INFECTION AND IMMUNITY, Jan. 1991, p. 7–17
0019-9567/91/010007-11$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 59, No. 1

Analysis of the Immunoglobulin A Protease Gene of *Streptococcus sanguis*

JOANNE V. GILBERT, ANDREW G. PLAUT,* AND ANDREW WRIGHT

Gastroenterology Division, Department of Medicine, Tufts-New England Medical Center Hospital, and Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111

Received 23 July 1990/Accepted 10 October 1990

The amino acid sequence T-P-P-T-P-S-P-S is tandemly duplicated in the heavy chain of human immunoglobulin A1 (IgA1), the major antibody in secretions. The bacterial pathogen *Streptococcus sanguis*, a precursor to dental caries and a cause of bacterial endocarditis, yields IgA protease that cleaves only the Pro-Thr peptide bond in the left duplication, while the type 2 IgA proteases of the genital pathogen *Neisseria gonorrhoeae* and the respiratory pathogen *Haemophilus influenzae* cleave only the P-T bond in the right half. We have sequenced the entire *S. sanguis iga* gene cloned into *Escherichia coli*. A segment consisting of 20 amino acids tandemly repeated 10 times, of unknown function, occurs near the amino-terminal end of the enzyme encoded in *E. coli*. Identification of a predicted zinc-binding region in the *S. sanguis* enzyme and the demonstration that mutations in this region result in production of a catalytically inactive protein support the idea that the enzyme is a metalloprotease. The *N. gonorrhoeae* and *H. influenzae* enzymes were earlier shown to be serine-type proteases, while the *Bacteroides melaninogenicus* IgA protease was shown to be a cysteine-type enzyme. The streptococcal IgA protease amino acid sequence has no significant homology with either of the two previously determined IgA protease sequences, that of type 2 *N. gonorrhoeae* and type 1 *H. influenzae*. The differences in both structure and mechanism among these functionally analogous enzymes underscore their role in the infectious process and offer some prospect of therapeutic intervention.

Immunoglobulin A (IgA) proteases are extracellular bacterial enzymes produced by numerous mucosal pathogens which cleave single peptide bonds in the hinge region of the human IgA1 immunoglobulin heavy chain (12, 22, 23, 35, 36, 42). The exact cleavage site is dependent on the bacterial strain but always follows one of the proline residues in the hinge (Fig. 1). Despite sharing substrate specificity for human IgA, the enzymes from various bacteria have different catalytic mechanisms. For example, the *Neisseria gonorrhoeae* and *Haemophilus influenzae* IgA proteases are apparently serine-type proteases because they have a consensus sequence for this class (1) and are inhibited by peptide boronic acid analogs of the IgA1 substrate (1). These two proteases are approximately 50% homologous (44), with extensive stretches of identity. In contrast, the IgA protease of *Bacteroides melaninogenicus* has been identified as a thiol-activated protease (35), and the IgA protease of *Streptococcus sanguis*, reported on in detail here, had been identified as a metalloprotease because it is inhibited by EDTA (28, 41). The dissimilarity in structure and catalytic mechanisms for enzymes that share substrate specificity could be a striking example of evolutionary convergence among pathogenic bacteria.

Cloning and expression of *iga* genes provides the opportunity for detailed study of these relationships. We have recently identified the *iga* gene in *S. sanguis* ATCC 10556 and cloned it into *Escherichia coli* (15), in which it encodes an active protease of relative molecular weight 190,000. The enzyme produced in *E. coli* has substrate specificity identical to that of the wild-type enzyme and is secreted to the periplasmic space. DNA probes from within the *S. sanguis iga* gene showed no demonstrable homology with *iga* genes of *Haemophilus* spp., *Neisseria* spp., or *Streptococcus pneumoniae* under conditions of high stringency.

To further define the biochemistry and genetics of these functionally equivalent enzymes, we have sequenced the entire *S. sanguis iga* gene. This sequence was compared with those of the other cloned proteases (43, 44) to search for similarities that explain their pronounced substrate specificity and to more securely classify the *S. sanguis* protease by identifying segments likely to be in the enzyme active site. Information on the *iga* gene has facilitated the construction of a protease-negative *S. sanguis* mutant (10556N3) for study of the role of the enzyme in the pathogenesis of dental caries and other bacterial infections. Also, insight into the catalytic mechanism is necessary for the rational design of inhibitors.

MATERIALS AND METHODS

Bacterial strains, plasmids, bacteriophage, and media. *E. coli* K-12 strains MM294 (*hrs hrm*$^+$ *thi endA supE*) and MC1000 [F$^-$ Δ(*ara-leu*)*7697 araD139* Δ*lac*(X74) *galU galK rpsL*] were used as hosts for plasmids, and JM103 was used as the host for M13 phage. *E. coli* CJ236 [*dut ung thi relA*(pCJ105 Cm^r)] and *E. coli* MV1190 [Δ(*lac-proAB*) *thi supE* Δ(*srl-recA*)*306*::Tn*10*(Tet^r) F′ *traD36 proAB lacI*^{q}ZΔM15)] were used for site-specific mutagenesis. Phage vectors of the M13mp series (59) and the recombinant plasmids pJG1 and pJG2 have been described in detail elsewhere (15). The *E. coli*-streptococcal shuttle vector pVA838 was kindly provided by Don Clewell and was used to construct pVA891 (31). *S. sanguis* of Lancefield group H was from the American Type Culture Collection (Rockville, Md.), catalog no. 10556, and is referred to below as 10556.

E. coli strains were grown in LB broth. Streptococcal strains were cultured in unmodified Todd-Hewitt broth

* Corresponding author.

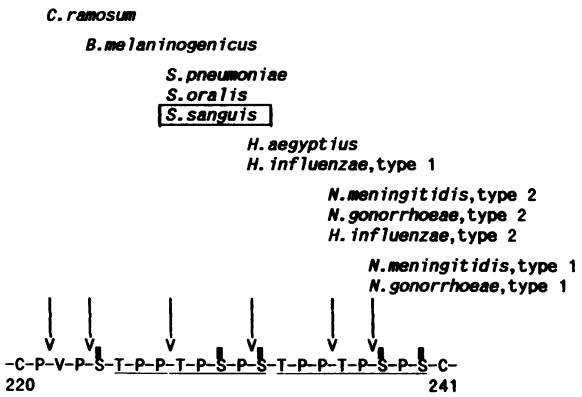

FIG. 1. Diagram of the human IgA1 heavy-chain hinge region showing sites cleaved by the various bacterial IgA proteases, highlighting that of *S. sanguis*. *Streptococcus oralis* was formerly *Streptococcus mitior*. The black rectangles above serine residues depict sites of glycosylation. Numbering of amino acids is from reference 55.

(BBL Microbiology Systems, Cockeysville, Md.). The following antibiotics were added to media when appropriate: ampicillin, 100 μg/ml; erythromycin, 200 μg/ml for *E. coli* and 10 μg/ml for *S. sanguis*; kanamycin, 30 μg/ml; and chloramphenicol, 30 μg/ml.

Cells in the logarithmic phase were subjected to cold osmotic shock and fractionated by the method of Neu and Heppel (38).

Preparation and processing of DNA. Methods for isolation of DNA of *S. sanguis* and of plasmid DNA from *E. coli* have been described previously (4, 6). Recombinant M13 plaques were purified and phage DNA was prepared as described elsewhere (34). Restriction endonucleases were purchased from New England BioLabs, Inc. (Beverly, Mass.), and Bethesda Research Laboratories Life Technologies, Inc. (Gaithersburg, Md.), and used as recommended by the suppliers. Gel electrophoresis was performed in 1% agarose gels. Streptococcal transformation was done by the method of Gaustad et al. (14), and *E. coli* transformations were done by the calcium chloride shock method of Berman et al. (3).

