THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 264, No. 19, Issue of July 5, pp. 11387-11393, 1989
Printed in U.S.A.

Multiple cDNA Sequences and the Evolution of Bovine Stomach Lysozyme*

(Received for publication, February 24, 1989)

David M. Irwin and Allan C. Wilson

From the Department of Biochemistry, University of California, Berkeley, California 94720

To investigate the origin of stomach expression of lysozyme in ruminants, we surveyed clones from a cow stomach cDNA library with a lysozyme cDNA probe. Ten percent of the clones in this library were lysozyme-specific. Thirty of the lysozyme clones were sequenced, and seven types of lysozyme mRNA sequence were found. They encode the three previously identified stomach isozymes of lysozyme. The seven sequences are closely related to one another and represent the products of a minimum of 4 of the approximately 10 cow lysozyme genes detected by genomic blotting. The most abundant form of stomach lysozyme (form 2) is encoded by at least two genes, whereas forms 1 and 3 are possibly each encoded by only one gene. The number of genes encoding each isozyme appears to contribute the largest factor in the relative abundance of each isozyme. The multiple lysozyme genes expressed in the cow stomach are the result of gene duplications that occurred during ruminant evolution. The recruitment of lysozyme as a major enzyme in the stomach may thus have involved an early regulatory event and a later 4–7-fold increase in expression allowed by gene amplification. During this period, the amino acid sequences of these lysozymes have been evolving more slowly than those of nonruminant lysozymes.

Regulatory changes in the expression of genes have been proposed as being of primary importance in the adaptive evolution of mammals (reviewed by Wilson *et al.*, 1977). The antibacterial enzyme lysozyme *c* appears to have been recruited as a digestive enzyme in two types of placental mammals: the ruminants and the colobine monkeys (Dobson *et al.*, 1984; Stewart *et al.*, 1987; Stewart and Wilson, 1987). In both cases, foregut fermentation has evolved, allowing the animal to utilize otherwise indigestible plant material (Langer, 1974; Janis, 1976). Microbial fermentation occurs in the first chamber (rumen) of the multichambered stomachs of these mammals and converts much of the plant material to volatile fatty acids. Lysozyme *c*, which is expressed in the last chamber, *i.e.* the true stomach (abomasum), probably functions to degrade the cell walls of bacteria coming from the foregut (Dobson *et al.*, 1984; Prieur, 1986; Stewart and Wilson, 1987) and thus makes their contents available for digestion by the enzymes conventionally present in the mammalian stomach and small intestine.

Previous studies have shown that within the foregut-fermenting ruminants and colobine monkeys, lysozyme is expressed at high levels within the true stomach (Dobson *et al.*, 1984). Lysozyme has been recruited as a digestive enzyme in some hindgut-fermenting animals such as the mouse, and in these animals lysozyme is secreted into the intestine (Hammer *et al.*, 1987). Several enzymatic properties of ruminant and colobine stomach lysozymes (*e.g.* their low pH optima at physiological ionic strength and resistance to pepsin) differ from those of conventional mammal or bird lysozymes (Dobson *et al.*, 1984; Stewart *et al.*, 1987). These changes in enzymatic properties allow the lysozyme to function in the acidic and protease-rich stomach and appear to be correlated with specific changes in amino acid sequence, including several amino acid residues arising in the two groups of foregut fermenters as a result of convergent evolution (Stewart *et al.*, 1987).

The simplest explanation for the recruitment of lysozyme as a digestive enzyme would be a change in the regulation of a single lysozyme gene by acquisition of a stomach regulatory signal (*e.g.* a stomach-specific enhancer). In contrast, a survey of mammals for stomach lysozyme revealed that the cow stomach contains a minimum of three nonallelic forms of lysozyme (Pahud and Widmer, 1982; Dobson *et al.*, 1984). In addition, there appear to be several non-stomach forms in other cow tissues (Dobson *et al.*, 1984; Stewart, 1986; Prieur, 1986; White *et al.*, 1988). Multiple forms of lysozyme are also observed in the stomachs of other ruminants (Dobson *et al.*, 1984).[1] In contrast, most birds and mammals appear to have a single lysozyme *c* that may be expressed in many tissues (Jollès and Jollès, 1984; Helm-Bychowski and Wilson, 1986). Thus, in addition to a change in the site of lysozyme expression, an amplification in the number of lysozyme genes has occurred in ruminants. The amplification event may have been involved in the recruitment of lysozyme as a stomach enzyme, with one copy retaining the conventional lysozyme functions, similar to the origin of the gene for lactalbumin from a lysozyme gene (Prager and Wilson, 1988). Additionally, amplification may have been an adaptive response to increase the expression of lysozyme, similar to the acquisition of drug resistance by gene amplification in tissue culture (Schimke, 1984, 1988). Gene duplications may have had roles both early and late in the evolution of ruminant stomach lysozymes. Genomic blotting experiments have shown that most of the gene duplication events occurred after the recruitment of lysozyme as a stomach enzyme and after the origin of the digestive foregut (Irwin *et al.*, 1989).

Lysozymes from both the stomach (Dobson *et al.*, 1984) and milk (White *et al.*, 1988) of the cow have been isolated and

* This work was supported by National Institutes of Health Grant GM-21509 (to A. C. W.) and fellowships from the Medical Research Council of Canada and the Izaak Walton Killam Memorial Foundation (to D. M. I.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04831.

[1] E. M. Prager, personal communications.

characterized. The amino acid sequence of the most abundant form of stomach lysozyme (isozyme 2) (Jollès *et al.*, 1984, 1989) and of the amino-terminal 39 residues of a milk lysozyme (White *et al.*, 1988) have been determined. Molecular cloning resulted in the isolation of part of one cow stomach lysozyme gene (White *et al.*, 1983).

In an attempt to study the origin and evolution of the stomach lysozyme genes in ruminants, we have screened a bovine abomasum cDNA library, resulting in the isolation of seven distinct lysozyme cDNA sequences. These cDNAs encode the three major cow stomach lysozymes and demonstrate that at least four lysozyme genes in the bovine genome are expressed in the stomach to provide the high levels of lysozyme *c*.

