Asn^1 might redirect signal peptidase to the secondary cleavage site at Gly^{-4}−Ile^{-3}. We found that preventing core oligosaccharide attachment to Asn^1 with tunicamycin, by Asn1Thr mutation or by Thr3Ala substitution, did not significantly alter the relative abundance of SP-A isoforms that were expressed. We were unable to find any precedent in the literature for cotranslational N-glycosylation affecting signal peptidase cleavage specificity, but neither did we uncover examples of secreted proteins with a glycosylated N-terminal Asn other than mouse SP-A. Indeed, previous in vitro studies have suggested that N-terminal glycosylation does not occur, based on the failure of partially purified *N*-glycosyltransferase to modify N-terminal Asn's of short peptides containing Asn-X-Thr/Ser sequences (*25*−*27*). The closest that a consensus Asn was found to the N-terminus of secreted protein was a report by Gavel of an acceptor Asn at residue 2 of the mature chain, but identity of protein was not mentioned in the study (*28*). For membrane-bound proteins, the Asn-X-Thr acceptor site must be spaced at least 12−14 residues from the lumenal end of the membrane-spanning segment for effective glycosylation to occur during in vitro translation across pancreatic microsomes (*24*), indicating that the active site of the membrane-bound *N*-glycosyltransferase is positioned at a defined distance above the membrane surface. N-Terminal glycosylation of SP-A occurred whether translation was initiated at Met^{-30} or Met^{-20}, based on the comparative yield of Asn^1 and Val^2 in the N-terminal sequence analysis. We cannot determine from our data whether glycosylation precedes or follows signal peptidase cleavage, but three lines of evidence presented indicate that N-terminal glycosylation of SP-A does not affect signal peptidase cleavage specificity.

We had previously reported that Cys−1Ser substitution resulted in a marked increase in the abundance of the IKC-SP-A isoform, suggesting to us that prevention of disulfide bond formation at this locus directed signal peptidase cleavage to the Gly^{-4}−Ile^{-3} bond (*5*). However, substitution of Cys^{-1} by Ala in this study actually promoted cleavage at the Cys^{-1}−Asn^1 bond, suggesting that cleavage at this site was more a function of the amino acid that occupies the −1 position than a function of disulfide bond formation. This result was not expected since Cys and Ser are both preferrred amino acids at the −1 position of proteins. We concluded that a suboptimal context for signal peptide cleavage at the Cys^{-1}−Asn^1 site, but not interchain disulfide formation, contributes to the N-terminal heterogeneity of rat SP-A.

The biological relevance of heterogeneity at the N-terminus of SP-A is not clear. Both human and rat SP-A are expressed as two or more isoforms with Cys containing N-terminal extensions (*4*). For rat SP-A, the Cys^{-1} forms interchain disulfide bonds and is known to be required for full multimeric assembly, but the functional importance of the additional Cys has not been established (*5*). We speculate that regulated expression of the longer SP-A isoform, or proteolytic or oxidative disruption of the Cys^{-1} disulfide bond, might perturb oligomeric assembly and thereby regulate the activity of SP-A to opsonize microorganisms, to prevent the conversion of surface-active large surfactant aggregates to inactive small aggregates, or to protect surfactant monolayers from inhibitors in the alveolar lining fluid. Studies to determine if SP-A isoform distribution is modulated by lung injury at cotranslational or posttranslational levels are in progress.

In summary, we found that individually introduced mutations in two in-frame translation initiation codons lead to synthesis of two primary translation products, identical in size to those of the rat SP-A cDNA. These same mutations alter the expression and isoform composition of SP-A synthesized in insect cells, indicating that alternative translation initiation plays a role in SP-A isoform generation. The context of amino acids or the secondary structure of the leader peptide flanking the primary cleavage site also appears to play a role in signal peptidase cleavage specificity, since both SP-A isoforms are synthesized to some degree even when only one of the initiating methionines is intact. Cotranslational modifications of the nascent SP-A at the N-terminus, including interchain disulfide bond formation at Cys^{-1} or glycosylation at the Asn^1 oligosaccharide attachment site, did not appear to play a major role in determining whether cleavage occurred at the primary or the secondary cleavage position.

ACKNOWLEDGMENT

Protein sequencing was performed by Midwest Analytical, St. Louis, MO.

REFERENCES

1. McCormack, F. X. (1998) *Biochim. Biophys. Acta 1408*, 109−131.
2. Voss, T., Eistetter, H., Schafer, K. P., and Engel, J. (1988) *J. Mol. Biol. 201*, 219−227.
3. McCormack, F. X., Pattanajitvilai, S., Stewart, J. J., Possmayer, F., Inchley, K., and Voelker, D. R. (1997) *J. Biol. Chem. 272*, 27971−27979.
4. Elhalwagi, B. M., Damodarasamy, M., and McCormack, F. X. (1997) *Biochemistry 36*, 7018−7025.

5. Zhang, M., Damodarasamy, M., Elhalwagi, B. M., and McCormack, F. X. (1998) *Biochemistry 37*, 16481−16488.
6. Izard, J. W., and Kendall, D. A. (1994) *Mol. Microbiol. 13*, 765−773.
7. Sano, K., Fisher, J., Mason, R. J., Kuroki, Y., Schilling, J., Benson, B., and Voelker, D. (1987) *Biochem. Biophys. Res. Commun. 144*, 367−374.
8. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989) *Gene 77*, 61−68.
9. McCormack, F. X., Calvert, H. M., Watson, P., Smith, D. L., Mason, R. J., and Voelker, D. R. (1994) *J. Biol. Chem. 269*, 5833−5841.
10. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.
11. Fornstedt, N., and Porath, J. (1975) *FEBS Lett. 57*, 187−191.
12. McCormack, F. X., Fisher, J. H., Suwabe, A., Smith, D. L., Shannon, J. M., and Voelker, D. R. (1990) *Biochim. Biophys. Acta 1087*, 190−198.
13. Laemmli, U. K. (1970) *Nature 227*, 680−685.
14. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990−7997.
15. Edman, P. (1956) *Acta Chem. Scand. 10*, 761−768.
16. Nielsen, H., Englebrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng. 10*, 1−6.
17. Pain, V. M. (1996) *Eur. J. Biochem. 236*, 747−771.
18. Kozak, M. (1984) *Nature 308*, 241−246.
19. Kozak, M. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 7134.
20. Kozak, M. (1989) *Mol. Cell. Biol. 9*, 5073−5080.
21. Bab, I., Smith, E., Gavish, H., Attar-Namdar, M., Chorev, M., Chen, Y. C., Muhlrad, A., Birnbaum, M. J., Stein, G., and Frenkel, B. (1999) *J. Biol. Chem. 274*, 14474−14481.
22. Malyak, M., Guthridge, J. M., Hance, K. R., Dower, S. K., Freed, J. H., and Arend, W. P. (1998) *J. Immunol. 161*, 1997−2003.
23. Kozak, M. (1991) *J. Biol. Chem. 266*, 19867−19870.
24. Nilsson, I. M., and von Heijne, G. (1993) *J. Biol. Chem. 268*, 5798−801.
25. Struck, D. K., Lennarz, W. J., and Brew, K. (1978) *J. Biol. Chem. 253*, 5786−5794.
26. Bause, E., and Lehle, L. (1979) *Eur. J. Biochem. 101*, 531−540.
27. Bause, E., and Hettkamp, H. (1979) *FEBS Lett. 108*, 341−344.
28. Gavel, Y., and von Heijne, G. (1990) *Protein Eng. 3*, 433−442.

BI000448C