DNA sequencing. DNA sequencing was done by the dideoxy-chain termination method of Sanger et al. (50) with the Sequenase kit (U.S. Biochemical Corp., Cleveland, Ohio), using the universal M13 primer or several oligonucleotide primers whose synthesis was based on sequence data. DNA oligonucleotides were prepared by automated DNA synthesis or were obtained commercially (Genosys Biotechnologies, Inc., The Woodlands, Tex.) and were used in most cases without further purification.

DNA sequence analysis. Sequence information was analyzed by the James M. Pustell DNA-protein sequencing program (International Biotechnologies, Inc., New Haven, Conn.) and by the Genetics Computer Group program from the University of Wisconsin. The NBRF protein data base and Transgen (which also contains translated nucleotide sequences present in GenBank) were used to compare our deduced amino acid sequence with other known sequences. Programs used for analysis were FastA, GAP, Compare, and Peptidestructure.

Southern hybridization. DNAs cut with appropriate restriction endonucleases were electrophoresed in 1% agarose gels, transferred to Nytran membranes (Schleicher & Schuell, Keene, N.H.), and hybridized (53) with digoxigenin-labeled DNA probes. DNA probes were labeled by random priming, using the Genius nonradioactive DNA labeling kit (Boehringer Mannheim Biochemicals, Indianapolis, Ind.). Labeled DNA on the Nytran membranes was detected by enzyme-linked immunoassay, using an anti-digoxigenin-alkaline phosphatase conjugate developed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate.

Primer extension. Total cellular RNA was isolated from both *E. coli* and *S. sanguis* from mid-log-phase cultures. Before use, RNA was treated with RNase-free DNase from Promega Corp. (Madison, Wis.). An 80-μg sample of RNA was heated at 95°C for 5 min with 100 pmol of synthetic primer that had been kinase labeled with [γ-^{32}P]ATP. The template-primer hybrids were incubated at 45°C for 30 min, after which was added 20 U of avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, Fla.) in 0.05 M Tris hydrochloride (pH 8.3) buffer containing 0.04 M KCl, 1.0 mM dithiothreitol, 6.0 mM $MgCl_2$, and 0.625 mM deoxynucleoside triphosphates. Following further incubation for 20 min, the samples were ethanol precipitated, resuspended in 5.0 M urea–0.1% bromophenol blue–0.1% xylene cyanol, and electrophoresed on 6% acrylamide-urea gels (32).

Polymerase chain reaction. To amplify double-stranded streptococcal DNA, 10 separate polymerase chain reactions were done (49). Each reaction mixture contained 100 ng of chromosomal DNA, 50 pmol (500 nM) of oligonucleotide primers A (5′-CAAACCAGGTCGTAAAAT-3′) and B (5′-GAATTCGCTCTTTCGGCTCTTGAGC-3′), 4 U of *Taq* polymerase (International Biotechnologies), and 25 μM each deoxynucleoside triphosphate in 10 mM Tris hydrochloride (pH 8.3)–50 mM KCl–1.5 mM $MgCl_2$–0.01% gelatin–0.01% Tween, 20–0.01% Nonidet P-40. Primer B was synthesized with six additional bases on the 5′ end to add an *Eco*RI restriction site for future work. Thirty cycles of the amplification reaction were carried out in a Perkin Elmer-Cetus (Emeryville, Calif.) DNA Thermal Cycler. DNA was denatured at 94°C for 1 min, annealed at 55°C for 2 min, and extended at 72°C for 2 min. The 10 samples were pooled and treated twice with phenol-chloroform, after which the DNA was ethanol precipitated, reconstituted in 50 μl of 0.01 M Tris hydrochloride (pH 8.0)–0.01 M EDTA (TE), and electrophoresed in low-melting-point agarose (SeaPlaque agarose; FMC Bioproducts, Rockland, Maine). DNA of the appropriate size was cut from the gel, eluted, extracted with phenol-chloroform, precipitated with ethanol, and resuspended in 50 μl of TE.

Site-specific mutagenesis. For site-specific mutagenesis, we used the Muta-Gene in vitro kits (Bio-Rad Laboratories, Richmond, Calif.) in a procedure based on the method described by Kunkel et al. (27). *E. coli* CJ236 (*dut ung*) was infected with an M13 clone containing a 400-bp *Sal*I-*Bam*HI fragment of pJG1 to prepare single-stranded uracil-containing DNA. This was used as a template for in vitro synthesis of the desired mutant DNA in conjunction with the mutagenic oligonucleotide 5′-CC TAT ACC TTT AAA ATG ACG C-3′. When the double-stranded DNA is transformed into *E. coli* MV1190, the uracil-containing, nonmutated template strand is inactivated, leaving the mutated strand to replicate. The mutant DNA introduced a new restriction site, *Dra*I, which was used to screen the transformants. The presence of the mutation was confirmed by the chain termination sequencing method of Sanger et al. (50). Plasmid pJG1 was reconstructed to include this mutant DNA by replacing its

*Sal*I-*Bam*HI fragment with the mutant *Sal*I-*Bam*HI fragment to create plasmid pJG8.

Western immunoblotting analysis. Samples were subjected to electrophoresis on a 5% polyacrylamide gel in the presence of 0.1% sodium dodecyl sulfate (SDS) according to the method of Laemmli (29) and blotted by the method of Towbin et al. (54), using a Transblot cell (Bio-Rad). For the immunological detection of IgA protease protein, we used the natural antiprotease antibody in human milk (16). Milk centrifuged to remove cells and lipids was used from a single volunteer as the first antibody, and goat anti-human IgA–alkaline phosphatase conjugate (Sigma Corp., St. Louis, Mo.) was used as the second antibody. Protein bands were developed by incubation with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate.

Amino acid sequencing. The amino acid sequence of protein blotted onto polyvinyl membranes was determined (33) by automatic Edman degradation, using a model 477A pulsed-liquid sequenator (Applied Biosystems, Foster City, Calif.), in the Microsequencing Facility at Tufts University School of Medicine.

Assay of IgA protease activity. Protease activity was assayed by using human IgA1 serum protein as the substrate, as described in detail elsewhere (15).

RESULTS

Sequence analysis of the *iga* gene. Plasmid pJG1 is a derivative of pBR322 carrying a 9-kb insert of *S. sanguis* DNA containing the *iga* gene (15). Specific restriction fragments of the streptococcal insert in pJG1 were cloned into the replicative form of the M13mp series of vectors and sequenced as described in Materials and Methods.

The sequence analysis identified a 5,655-bp open reading frame, at least a part of which encodes the *S. sanguis* IgA protease. The nucleotide and deduced amino acid sequences are shown in Fig. 2.

To determine the translation start site for IgA protease within the open reading frame, and to verify that the gene was translated in the correct reading frame, we did limited microsequence analysis of the amino terminus of the active enzyme protein obtained from a periplasmic fraction of *E. coli* containing pJG1. The amino terminus had the sequence -Met-Asp-Lys-Glu-Ala-Leu-Asn-Gln-Asn-. Nucleotides encoding the stretch Asp through Asn were identified in the *iga* gene sequence (Fig. 2). The codon immediately amino terminal to this segment was GTG, which could be identified as the translation start site because GTG, which typically encodes valine, specifies methionine only when it occurs in the first translated position (25). A putative ribosome-binding site having the sequence AGAATG was identified 6 bases upstream of the initiation codon.