EXPERIMENTAL PROCEDURES

Tissues and DNAs—Cow (*Bos taurus*), sheep (*Ovis aries*), and house mouse (*Mus castaneus*) genomic DNAs were provided by G. A. Cortopassi (Dept. of Biochemistry, University of California, Berkeley). Human (*Homo sapiens*) genomic DNA was provided by M. A. Stoneking (Dept. of Biochemistry, University of California, Berkeley). Rhesus (*Macaca mulatta*) and colobus (*Colobus polykomos*) monkey genomic DNAs, provided by S. L. Martin, were those described before (Martin *et al.*, 1983). Genomic DNA from fallow deer (*Dama dama*), chevrotain (*Tragulus napu*), and zebra (*Equus grevyi*) were prepared as described (Irwin *et al.*, 1989). All genomic DNAs were from a single individual of each species. Abomasum tissues for RNA isolation from cow, sheep, and axis deer (*Axis axis*), likewise from a single individual of each species, were from previous studies (Dobson *et al.*, 1984). An aliquot of an amplified bovine cDNA library in λgt10, prepared from mRNA from the abomasum of a single cow, and a bovine stomach lysozyme cDNA clone, λBL3, were generously provided by M. E. Williams and M. Harpold (SIBIA, San Diego, CA). A human lysozyme cDNA clone (Castañón *et al.*, 1988) was from M. J. Castañón (Ernst-Boehringer Institut für Arzneimittelforschung, Vienna, Austria). A cloned probe to cow lysozyme exon 2 was from a previous study (White *et al.*, 1983).

Enzymes and Chemicals—Restriction endonucleases, T4 DNA ligase, DNA polymerase I large fragment (Klenow), and modified T7 DNA polymerase (Sequenase) were purchased from Bethesda Research Laboratories, New England BioLabs, or the United States Biochemical Corp. [^{32}P]dCTP (3000 Ci/mmol) and [^{35}S]dATP (>1000 Ci/mmol) were supplied by Amersham Corp. Nitrocellulose (BA85) was purchased from Schleicher & Schuell. Agarose (SeaKem and SeaPlaque) was from FMC Corp.

Southern Blot Analysis—Genomic DNA (10 μg) was digested with various restriction endonucleases (20–30 units) according to the suppliers' specifications at 37 °C overnight, and the resulting DNA fragments were separated by electrophoresis in 0.8% agarose gels. After denaturation, the DNA was transferred to nitrocellulose (Southern, 1975). Blots were hybridized with ^{32}P-labeled probes and washed according to the procedure of Kan and Dozy (1978). Probes for hybridizations were labeled with [^{32}P]dCTP by the procedure of Feinberg and Vogelstein (1983).

Isolation of Stomach mRNA—Total RNA of cow, sheep, and axis deer abomasum was extracted from frozen tissue, which had been ground to a powder in liquid nitrogen, by the guanidinium/hot phenol method (Feramisco *et al.*, 1982). Messenger RNA was selected from total RNA by chromatography on an oligo(dT)-cellulose column (Edmonds *et al.*, 1971; Aviv and Leder, 1972).

Northern Blot Analysis of Ruminant Stomach mRNA—Messenger RNAs from the cow, sheep, and axis deer were denatured with formamide and formaldehyde, separated in 1.2% agarose-formaldehyde gels, and transferred to nitrocellulose (Maniatis *et al.*, 1982). Blots were hybridized with a 300-base pair *Eco*RI-*Pst*I fragment from the 5′ end of the bovine lysozyme cDNA (clone λBL3) and washed as described above for genomic blots.

Screening a Bovine Stomach cDNA Library—A total of 2000 phage clones from the bovine abomasum cDNA library was plated, and duplicate lifts on nitrocellulose were obtained. The replica filters were hybridized with the 300-base pair *Eco*RI-*Pst*I fragment described above. Small scale cultures of 30 positive clones were prepared (Maniatis *et al.*, 1982). Insert DNA from these clones was released by digestion with *Eco*RI and separated from vector DNA by electrophoresis through agarose with a low gelling temperature and purified by phenol extraction (Maniatis *et al.*, 1982). The cDNA inserts were subsequently ligated into M13 sequencing vectors (Messing, 1983) for the initial partial DNA sequence analysis.

DNA Sequence Analysis—λ Phages containing unique cDNA sequences (see Table I and Fig. 3) were purified with CsCl density gradients (Maniatis *et al.*, 1982). λ Clone DNA was digested with restriction endonucleases, ligated into M13 sequencing vectors mp18 and mp19, and subsequently transformed into an *Escherichia coli* JM101 host (Messing, 1983). Nitrocellulose plaque lifts of the M13 transformants were screened as described above using ^{32}P-labeled lysozyme cDNA. Single-stranded DNA was isolated from positive M13 phage plaques (Messing, 1983) and the DNA sequence determined by the chain termination method (Sanger *et al.*, 1977) as modified for ^{32}S (Biggin *et al.*, 1983) using either Klenow or Sequenase. DNA sequence data were compiled by using Intelligenetics computer programs.

RESULTS

Genomic Southern Blotting—As an initial step toward the characterization of lysozyme genes in ruminants, genomic DNAs from cow and sheep were digested with *Eco*RI, separated in an agarose gel, and blotted to nitrocellulose. Autoradiography after hybridization with ^{32}P-labeled bovine lysozyme cDNA revealed multiple fragments in both animals (Fig. 1*A*).

To show that the multiple fragments found in the cow genome were due to the presence of multiple lysozyme-like genes that were the result of amplification of lysozyme genes in the ruminant lineage, genomic DNAs from four ruminants (cow, sheep, fallow deer, and chevrotain) and three nonruminant mammals (zebra, rhesus monkey, and a mouse) were digested separately with two different enzymes (*Xba*I and *Bgl*II) and hybridized with a single exon (exon 2) from a cloned bovine lysozyme gene (White *et al.*, 1983). As shown in Fig. 1, *B* and *C*, both enzymes demonstrated that there were multiple (approximately 10) lysozyme-like genes in the ruminants and only one or a few genes in nonruminants. Similar results were observed using a human lysozyme cDNA clone (Fig. 1*D*). Hybridization conditions were such that conventional lysozyme *c* genes (Prager and Wilson, 1988) of any placental mammal could be detected with the bovine lysozyme cDNA probe, and therefore any gene duplication event within the artiodactyl or primate lineages should be detectable.

Northern Blot Analysis of Ruminant Stomach mRNA—The size of the major mRNA encoding stomach lysozyme in the cow, sheep, and axis deer is about 1100 nucleotides in length (Fig. 2). The cow appears to contain additional hybridizing messages that are longer.

Abundance of Stomach Lysozyme cDNA Types—In order to identify which of the multiple bovine lysozyme genes correspond to those expressed in the true stomach, an abomasum cDNA library was screened for lysozyme sequences with a cloned cow stomach lysozyme cDNA. Approximately 10% of the recombinant clones hybridized to this cow cDNA clone. The proportion of cDNAs which encodes lysozyme is similar to the proportion of the mRNA of the stomach mucosa which encodes lysozyme (White *et al.*, 1983); therefore, it appears that this cDNA library may truly reflect the lysozyme mRNA population.

Partial DNA sequence analysis of 30 purified cDNA clones led to the identification of seven distinct types of cDNA sequence (designated types 1a, 1b, 2a, 2b, 2c, 2d, and 3a in Figs. 3 and 4 and in Table I). Each type was identified at least twice but not more than eight times among the 30 sequenced clones (Table I). The two clones encoding sequence 2d were identical both in DNA sequence and in the location of the 5′ end and may therefore represent the same initial cDNA from the library construction.

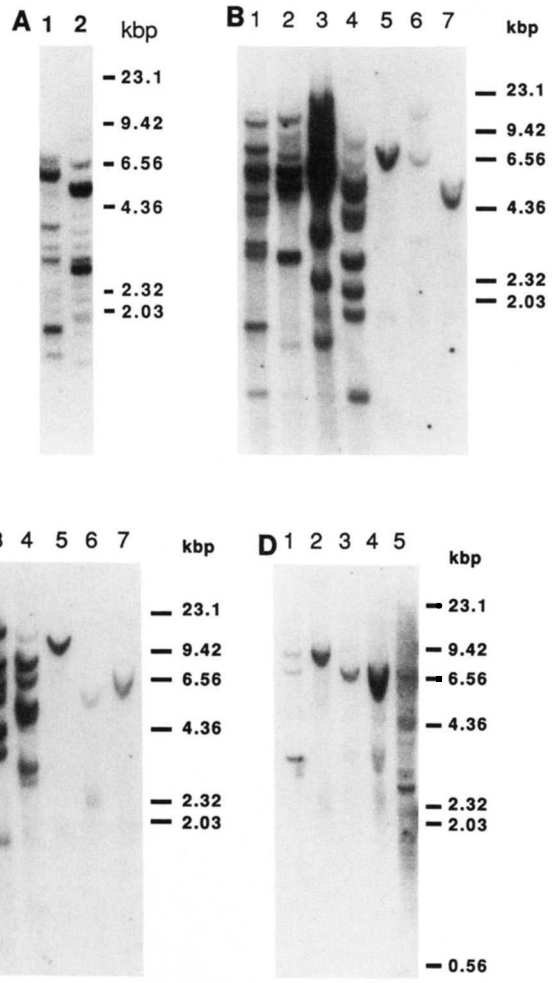

FIG. 1. **Lysozyme genes detected by genomic blotting.** Genomic DNA (10 μg) was digested with restriction endonucleases and electrophoresed in 0.8% agarose gels. After denaturation, the DNA was transferred to nitrocellulose and hybridized with ^{32}P-labeled lysozyme probes. The sizes in kilobase pairs (*kbp*) of the molecular weight markers (λ DNA digested with *Hin*dIII) appear to the *right* of each gel. (*A*) DNA from cow (*lane 1*) and sheep (*lane 2*) digested with *Eco*RI and probed with the bovine cDNA. (*B*) DNA from cow (*lane 1*), sheep (*lane 2*), fallow deer (*lane 3*), chevrotain (*lane 4*), zebra (*lane 5*), mouse (*lane 6*), and rhesus (*lane 7*) digested with *Bgl*II and probed with bovine exon 2. (*C*) same as *B* except digested with *Xba*I. (*D*) DNA from human (*lanes 1* and *2*), rhesus (*lane 3*), colobus (*lane 4*), and cow (*lane 5*) digested with *Eco*RI (*lanes 1* and *5*) or *Bgl*II (*lanes 2–4*) and probed with human cDNA.

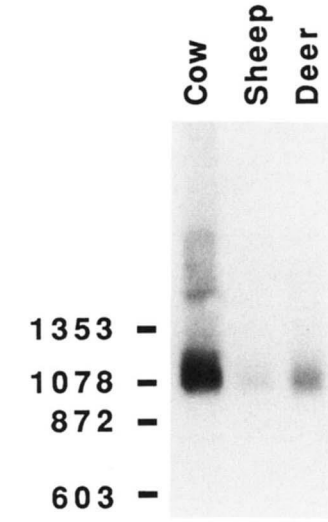

FIG. 2. **Northern blot analysis of ruminant stomach lysozyme mRNA.** Messenger RNA from several ruminants was denatured and fractionated in 1.2% agarose. The blot was hybridized to bovine stomach lysozyme cDNA and exposed to x-ray film for 4 h. The amounts of mRNAs used are 1 μg for cow and 2 μg for sheep and axis deer. The sizes in nucleotides of the molecular weight markers (ϕX174 DNA digested with *Hae*III) appear to the *left*.