The amino acid sequence deduced from the cloned gene gave a calculated molecular mass for the protease of 186 kDa. Table 1 shows the amino acid composition of the protein. Typical for extracellular proteins, the *S. sanguis* IgA protease was found to contain no cysteine residues. Codon usage (Table 1) showed a mol% G+C content of 40.2. The calculated mol% G+C content for *S. sanguis* 10556 had earlier been reported as 45.3 (7). A or T was found in the third position in 65% of the codons, a high figure characteristic of streptococcal genes.

Primer extension. We mapped the 5′ end of the *iga* mRNA expressed in *E. coli* by primer extension, using as the primer the synthetic oligonucleotide 5′-GTATTGACGCCAACC GT-3′, which is the complement to nucleotides 751 through 767 in the *iga* gene sequence (Fig. 2). Primer extension was carried out with RNA both from *E. coli* MM294 containing pJG1 and from *S. sanguis* 10556 (Fig. 3). Primer extension using *E. coli* RNA gave rise to a single DNA product corresponding to a transcriptional start site at position 538 ± 2 bases. Just upstream of this position is a structure, typical of *E. coli* promoters, with a TATAAT at the −10 position and separated by 17 bp from a TGGAAA sequence at the −35 position. The same DNA product was not obtained by primer extension using *S. sanguis* RNA. In fact, only minor amounts of larger DNA products were obtained, giving no clear indication of the transcriptional start site in *S. sanguis*. Therefore, it seems likely that the transcription start points of the *iga* gene in *E. coli* and *S. sanguis* are different.

Identification of a lengthy tandem repeat in the enzyme protein. Analysis of the deduced amino acid sequence showed the presence of a region containing multiple tandem repeats extending from residues 139 to 336. This region consists of 10 segments of 20 amino acids each, having a high degree of homology to one another. Where these repeats are aligned for maximum homology, at 6 of the 20 positions the specified amino acids are invariant; at 4 additional positions there is 90% homology, and at 4 of the remaining positions there is 80% homology (Fig. 4). Also notable is that 48 of the 198 amino acids in the repeat section are either glycine or proline, and 28 are glutamic acid. The only positively charged residue in the repeat region is the lysine at position 258. Although the high content of glutamic acid would be expected to impart an overall negative charge to this region, this region in the hydropathy plot (Fig. 5) is, in general, largely hydrophobic.

The presence of a tandem repeat of this length raised the possibilities that duplications or deletions had occurred during cloning and that the DNA sequence determined by using M13 clones might not accurately define the nucleotide sequence of the wild-type gene. To verify that the cloned repeat segment present in pJG1 and in the M13 derivative was identical to chromosomal DNA of *S. sanguis* 10556, we used the polymerase chain reaction technique to examine the chromosomal repeats. The objective was to amplify the region of *S. sanguis* chromosomal DNA that included the entire repeat to allow comparison of its length with that of the corresponding segment in pJG1. Using two oligonucleotides, primer A, which hybridized 166 bases upstream of the start of the repeat, and primer B, which hybridized to the opposite DNA strand 27 bases downstream of the end of the tandem repeats, we amplified the repeat region present in *S. sanguis* chromosomal DNA. The amplified DNA was digested with the restriction enzymes *Pst*I and *Hin*dIII to generate a 630-bp fragment whose size was compared with those of the corresponding fragments obtained both from restriction digests of pJG1 DNA and from an M13 clone containing the repeat segment. The sizes of all fragments were found to be identical, within the limits of resolution of the gel system (1% agarose). The loss or gain of one or more repeat segments (60 nucleotides) would have been readily detected by using this system.

Apparent absence of a signal sequence in the enzyme encoded by *iga* DNA in *E. coli*. The deduced amino acid sequence of the *S. sanguis* IgA protease lacks an N-terminal signal sequence, although the enzyme produced in *E. coli* is localized mainly in the periplasm. A hydropathy plot of the deduced amino acid sequence specified by the entire open reading frame (Fig. 5) shows three lengthy hydrophobic sequences upstream of the GTG used as the translational start in *E. coli*. The hydrophobic sequence farthest upstream