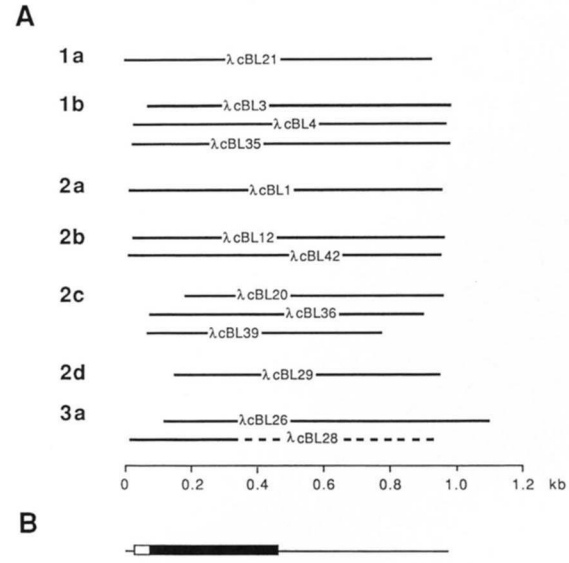

FIG. 3. **Sequencing strategy for cow stomach lysozyme sequences.** (*A*) the cDNA clones used to determine the sequence of the seven types of cow stomach lysozyme. The complete sequence of each of these 13 clones was obtained except for the part of λcBL28 which is shown as a *dashed line* (see Table I). Clones were identified as overlapping by the presence of uniquely shared nucleotide differences. The *scale* at the *bottom* indicates the length of the mRNA and cDNAs in nucleotides; *kb*, kilobase(s). (*B*) a representation of the lysozyme cDNA sequence with the signal peptide (*open box*), the coding region (*solid box*), and the untranslated regions (*line*).

Sequences of Cow Stomach Lysozymes—The DNA sequences of the seven distinct types of bovine stomach lysozyme cDNA (Fig. 4) were obtained from the cDNA clones shown in Fig. 3. The sequences 1b, 2b, 2c, and 3a were obtained from the overlapping clones (Fig. 3) which were identical to each other in the characterized regions and differed from other cDNA types by at least one nucleotide substitution. Some of the cDNA clones contained poly(A) tails that generally were located near nucleotide 970 (Fig. 4). Most of the cDNA clones did not contain poly(A) tails but still terminated near nucleotide 970. The absence of poly(A) tails is probably the result of failure to achieve full copying of the second strand during cDNA synthesis. Variability in the site of polyadenylation may account for some of the observed differences at the 3′ end (Fig. 4). One representative of sequence 3a (λcBL26, Table I) contained approximately 150 nucleotides of additional 3′ sequence. This additional sequence may be the result of the use of a rare polyadenylation site further downstream. Clone λcBL26 was the only one of 30 which appeared to extend beyond nucleotide 970 (Table I).

There is little variation in the size of lysozyme mRNA in bovine abomasum tissue, with the size of the mRNA being about 1100 nucleotides (Fig. 2, see above). Considering that poly(A) tails are usually 150–200 nucleotides in length (Perry, 1976), this implies that very little of the 5′ end of the mRNA sequence is absent from the longest cDNAs shown in Fig. 4. Polyadenylation signals AATAAT or AATAAC, AATACA, and ATTAAA (Birnstiel *et al.*, 1985) are observed in the

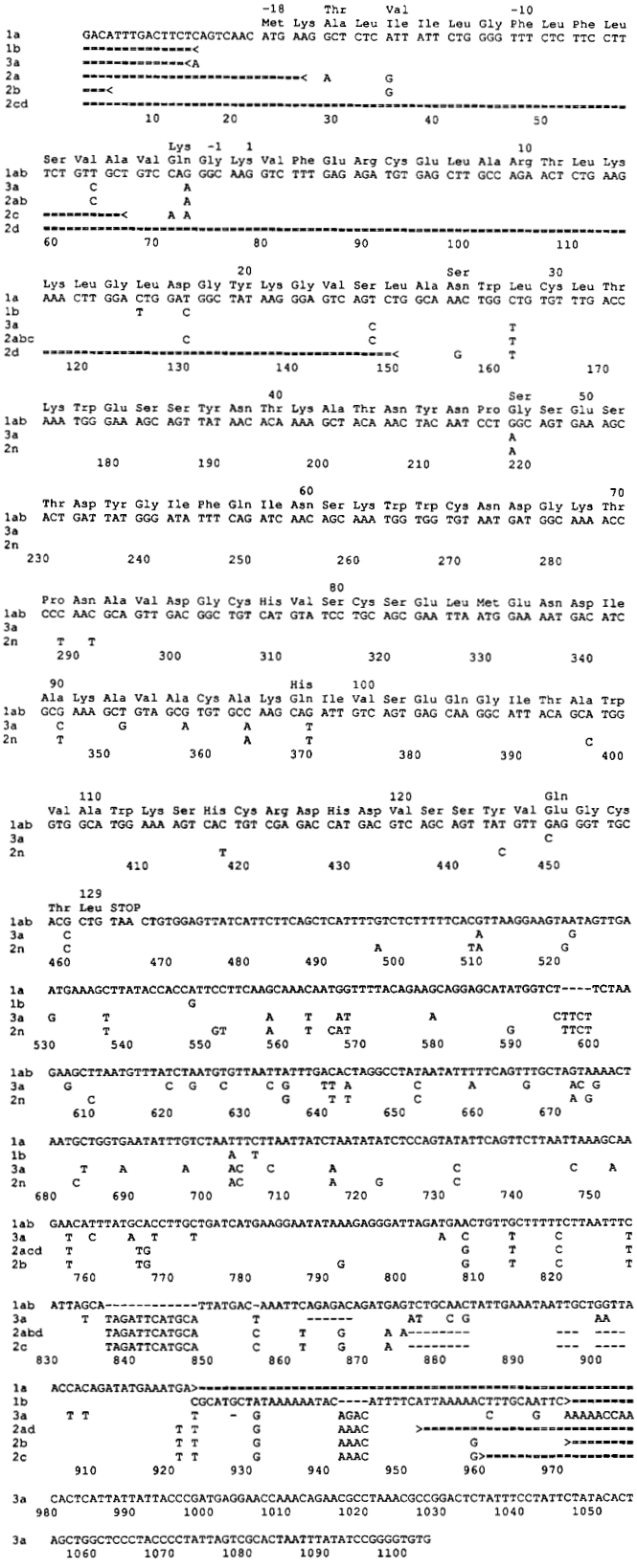

FIG. 4. **Sequences of seven cow stomach lysozyme cDNAs.** The nucleotide sequence of 1a is shown, with nucleotide 1 being the 5′ end of the cDNA. Differences in the nucleotide sequences found in the other cDNA clones appear below the first sequence, and gaps are indicated by *dashes*. (From positions 923 to 970, the sequence of 1b is shown in full and differences found in 2a–2d and 3a indicated; from 971 to 1103 the sequence of 3a is shown in full.) Sequences that were not isolated within the cDNAs are represented by ===< or >===. The amino acid sequence predicted from cDNA 1a is shown above the DNA sequence, with the amino acid differences predicted by the other cDNA sequences shown above the sequence for 1a. The amino acids are numbered from the amino terminus of the mature protein; the signal peptide is numbered backward. The amino acid sequence predicted by cDNAs 2a, 2b, and 2c is identical to the amino acid sequence of cow lysozyme 2 (Jollès *et al.*, 1984, 1989); a lysozyme with the amino acid sequence predicted from 2d would not have been observed as an electrophoretic variant of cow 2. The sequence predicted from cDNAs 1a and 1b is consistent with the electrophoretic properties of cow lysozyme 1, whereas the predicted product of cDNA 3a has a sequence consistent with the properties of cow lysozyme 3 (see text). The different cDNAs encoding cow 1 (a or b) and cow 2 (a, b, c, or d) are shown on separate lines only when they differ from each other, with *2n* representing all four cow 2-like cDNAs.

TABLE I

30 cow lysozyme cDNA clones that were sequenced

Type[a]	λcBL clone	Residues sequenced[b]
1a	21	1–922
	22	75–294
	25	166–305, 375–896
1b	3	68–970
	4	17–958
	8	334–757
	35	16–970
	38	32–316
1a or b	23	161–305
	31	142–298
2a	1	28–951
	10	113–298, 358–907
	30	32–305, 337–945
2b	12	15–970
	42	5–955
2c	20	182–959
	33	331–941
	36	76–893
	39	68–769
2d	27	151–292
	29	151–951
2a or b	2	32–298
2a, c, or d	18	338–795
2a, b, c, or d	32	337–639
	40	352–721
3a	11	237–961
	26	122–1103
	28	15–321
	37	361–832
1a, b, or 3a	41	241–298

[a] Numbers (1, 2, or 3) refer to the form of lysozyme encoded; letters (a, b, c, or d) distinguish different DNA sequences encoding the same mature lysozyme protein. Some clones could not be assigned to a unique type, as several types were identical within the sequenced region of the clone.

[b] See Figs. 3 and 4.

cDNAs at positions 890–895/8, 937–942/6, and 952–957 (see Fig. 4). The existence of these polyadenylation signals may account for some of the variability in the site of polyadenylation.

Predicted Amino Acid Sequences—In the cDNAs (Fig. 4), an initiator methionine codon is located at nucleotides 23–25, with a termination codon at nucleotides 464–466 (Fig. 4). Secreted cow stomach lysozyme is 129 residues long and is preceded by an 18-residue signal peptide responsible for secretion (Steiner *et al.*, 1980). The amino acid sequences of the mature lysozymes predicted from the seven cDNA sequences differ from one another by 1 to 3 amino acid replacements (Fig. 4) and are consistent with the information reported by Dobson *et al.* (1984) and Jollès *et al.* (1984) for the three previously identified stomach isozymes. The amino acid sequences predicted from cDNAs 2a, 2b, and 2c agree completely with that reported for isozyme 2 (Jollès *et al.*, 1984, 1989).

TABLE II

Correspondence between cDNA abundance and amount of lysozyme in cow stomach

Form of lysozyme	No. of sequences/ lysozyme	cDNA abundance[a]		Relative amounts of lysozyme protein[b]
		No.	%	
1	2	10	34	34 ± 4
2	4	15	52	56 ± 5
3	1	4	14	10 ± 4

[a] Only 29 of the 30 correspond to a single form. The remaining clone corresponded in sequence to either form 1 or form 3 (Table I).

[b] The relative amounts of each of the three lysozymes from 18 individual cows, with standard deviations, as detected by gel electrophoresis and protein staining (Dobson *et al.*, 1984).

The sequence predicted by 2d differs from the cow 2 sequence at position 27 and may be a variant of isozyme 2. (This difference was identified in two identical cDNA clones, and thus the possibility of an artifact cannot be excluded. Alternatively, this enzyme sequence may represent a true isozyme found at low levels (possibly 13%, see Table I) compared with isozyme 2 or a rare allele (functional or nonfunctional) found in the individual from which the cDNA library was constructed. Serine at position 27 has not been reported in any other lysozyme *c* (Jollès *et al.*, 1989 and references therein).)

The sequences predicted from the remaining cDNAs are consistent with the relative electrophoretic mobilities of isozymes 1, 2, and 3 at pH values of 4.3 and 8.9 (Dobson, 1981; Dobson *et al.*, 1984)[2] and also with amino acid compositions reported for the three forms (Jollès *et al.*, 1984). In addition to the amino acid differences observed within the mature enzyme, there are three positions within the signal peptide where the sequences differ (Fig. 4). Each of these replacements appears to be a conservative change and probably does not interfere with signal peptide function (von Heijne, 1985).

Half of the cDNAs isolated in this library screening encode cow stomach lysozyme 2-like sequences (Tables I and II). Of the remaining half, 10 encode isozyme 1, four encode isozyme 3, and one may encode either 1 or 3 (see Table I). The proportions of the cDNAs encoding each of the isozymes of lysozyme are very similar to the average fractions of each isozyme found by Dobson *et al.* (1984) in stomach mucosa (Table II). The proportion of the total cDNA which is specific for lysozyme is similar also to the proportion of mRNA encoding lysozyme in the stomach mucosa. Thus, the lysozyme cDNAs appear to reflect the mRNA population. This result implies that the level of each isozyme found in the stomach is a reflection of the amount of mRNA encoding each and that these levels are not greatly affected by different translation or secretion efficiencies. The differences observed in the sequences of the signal peptides of these isozymes (Fig. 4) probably have little effect on secretion, as expected from previous studies on signal peptides (Steiner *et al.*, 1980).