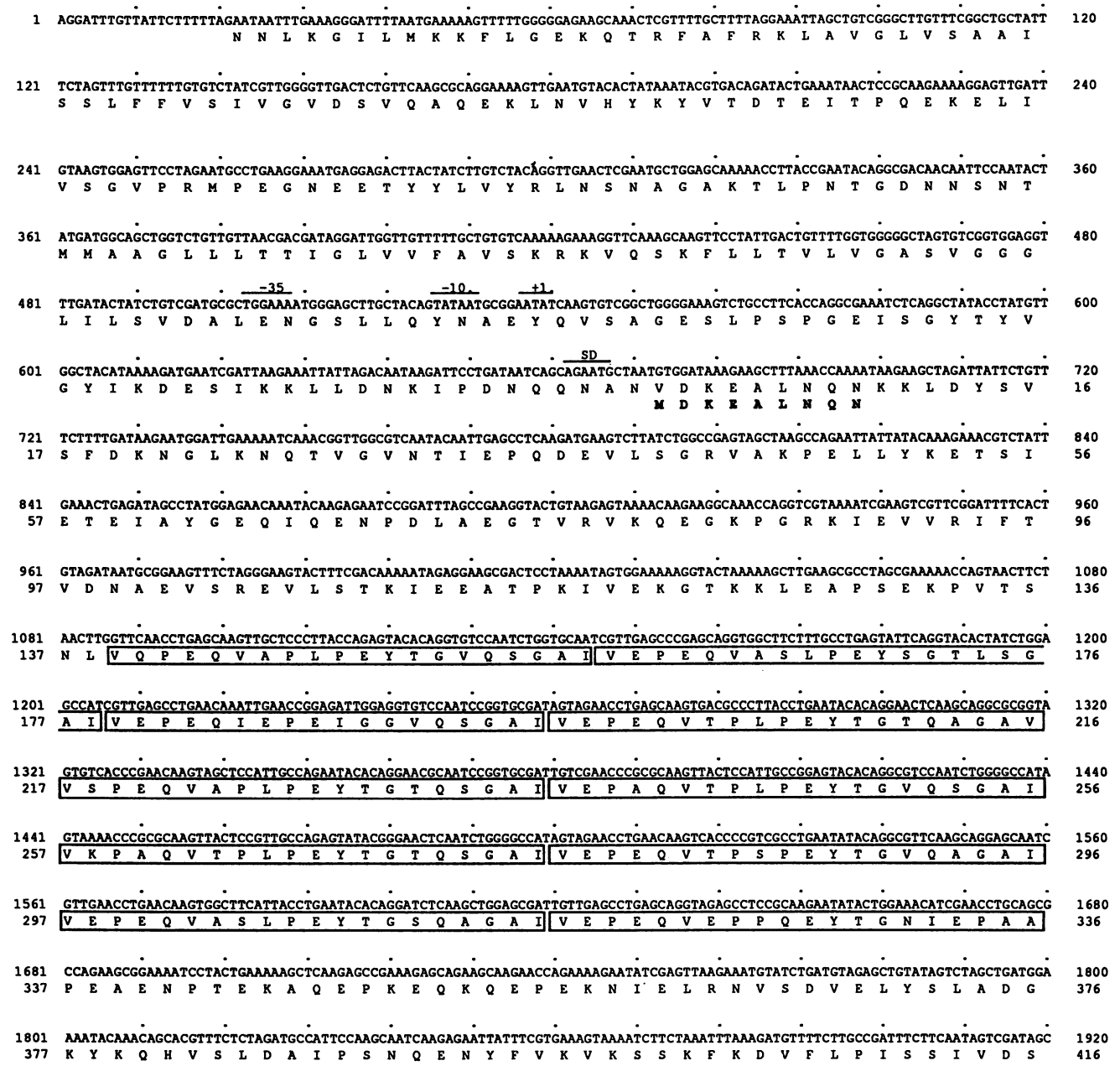

FIG. 2. Nucleotide and deduced amino acid sequences of the *S. sanguis* 10556 *iga* gene. Sequence was determined by the dideoxy-chain termination method on M13 clones containing restriction fragments of plasmid pJG1. Shown are the +1 start site for transcription, the −10 and −35 promoter region, and the proposed ribosome-binding site (SD). The amino acid sequence MDKEALNQN is the amino terminus of the enzyme protein derived from pJG1 in *E. coli*, as determined by microsequencing. The boxed segments beginning at residue 139 are the 10 tandem repeats. The putative zinc-binding site is at the boxed residues HEMTH and at E, at positions 1284 to 1288 and 1308, respectively. F and K are the amino acids that replaced H and E at positions 1284 and 1285 in pJG8 to create an inactive IgA protease.

immediately follows a methionine and has features typical of a streptococcal signal sequence. This region begins at nucleotide 43 in the DNA sequence (Fig. 2). The role of these sequences in expression of IgA protease in *S. sanguis* has yet to be established.

Relationship of the amino acid sequence of *S. sanguis* IgA protease to the sequences of other IgA proteases. Analysis of the amino acid sequence of the *S. sanguis* protease showed no regions of any length that were obviously homologous to the sequences of other known proteins, including IgA proteases of other bacterial genera. When the *S. sanguis* protease sequence was directly compared with that of *N. gonorrhoeae* type 2 protease in the NBRF data base (accession number A2069) by using program GAP, 19% of the residues were found to be identical. Using this figure, we calculated a normalized alignment score (9) of 95, indicating that a significant homology is improbable.

Identification and site-specific mutagenesis of a putative zinc-binding region in the *S. sanguis* IgA protease. The *S. sanguis* gene encodes the internal pentapeptide sequence

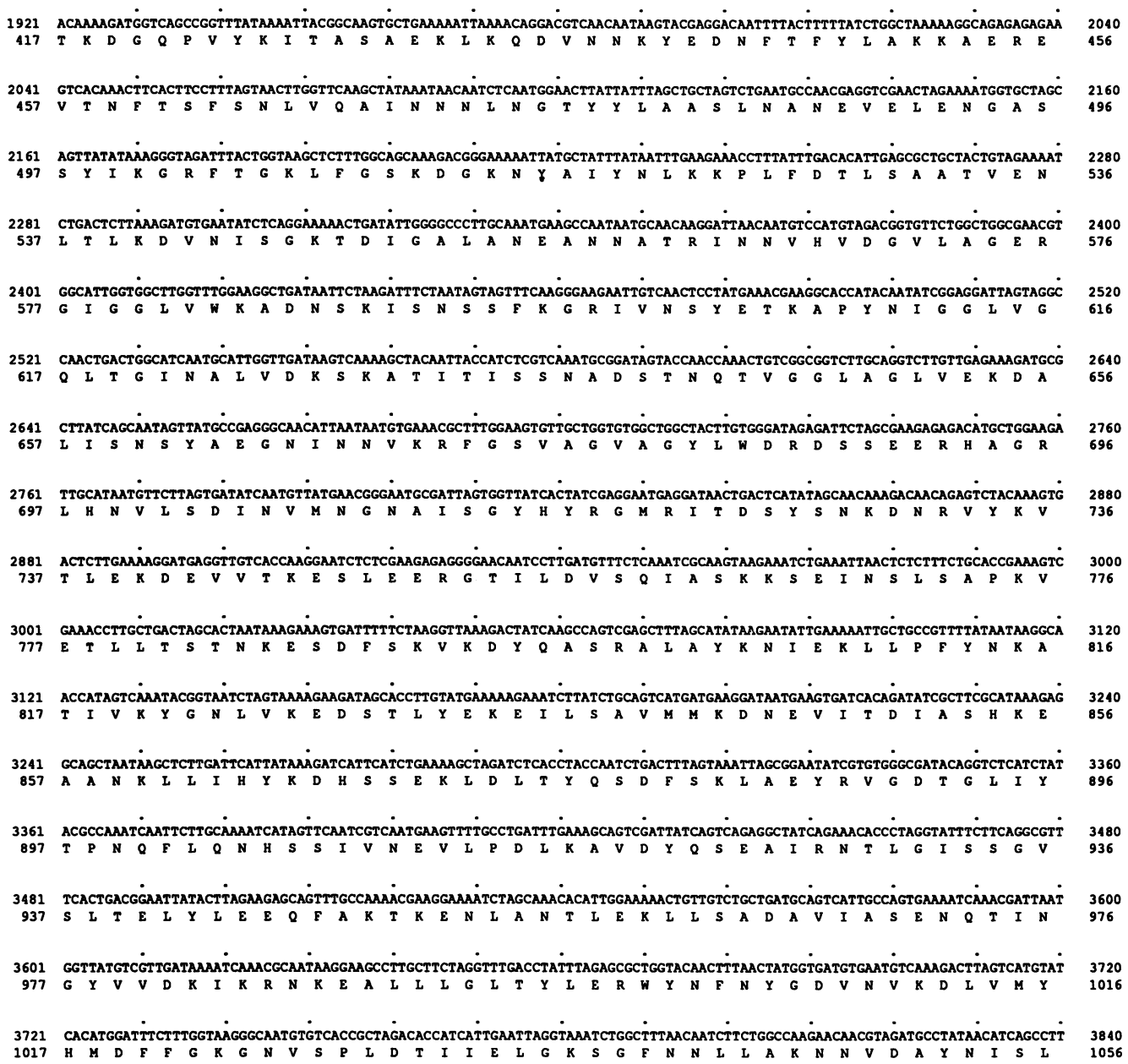

FIG. 2—*Continued.*

His-Glu-Met-Thr-His near the carboxy terminus, beginning with amino acid 1284. Twenty residues carboxy terminal to histidine 1288 is a glutamic acid residue. These elements of the primary structure conform to a zinc-binding signature that is characteristic of metalloproteases of both gram-negative and gram-positive bacteria (21, 57). Analyses of X-ray crystal structures of other bacterial metalloproteases, such as thermolysin, have confirmed that the pentapeptide and an appropriately spaced glutamic acid collaborate to form the zinc-binding site. The *S. sanguis* sequence in this region is highly similar to sequences of other known neutral metalloendoproteases of gram-positive bacteria (Table 2).

To obtain further evidence that the closely spaced histidine residues in the pentapeptide are essential for enzyme activity, we used the technique of oligonucleotide-directed in vitro mutagenesis (see Materials and Methods) to substitute phenylalanine for histidine at position 1284 and lysine for glutamic acid at position 1285, thus changing the consensus pentapeptide to Phe-Lys-Met-Thr-His. *E. coli* cells containing plasmid pJG8 with the mutant *iga* gene were assayed for protease activity and found to be negative. To verify that a mutant protein was synthesized, we examined periplasmic proteins encoded by pJG8 in *E. coli* and compared them by Western blotting with those encoded by pJG1. With human milk used as the first antibody, a protein of identical size (approximately 200 kDa) was detected in each preparation but was not present in *E. coli* MC1000 having no plasmid (Fig. 6).

Construction of an IgA protease-negative mutant of *S. sanguis* 10556. Using Tn*5* insertions, we had previously

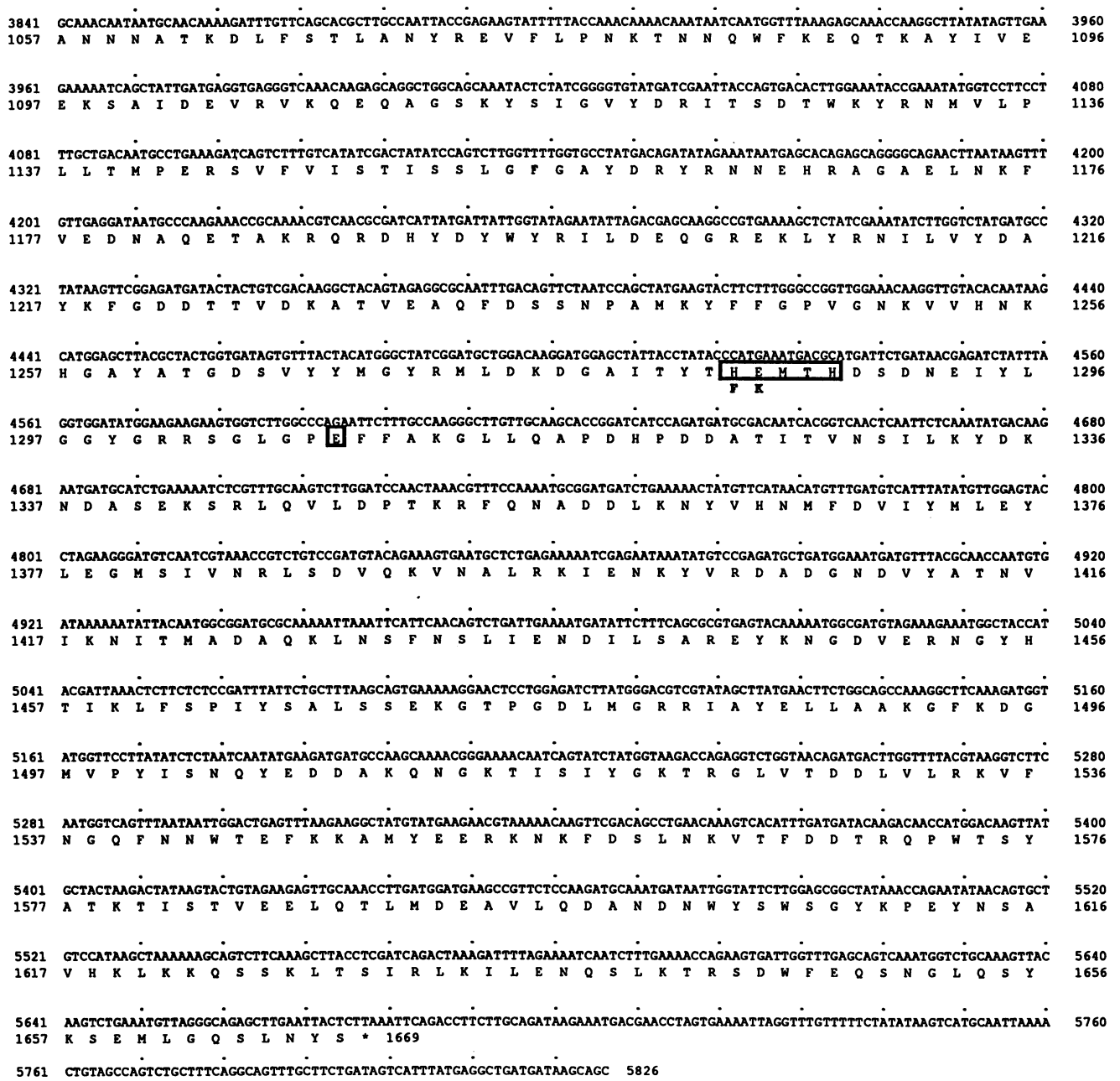

FIG. 2—*Continued.*

shown that the *Xba*I-*Bam*HI fragment of pJG1 and pJG2 was essential for protease activity (15). To introduce a deletion mutation into the *iga* gene, we replaced the *Xba*I-*Bam*HI fragment in pJG2 with an *Xba*I-*Bam*HI fragment from plasmid pVA891, encoding erythromycin resistance, to create pJG891. This construction (Fig. 7) has the potential to produce a protein containing the first 384 amino acids of the IgA protease, but in *E. coli*(pJG891) it lacked the ability to encode active protease.

To construct an enzymatically inactive mutant strain of *S. sanguis* 10556, we transformed this strain with pJG891 carrying the defective gene. Fifty recombinants selected on Todd-Hewitt agar plates containing 10 μg of erythromycin per ml were all found to be IgA protease negative. One chosen for further study was designated 10556N3. DNA hybridization analysis using pJG1 as a probe was used to confirm the genotype of 10556N3 (Fig. 8). Deletion of the *iga* sequences had no apparent effect on cell growth and caused no morphologic change of either cell shape or chain length, nor did it change the score in the Apizyme streptococcal screening assay (API 20S system; Analytab Products, Plainview, N.Y.).

DISCUSSION

S. sanguis is clinically important as a cause of bacterial endocarditis and as an early constituent of dental plaque, the microbial mass leading to caries formation. Because immu-

TABLE 1. Codon usage in the *S. sanguis* 10556 *iga* gene

Codon	Amino acid[a]	No.[b]	%[c]	Codon	Amino acid	No.	%	Codon	Amino acid	No.	%	Codon	Amino acid	No.	%
TTT	F	29	1.7	TCT	S	45	2.7	TAT	Y	61	3.7	TGT	C	0	0
TTC	F	17	1.0	TCC	S	6	0.4	TAC	Y	28	1.7	TGC	C	0	0
TTA	L	29	1.7	TCA	S	22	1.3	TAA		0		TGA		0	
TTG	L	38	2.3	TCG	S	6	0.4	TAG		0		TGG	W	11	0.7
CTT	L	23	1.4	CCT	P	22	1.3	CAT	H	15	0.9	CGT	R	13	0.9
CTC	L	10	0.6	CCC	P	6	0.4	CAC	H	5	0.3	CGC	R	4	0.2
CTA	L	14	0.8	CCA	P	20	1.2	CAA	Q	54	3.2	CGA	R	3	0.5
CTG	L	21	1.3	CCG	P	15	0.9	CAG	Q	14	0.8	CGG	R	2	0.1
ATT	I	37	2.2	ACT	T	36	2.2	AAT	N	86	5.2	AGT	S	31	1.9
ATC	I	33	2.0	ACC	T	21	1.3	AAC	N	37	2.2	AGC	S	18	1.1
ATA	I	19	1.1	ACA	T	29	1.7	AAA	K	81	4.9	AGA	R	23	1.4
ATG	M	21	1.3	ACG	T	15	0.9	AAG	K	53	3.2	AGG	R	5	0.3
GTT	V	40	2.4	GCT	A	44	2.6	GAT	D	76	4.6	GGT	G	37	2.2
GTC	V	39	2.3	GCC	A	24	1.4	GAC	D	20	1.2	GGC	G	28	1.7
GTA	V	28	1.7	GCA	A	31	1.9	GAA	E	88	5.3	GGA	G	31	1.9
GTG	V	19	1.1	GCG	A	21	1.2	GAG	E	48	2.9	GGG	G	11	0.7

[a] Given in single-letter code.
[b] Total number of amino acids specified.
[c] Percent amino acids in the enzyme specified by the codon.