DISCUSSION

Multiple Genes for Lysozyme—Three nonallelic forms of stomach lysozyme *c* have been described previously in the cow, and additional forms occur in non-stomach tissues (Pahud and Widmer, 1982; Dobson *et al.*, 1984; Stewart, 1986; White *et al.*, 1988). In contrast, most birds and mammals have had fewer lysozymes, usually only one, identified in all tissues. Genomic Southern blot analysis of mammalian DNA hybridized with stomach lysozyme cDNA indicates the existence of multiple lysozyme-like sequences within the ruminant genomes (Fig. 1). Hybridization of these genomic blots with a single exon (exon 2) from a cow lysozyme gene indicated that there were about 10 genes in the ruminants, whereas nonruminants had only one or a few genes (Fig. 1, *B* and *C*). As similar results were observed with a human lysozyme cDNA (Fig. 1*D*), the implication is that the amplification in the number of lysozyme genes occurred only on the ruminant lineage.

The existence of 10 lysozyme-like genes within the cow genome is consistent with the identification of several forms of lysozyme in different cow tissues. The existence of the three forms of lysozyme found in the stomach is explained by the duplication of lysozyme genes within the ruminant lineage. The duplications of stomach lysozyme genes may have been an adaptive response to a requirement for greater amounts of lysozyme in the stomach. Additional lysozyme genes would then be available for expression in non-stomach tissues.

Size of Stomach Lysozyme mRNA—Northern blot analysis of mRNA from several ruminant species showed that the mRNA is approximately 1100 nucleotides long (Fig. 2). It had been shown previously that mRNA of a similar length is a major component in the poly(A) fraction extracted from the cow stomach and can be translated into lysozyme (White *et al.*, 1983). The length of mRNA for lysozyme is more than twice that needed to encode a protein of 147 residues and thus suffices to encode the entire mature protein plus signal peptide. Additional longer mRNAs were observed in the cow (Fig. 2), suggesting either that multiple sites of polyadenylation exist in a single gene or that different genes encode lysozyme mRNAs of slightly differing lengths (or both).

Number of Stomach Lysozyme Genes—Based on the number of different cDNA sequences identified and the fact that all of these cDNAs were isolated from one individual, there must be a minimum of four genes expressed in stomach tissue, with the other three sequences possibly representing alleles of three of these genes. There are four sequences encoding isozyme 2, which at a minimum represents two different genes, each with two alleles; and the two cDNAs for isozyme 1, in turn, may represent two alleles from a single gene. If only four genes for stomach lysozyme exist, with two of them encoding isozyme 2, this indicates that the four genes are all expressed at roughly equal levels (see Tables I and II). Alternatively, each of the cDNAs may represent a single gene, but again, all would be expressed at similar levels. The majority if not all of the major stomach lysozyme genes have thus been identified, though it is possible that other lysozyme genes are expressed at low levels in stomach tissue and thus would be found at low levels within the cDNA library and unlikely to be among the identified clones.

Since the proportion of cDNAs encoding each isozyme of lysozyme is similar to the proportion of each protein in the stomach, it appears that each of the cDNAs for lysozyme is translated at equal efficiency, as shown in Table II. The most abundant form of lysozyme (cow 2) found in stomach mucosa from the viewpoints of both protein and mRNA (Table II) is encoded by a minimum of two genes, whereas cow 1 and cow 3 may each be expressed by a single gene. The stomach lysozyme genes are all expressed at high levels to an extent

[2] Isozyme 3a differs from isozyme 2 by a single amino acid residue with glutamine replacing glutamate at position 125 in the predicted amino acid sequence (Fig. 4). This replacement results in the loss of a negative charge, consistent with a relative mobility of 103 at pH 4.3 and 64 at pH 8.9 (with a mobility of 100 assigned to isozyme 2 at each pH). The predicted amino acid sequence of 1a and 1b differs from that of the isozyme 2 sequence at position 48 (glycine in place of serine) and at position 98 (glutamine in place of histidine) (Fig. 4). The latter change results in the loss of a positive charge at low pH values, in agreement with a relative mobility of 93 at pH 4.3 and 100 at pH 8.9 (Dobson, 1981; Dobson et al., 1984; and Footnote 1).

such that 10% of stomach mucosal mRNA encodes lysozyme. It therefore appears that the greatest factor contributing to the differences in amount of each of the three forms of stomach lysozyme is the number of genes encoding each of them. Changes in the number of lysozyme genes expressed in the stomach may be the easiest way to elevate the level of expression of lysozyme in ruminants such as cows.

Relationships between the cDNAs and Isozymes—As shown in Fig. 4 and Table III, the different cDNAs that encode the same isozyme are very similar in nucleotide sequence. The stomach isozymes themselves differ from each other by only 1–3 amino acid replacements within the mature secreted enzyme, with additional changes occurring within the signal peptide (see above and Fig. 4).

The relationships among the cow stomach isozymes were analyzed as shown in Fig. 5, based on the amino acid sequences of the mature proteins. The axis deer and pig lysozyme sequences (Jollès *et al.*, 1989) were included in the analysis, with the pig sequence as the outgroup, to permit determination of the order of gene divergences and the lineages on which particular amino acid replacements occurred. The evolutionary tree in Fig. 5 indicates that the cow stomach isozyme genes duplicated within the bovine lineage after the divergence of the cow and deer, which occurred approximately 25 million years ago (Savage and Russell, 1983). (Since deer appear to have multiple lysozyme genes (Fig. 1), it may be expected that they would also have multiple stomach lysozymes, but initial studies revealed only a single lysozyme (Dobson *et al.*, 1984; Jollès *et al.*, 1989).)

Since all the cow stomach lysozyme sequences are more similar to one another than any one is to the axis deer sequence, the amplification of stomach lysozyme genes might have occurred after the cow-deer divergence; alternatively, during this time period concerted evolution may have homogenized the stomach lysozyme sequences within each species.