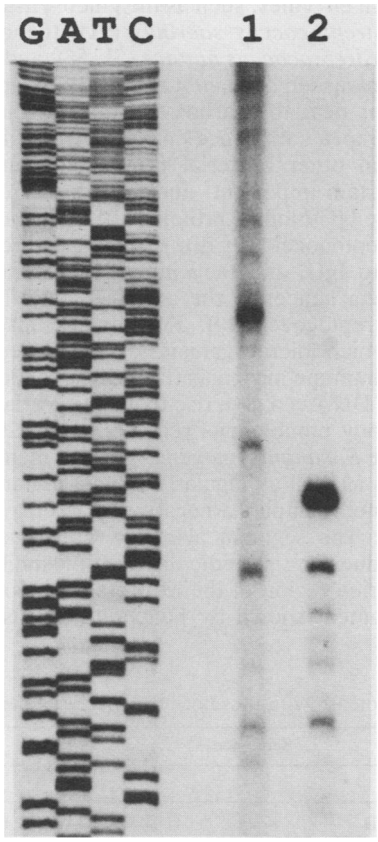

FIG. 3. Primer extension analysis of *iga* gene transcripts from *S. sanguis* 10556 (lane 1) and *E. coli* containing pJG1 (lane 2). Lanes G, A, T, and C represent a series of four sequencing reactions carried out by using an M13 clone of pJG1. Those were run alongside the primer extension reactions to determine the precise transcriptional start sites.

nity in the oral cavity is in part dependent on salivary secretory IgA antibodies, proteolytic cleavage of IgA1 in plaque and mixed saliva can potentially contribute to infection. A recent reclassification of oral *S. sanguis* strains has shown that colonization of the tooth surface is nearly always by species that elaborate IgA proteases (46). Such IgA proteolysis would not only reduce effective levels of secretory antibodies but also yield Fab fragments that can retain antigen-binding function. The binding of IgA Fab fragments can block access of other immunoglobulin isotypes mediating host effector functions against these microbial pathogens

V Q P E Q V A P L P E Y T G V Q S G A I
139

V E P E Q V A S L P E Y S G T L S G A I
159

V E P E Q - I - E P E I G G V Q S G A I
179

V E P E Q V T P L P E Y T G T Q A G A V
197

V S P E Q V A P L P E Y T G T Q S G A I
217

V E P A Q V T P L P E Y T G V Q S G A I
237

V K P A Q V T P L P E Y T G T Q S G A I
257

V E P E Q V T P S P E Y T G V Q A G A I
277

V E P E Q V A S L P E Y T G S Q A G A I
297

V E P E Q V E P P Q E Y T G N I E P A A
317

FIG. 4. Tandem repeats in *S. sanguis* IgA protease deduced from the nucleotide sequence of the *iga* gene in pJG1. Numbers are from the enzyme sequence in Fig. 2. Gaps flanking residue 184 are entered to maximize alignment.

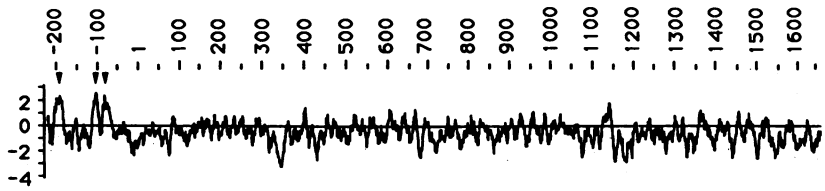

FIG. 5. Kyte-Doolittle hydropathy plot of the amino acid sequence deduced from the entire open reading frame of the *iga* gene in *S. sanguis*. The *y* axis represents the hydrophobic index. There is an interval index of 11 amino acids. Hydrophobic regions are above and hydrophilic regions below the line. Amino acid numbers are the same as in Fig. 2, so 1 is the GTG translational start in *E. coli*. Negatively numbered amino acids are in the open reading frame but are not numbered in Fig. 2, being upstream of the translational start. The arrows identify the hydrophobic regions discussed in the text.

(45, 48). Also, Reinholdt and Kilian (45) have suggested that hydrophobic, monomeric Fab-alpha fragments bound to the bacterial cell surface can foster bacterial adherence to tissues in the oral cavity.

The most interesting feature of the deduced amino acid sequence of the *S. sanguis* IgA protease presented in this work is its lack of significant homology with the IgA proteases of *Haemophilus* and *Neisseria* species. This finding suggests an independent evolution of these enzymes and emphasizes the role for this proteolytic activity in colonization or infection. Whether the *S. sanguis iga* gene shares homology with the corresponding gene in other gram-positive bacteria, such as pneumococci, is not yet known.

An important structural feature of the *S. sanguis* enzyme is the pentapeptide H-E-M-T-H, analogous to the H-x-x-T-H zinc-binding signature in other metalloproteases (21). Direct evidence for the involvement of this pentapeptide in catalytic activity of the *S. sanguis* enzyme was obtained by altering it by site-specific mutagenesis to the sequence F-K-M-T-H. This led to complete loss of enzyme activity despite the fact that normal amounts of the mutant polypeptide were produced in *E. coli*. This approach was used by Devault et al., who found that mutations in a similar zinc-binding pentapeptide region markedly reduced activity of the neutral endopeptidase 24.11 (8). However, crystallographic studies such as those with the metalloprotease thermolysin will be needed to definitively localize the pentapeptide to the active site and exclude nonspecific conformational changes as responsible for loss of catalytic activity. Such studies may also reveal which residues near the enzyme active site are involved in substrate recognition, important information that will guide the design of inhibitors for the *S. sanguis* enzyme.

Early difficulties in defining proteolytic enzymes on the basis of their substrate specificities led Hartley (18) and Barrett (2) to classify proteases according to their catalytic mechanisms into serine, cysteine, metallo-, and aspartic types. Having identified the *S. sanguis* IgA protease as a metalloenzyme, it is now clear that three of the four known mechanisms are represented among these enzymes. Mortensen and Kilian (35) have identified the *B. melaninogenicus* IgA protease as a thiol-activated enzyme; as discussed earlier, the *Haemophilus* and *Neisseria* proteases are of the serine type. This leaves only the aspartic type not represented among the known bacterial IgA proteases, but we know of no example of an aspartic protease of any specificity having been identified in bacteria (5).

The tandem repeats in the *S. sanguis iga* gene are of special interest. No similar structure has been found in the full-length *iga* gene sequences of other species, but repeated regions are being increasingly recognized in extracellular and surface-associated proteins of other pathogenic streptococci. For example, the nucleotide sequences of a number of streptococcal enzymes, such as the glucosyltransferase gene *gtfI* from *Streptococcus sobrinus* (11), the *gtfB* and *gtfC* genes from *Streptococcus mutans* (51, 56), and the *lytA* gene encoding amidase in *S. pneumoniae* (13), all specify tandem repeats at or near the carboxy termini of these enzymes. How these repeats influence enzyme activity is not entirely clear, but in other bacterial proteins repeats have been assigned certain important, and varied, functions. For example, in the IgG-binding proteins of *Staphylococcus aureus* (52) and streptococci (10, 19), repeats are directly involved in binding to IgG, and they impart structural stability by creating alpha helices in the antiphagocytic M proteins of group A streptococci (40). Repeats are also a potential means by which microorganisms and their derived proteins evade host immune mechanisms because tandemly repeated stretches of DNA can give rise to genetic variability through high-frequency homologous recombination (20). The DNA repeat in the *S. sanguis iga* gene specifies mainly hydrophobic residues as well as regularly spaced glutamic acids that will undoubtedly impart a negative charge to this region of the protein. The segment also has numerous proline and glycine residues in a periodic distribution similar to those in the proline-rich region of the M protein of group A streptococci, a segment shown by Pancholi and Fischetti (39) to

TABLE 2. Comparison of the zinc-binding components in *S. sanguis* 10556 IgA protease with those of other bacterial metalloproteases

Organism	Sequence																				
	1284[a]					1288													1308		
Streptococcus sanguis	T	Y	T	H	E	M	T	H	D	S	D				G	L	G	P	E	F	F	A	K
Bacillus thermoproteolyticus	V	V	A	H	E	L	T	H	A	V	T				G	A	I	N	E	A	I	S	D
Bacillus cereus	V	I	G	H	E	L	T	H	A	V	T				G	A	L	N	E	A	I	S	D
Bacillus stearothermophilus	V	V	G	H	E	L	T	H	A	V	T				G	A	I	N	E	A	M	S	D
Bacillus subtilis	V	T	A	H	E	M	T	H	G	V	T				G	A	L	N	E	S	F	S	D
Bacillus amyloliquefaciens	V	T	A	H	E	M	T	H	G	V	T				G	A	L	N	E	S	F	S	D

[a] Numbered amino acids refer only to the *S. sanguis* sequence.

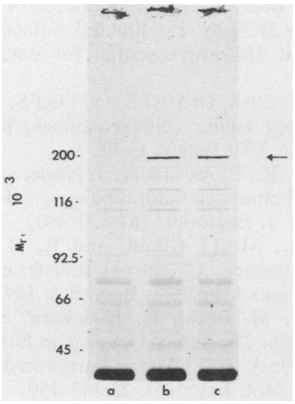

FIG. 6. Western blots of SDS-polyacrylamide gel electrophoresis-fractionated extracts of *E. coli* containing no plasmid (lane a), pJG1 (lane b), and pJG8 (lane c) developed with human milk as the first antibody. The enzyme proteins (arrow) are both approximately 200 kDa, although only the protease of pJG1 has activity.

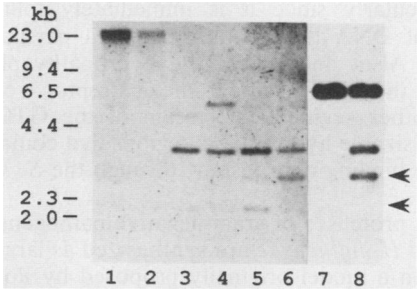

FIG. 8. Southern hybridization of wild-type and mutant *S. sanguis* chromosomal DNA digested with restriction enzymes and probed with digoxigenin-labeled pJG1. Arrows show the *Xba*I-*Bam*HI fragments whose sizes verify the replacement of a segment of the wild-type *iga* gene. Lanes: 1, 10556N3 digested with *Bam*HI; 2, 10556 (*Bam*HI); 3, 10556N3 (*Xba*I); 4, 10556 (*Xba*I); 5, 10556N3 (*Xba*I and *Bam*HI); 6, 10556 (*Xba*I and *Bam*HI); 7, pJG891 (*Xba*I and *Bam*HI); 8, pJG1 (*Xba*I and *Bam*HI). Size markers are lambda DNA digested with *Hin*dIII.

interact with the cross-linked peptidoglycan of the bacterial cell wall. We do not yet know whether the repeats in the *S. sanguis iga* gene are present on the fully processed and secreted enzyme. Further information on the fate of the repeat sequence in *S. sanguis* is needed before the role that it plays in the interaction of *S. sanguis* and its host can be determined.

We have noted certain unusual aspects of the synthesis and secretion of the *S. sanguis* IgA protease in *E. coli*, including the absence of an amino-terminal signal sequence and the absence of a precursor that gives rise to the active enzyme. As recently emphasized by Wandersman (58), these two conditions have invariably been met by all known extracellular proteases of gram-positive bacteria. The lack of an amino-terminal signal sequence leads to uncertainty as to how the enzyme enters the periplasmic space in recombinant *E. coli* or how it is secreted extracellularly by wild-type *S. sanguis*. Recognizable signal sequences are lacking in a number of proteins synthesized by both procaryotes and eucaryotes, e.g., proteases B and C of *Erwinia* species (30), metalloprotease of *Serratia marcescens* (37), alpha-hemolysin of *E. coli* (24), yeast a-factor (26), and the cytokine interleukin-1β (47), and secretion of these proteins apparently depends on carboxy-terminal residues or accessory gene products.

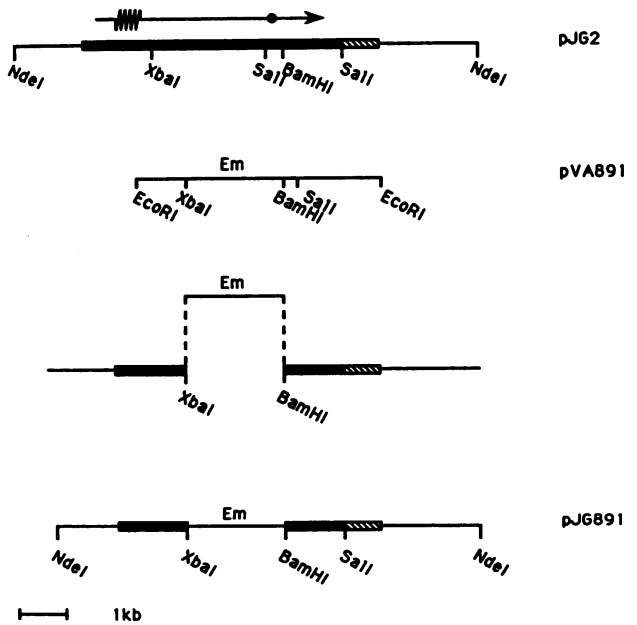

FIG. 7. Steps used to construct pJG891, a deletion mutation of pJG2, used to transform wild-type *S. sanguis* 10556 to protease-negative 10556N3. Symbols: ⟶, Length of the *iga* gene in pJG2; ∿∿∿, the repeat segment, ●, the putative zinc-binding site; ■, *S. sanguis* 10556; ▨, λ EMBL3.

Several lines of evidence indicate that pJG1 in *E. coli* does not encode a larger precursor. Direct evidence that the amino-terminal end of the encoded protein is not processed comes from the fact that the amino-terminal residue of the pJG1 enzyme recovered from the *E. coli* periplasm is a methionine encoded by GTG. Because GTG usually codes for valine but can encode methionine to initiate a polypeptide chain, we conclude that the recovered enzyme has not undergone amino-terminal processing. Thus, if processing of the encoded protein occurs at all, it would be at the carboxy-terminal end. Plasmid pJG8, which contains the *iga* gene having a site-specific mutation in the proposed zinc-binding region, encodes an inactive protein the same size as the active enzyme encoded by pJG1. This indicates that the size of the enzyme in *E. coli* is clearly independent of its catalytic activity, persuasive evidence that an autocatalytic-type processing does not occur. These data do not, however, exclude the possibility that other cellular mechanisms in *E. coli* carry out identical carboxy-terminal modifications of both the active and inactive enzyme proteins.