TABLE III

Matrix of sequence differences in the 3′-noncoding region of cow stomach lysozymes

Percent difference for all possible pairs of cDNAs appears above the diagonal; below the diagonal is the number of nucleotide differences followed by the number of base pairs compared. Gaps were not counted so that the table summarizes point mutational differences within the region 467–970 in Fig. 4.

Sequences compared	1a	1b	2a	2b	2c	2d	3a
1a		0.68	8.96	9.20	8.73	8.96	12.47
1b	3/439		9.13	9.19	8.97	9.13	12.39
2a	38/424	41/449		0.21	0.21	0.00	10.58
2b	39/424	43/468	1/470		0.42	0.21	11.00
2c	37/424	41/457	1/470	2/478		0.21	10.40
2d	38/424	41/449	0/470	1/470	1/470		10.58
3a	54/433	59/476	49/463	53/482	49/471	49/463	

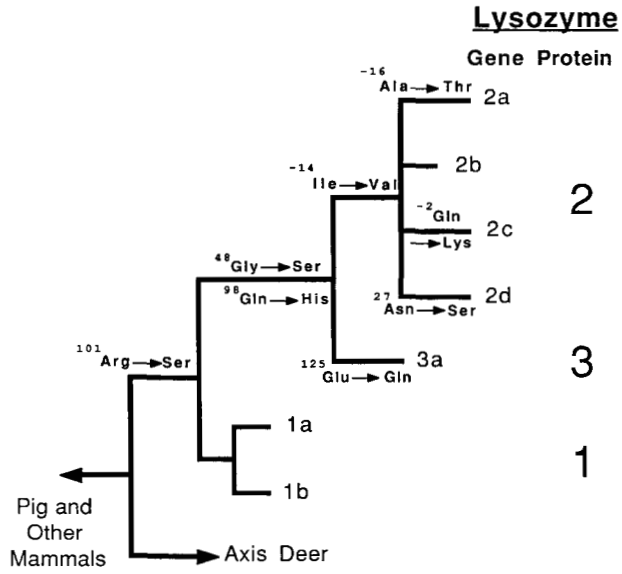

FIG. 5. **Tree analysis of amino acid substitutions among ruminant lysozymes.** The relationships among the amino acid sequences of cow stomach lysozymes are depicted. The direction of the amino acid substitutions and the lineages on which they occurred are indicated, with the numbering of positions in the amino acid sequences as in Fig. 4. Branching order was determined by the parsimony method, which means that it accounts for the observed diversity with the fewest mutations. The amino acid sequences of the mature lysozymes from axis deer and pig stomachs (Jollès *et al.*, 1989) were used, with pig as an outgroup to root the tree. (Because the complete sequences of the signal peptides of 2a, 2c, and 2d were not determined (Fig. 4), there is some uncertainty regarding the lineage to which the changes at positions −2, −14, and −16 should be assigned, *i.e.* the missing information might lead to some resolution of the branching order among 2a–2d.)

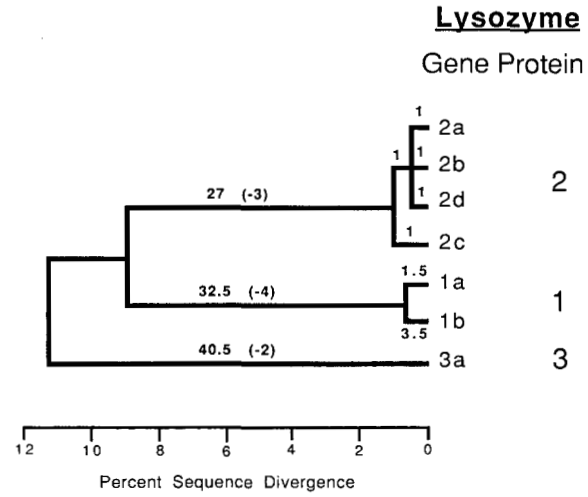

FIG. 6. **Tree analysis of nucleotide substitutions among stomach lysozyme genes.** The tree was constructed from the sequences in Fig. 4, and nucleotide substitutions were assigned to specific lineages based on parsimony. The number of nucleotide substitutions in the complete sequence is shown on each lineage. The number of places where deletions (−) of nucleotides occur appears in *parentheses*; contiguous deletions of more than 1 base pair are counted as single events. The *scale* below indicates the amount of difference between the 3′-noncoding regions (nucleotides 467–970) of the different lysozymes (Table III). At 13 of the 106 positions with point mutational differences, two or three equally parsimonious assignments of nucleotide substitutions exist; these alternative solutions were averaged. At the few positions at which some of the sequences were unavailable, it was assumed that 1a and 1b were the same and 2a–2d were identical to one another or (positions 29 and 71) that only one among 2a–2d was variable.

TABLE IV

Silent and replacement substitutions in genes for cow stomach prelysozymes

The numbers of silent and replacement nucleotide substitutions within the amino acid coding region (nucleotides 23–466 in Fig. 4) are shown. The signal peptide is encoded by nucleotides 23–76 and the mature secreted enzyme by nucleotides 77–466.

Type of substitution	No. of substitutions observed		
	Signal peptide	Mature enzyme	Total
Silent	2	16	18
Replacement	3	4	7

Concerted evolution may be the more likely explanation.[3]

In addition to the assignment of amino acid replacements to specific lineages as shown in Fig. 5, it is also possible to assign nucleotide changes in the cDNAs to specific lineages (Fig. 6). The relationships shown in Fig. 6 represent an unrooted network and do not differ significantly from the relationships shown in Fig. 5. The different cDNAs that encode the same isozyme differ by two to five nucleotide substitutions (<0.6% difference, Figs. 4 and 6 and Table III), numbers that indicate that these could be allelic differences (except that there must be at least two genes for lysozyme 2). The sequences for the different isozymes, in contrast, differ from one another by about 9–12% in their 3′-noncoding sequences (Fig. 4 and Table III).

Anomalous Ratio of Silent to Replacement Changes—Cow stomach lysozymes have an unusual ratio of silent to replacement substitutions. Within the region encoding amino acids, there are 25 nucleotide substitutions among the seven cDNA sequences (Fig. 4 and Table IV). Of the 25 changes, 7 result in amino acid replacements, and 18 are silent. Differences in the ratio of silent to replacement changes are observed between the nucleotide sequences of the signal peptide and the mature secreted enzyme. Within the signal peptide, there are three replacements and two silent changes (Table IV), a ratio of silent to replacement substitutions not different from that of the signal peptides of bird lysozymes (Weisman *et al.*, 1986) and from the ratio expected for typical mammalian (non-stomach) lysozymes. The mature secreted enzyme, in contrast, has 4 replacements and 16 silent changes (Table IV). There are statistically fewer replacements than expected if stomach lysozyme were evolving like its signal peptide or like conventional bird or mammalian lysozymes in general.

Comparison of amino acid sequences of lysozymes from cow and axis deer stomachs with other mammalian lysozymes has indicated that ruminant stomach lysozymes initially evolved faster than the lysozymes of other creatures but afterward evolved at a subnormal rate (Jollès *et al.*, 1989). Cow stomach lysozyme amino acid sequences appear to be evolving at about one-fourth the rate of typical lysozymes. The reduced amino acid sequence evolution is coupled with a 3–4-fold increase in the ratio of silent to replacement nucleotide substitutions. This observation supports the proposal that ruminant lysozymes are evolving at a reduced rate as a consequence of an increase in purifying selection due to functional constraints on the well adapted stomach enzyme, whereas synonymous substitutions accumulate at the standard rate, presumably because they are neutral or nearly so.

Acknowledgments—We thank all those who provided materials as well as G. A. Cortopassi, D. E. Dobson, U. Gyllensten, T. D. Kocher, C.-B. Stewart, L. Vigilant, R. T. White, and M. E. Williams for helpful advice and discussion. We also thank E. M. Prager for discussions on multiple lysozymes in various ruminants as well as for computational and editorial assistance.

[3] D. M. Irwin, unpublished observations.

REFERENCES

Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69,** 1408–1412
Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 3963–3965
Birnstiel, M. L., Busslinger, M., and Strub, K. (1985) *Cell* **41,** 349–359
Castañón, M. J., Spevak, W., Adolf, G. R., Chlebowicz-Śledziewska, E., and Śledziewski, A. (1988) *Gene (Amst.)* **66,** 223–234
Dobson, D. E. (1981) Ph.D. dissertation, University of California, Berkeley
Dobson, D. E., Prager, E. M., and Wilson, A. C. (1984) *J. Biol. Chem.* **259,** 11607–11616
Edmonds, M., Vaughn, M. H., Jr., and Nakazato, H. (1971) *Proc. Natl. Acad. Sci. U. S. A.* **68,** 1336–1340
Feinberg, A. P., and Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13
Feramisco, J. R., Smart, J. E., Burridge, K., Helfman, D. B., and Thomas, G. P. (1982) *J. Biol. Chem.* **257,** 11024–11031
Hammer, M. F., Schilling, J. W., Prager, E. M., and Wilson, A. C. (1987) *J. Mol. Evol.* **24,** 272–279
Helm-Bychowski, K. M., and Wilson, A. C. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 688–692
Irwin, D. M., Sidow, A., White, R. T., and Wilson, A. C. (1989) in *The Immune Response to Structurally Defined Proteins: The Lysozyme Model* (Smith-Gill, S., and Sercarz, E., eds) pp. 73–85, Adenine Press, Guilderland, NY
Janis, C. (1976) *Evolution* **30,** 757–774
Jollès, J., Jollès, P., Bowman, B. H., Prager, E. M., Stewart, C.-B., and Wilson, A. C. (1989) *J. Mol. Evol.* **28,** 528–535
Jollès, P., and Jollès, J. (1984) *Mol. Cell. Biochem.* **63,** 165–189
Jollès, P., Schoentgen, F., Jollès, J., Dobson, D. E., Prager, E. M., and Wilson, A. C. (1984) *J. Biol. Chem.* **259,** 11617–11625
Kan, Y. W., and Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 5631–5635
Langer, P. (1974) *Mammalia* **38,** 295–314
Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
Martin, S. L., Vincent, K. A., and Wilson, A. C. (1983) *J. Mol. Biol.* **164,** 513–528
Messing, J. (1983) *Methods Enzymol.* **101,** 20–78
Pahud, J.-J., and Widmer, F. (1982) *Biochem. J.* **201,** 661–664
Perry, R. P. (1976) *Annu. Rev. Biochem.* **45,** 605–629
Prager, E. M., and Wilson, A. C. (1988) *J. Mol. Evol.* **27,** 326–335
Prieur, D. J. (1986) *Comp. Biochem. Physiol.* **85B,** 349–353
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
Savage, D. E., and Russell, D. E. (1983) *Mammalian Paleofaunas of the World*, Addison-Wesley Publishing Company, Reading, MA
Schimke, R. T. (1984) *Cell* **37,** 705–713
Schimke, R. T. (1988) *J. Biol. Chem.* **263,** 5989–5992
Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517
Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J., and Tager, H. S. (1980) *Ann. N. Y. Acad. Sci.* **343,** 1–16
Stewart, C.-B. (1986) Ph.D. dissertation, University of California, Berkeley
Stewart, C.-B., and Wilson, A. C. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 891–899
Stewart, C.-B., Schilling, J. W., and Wilson, A. C. (1987) *Nature* **330,** 401–404
von Heijne, G. (1985) *J. Mol. Biol.* **184,** 90–105
Weisman, L. S., Krummel, B. M., and Wilson, A. C. (1986) *J. Biol. Chem.* **261,** 2309–2313
White, F. H., Jr., McKenzie, H. A., Shaw, D. C., and Pearce, R. J. (1988) *Biochem. Int.* **16,** 521–528
White, R. T., Dobson, D. E., Wilson, A. C., Jollès, J., Schoentgen, F., and Jollès, P. (1983) *Fed. Proc.* **42,** 1971
Wilson, A. C., Carlson, S. S., and White, T. J. (1977) *Annu. Rev. Biochem.* **46,** 573–639