Our results with the streptococcal *iga* gene cloned into *E. coli* do not allow firm conclusions as to the translation start site and the secretion and processing steps in wild-type *S. sanguis* because we do not yet have an amino-terminal sequence of the native protein. Primer extension experiments show that the locations of the promoters in *E. coli* and *S. sanguis* are not the same, indicating that the *S. sanguis* protease is part of an operon or that transcription in *S. sanguis* actually does begin upstream of the start site in pJG1. In the 651-bp untranslated region of the open reading frame upstream of the GTG initiation codon used by *E. coli* there is an ATG (bp 43) that could be a translational start

site, particularly since it is immediately followed by a segment of DNA that would specify a persuasive signal sequence. Also, analysis of the hydropathy plot (Fig. 5) shows that the open reading frame of streptococcal *iga* DNA has two other segments upstream of the GTG start site specifying sizable hydrophobic regions that could contribute to enzyme binding or transport through the *S. sanguis* cell membrane.

The IgA proteases of gram-negative pathogens *N. gonorrhoeae* and *H. influenzae* are synthesized as large precursor proteins. In a model originally proposed by Pohlner et al. (43), once the signal peptide directs the precursor to the periplasm, the carboxy-terminal domain forms a pore in the bacterial outer membrane through which the active enzyme (the amino-terminal domain) is excreted. Final maturation and release are by autoproteolysis at proline-rich sequences within the precursor that are homologous with the IgA1 hinge region substrate (17, 43, 44). There is no information on the secretion mechanism for IgA proteases in gram-positive bacteria, but the published estimate for the mature *S. sanguis* enzyme as 100 kDa (28) suggests that processing occurs. We have analyzed the *iga* sequence in *S. sanguis* for possible autoproteolytic sites, and two areas are identified. One is the tetrapeptide -T-P-S-P- (residues 283 to 286) in the repeat, identical to the sequence in human IgA1 that is immediately carboxy terminal to the peptide bond cleaved by the *S. sanguis* protease (Fig. 1). Because this is clearly not cleaved by the active enzyme accumulating in the *E. coli* periplasm, we can conclude that it is not a processing site for the recombinant enzyme or that the periplasmic environment in *E. coli* is not optimal for processing. In *S. sanguis*, a propeptide form specified by additional upstream DNA may be required for autoproteolytic attack on this tetrapeptide. The other potential autocatalytic processing site, the tripeptide P-S-P 36 amino acids upstream of the GTG start site, is not translated in *E. coli*. Verification of either of these potential cleavage sites will require identification of the amino-terminal sequence of the *S. sanguis* enzyme.

ACKNOWLEDGMENTS

We thank Jiazhou Qiu, Rina Ben Yaacov, Carol Nichols, and Anne Kane for technical support and helpful suggestions.

This work was supported by Public Health Service grants DE07257, AI 20337, and P30-AM 39428 (the GRASP Center for Gastroenterology Research).

REFERENCES

1. **Bachovchin, W. W., A. G. Plaut, G. R. Flentke, M. Lynch, and C. A. Kettner.** 1990. Inhibition of IgA1 proteases from *Neisseria gonorrhoeae* and *Haemophilus influenzae* by peptide prolyl boronic acids. J. Biol. Chem. **265**:3738–3743.
2. **Barrett, A. J.** 1980. Introduction: the classificatin of proteinases. CIBA Found. Symp. **75**:1–13.
3. **Berman, M. L., L. W. Enquist, and T. J. Silhavy.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Birnboim, H., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–23.
5. **Bond, J. S., and P. E. Butler.** 1987. Intracellular proteases. Annu. Rev. Biochem. **56**:333–364.
6. **Coykendall, A. L.** 1970. Base composition of deoxyribonucleic acid isolated from cariogenic streptococci. Arch. Oral Biol. **15**:365–368.
7. **Coykendall, A. L., and P. A. Specht.** 1975. DNA base sequence homologies among strains of *Streptococcus sanguis*. J. Gen. Microbiol. **91**:92–98.
8. **Devault, A., V. Sales, C. Nault, A. Beaumont, B. Roques, P. Crine, and G. Boileau.** 1988. Exploration of the catalytic site of endopeptidase 24.11 by site-directed mutagenesis: histidine residues 583 and 587 are essential for catalysis. FEBS Lett. **231**:54–58.
9. **Doolittle, R. F.** 1986. Of URFS and ORFS: a primer on how to analyze derived amino acid sequences, p. 3–24. University Science Books, Mill Valley, Calif.
10. **Fahnestock, S. R., P. Alexander, J. Nagle, and D. Filpula.** 1986. Gene for an immunoglobulin-binding protein from a group G streptococcus. J. Bacteriol. **167**:870–880.
11. **Ferretti, J. J., M. L. Gilpin, and R. R. B. Russell.** 1987. Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. J. Bacteriol. **169**:4271–4278.
12. **Fujiyama, Y., M. Iwaki, K. Hodohara, S. Hosoda, and K. Kobayashi.** 1986. The site of cleavage in human alpha chains of IgA1 and IgA2: $A_2m(1)$ allotype paraproteins by the clostridial IgA protease. Mol. Immunol. **23**:147–150.
13. **Garcia, E., J. L. Garcia, P. Garcia, A. Arraras, J. M. Sanchez-Puelles, and R. Lopez.** 1988. Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. Proc. Natl. Acad. Sci. USA **85**:914–918.
14. **Gaustad, P., J. Eriksen, and S. D. Henriksen.** 1979. Genetic transformation in *Streptococcus sanguis*. Acta Pathol. Microbiol. Scand. **87**:117–122.
15. **Gilbert, J. V., A. G. Plaut, Y. Fishman, and A. Wright.** 1988. Cloning of the gene encoding streptococcal immunoglobulin A protease and its expression in *Escherichia coli*. Infect. Immun. **56**:1961–1966.
16. **Gilbert, J. V., A. G. Plaut, B. Longmaid, and M. E. Lamm.** 1983. Inhibition of microbial IgA proteases by human secretory IgA and serum. Mol. Immunol. **20**:1039–1049.
17. **Grundy, F. J., A. G. Plaut, and A. Wright.** 1987. *Haemophilus influenzae* immunoglobulin A1 protease genes: cloning by plasmid integration-excision, comparative analysis, and localization of secretion determinants. J. Bacteriol. **169**:4442–4450.
18. **Hartley, B. S.** 1960. Proteolytic enzymes. Annu. Rev. Biochem. **29**:45–72.
19. **Heath, D. G., and P. P. Cleary.** 1989. Fc-receptor and M protein genes of group A streptococci are products of gene duplication. Proc. Natl. Acad. Sci. USA **86**:4741–4745.
20. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207**:196–203.
21. **Jongeneel, C. V., J. Bouvier, and A. Bairoch.** 1989. A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett. **242**:211–214.
22. **Kilian, M., and K. Holmgren.** 1981. Ecology and nature of immunoglobulin A1 protease-producing streptococci in the human oral cavity and pharynx. Infect. Immun. **31**:868–873.
23. **Kilian, M., J. Mestecky, and R. E. Schrohenloher.** 1979. Pathogenic species of the genus *Haemophilus* and *Streptococcus pneumoniae* produce immunoglobulin A1 protease. Infect. Immun. **26**:143–149.
24. **Koronakis, V. E., and C. Hughes.** 1989. Isolation and analysis of the c-terminal signal directing export of *Escherichia coli* hemolysin protein across both bacterial membranes. EMBO J. **8**:595–605.
25. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
26. **Kuchler, K., R. E. Sterne, and J. Thorner.** 1989. *Saccharomyces cerevisiae* STE 6 gene product: a novel pathway for protein export in eukaryotic cells. EMBO J. **8**:3973–3984.
27. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.
28. **Labib, R. S., N. J. Calvanico and T. B. Tomasi, Jr.** 1978. Studies on extracellular proteases of *Streptococcus sanguis*: purification and characterization of a human IgA1 specific protease. Biochim. Biophys. Acta **526**:547–559.
29. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London)

227:680–685.
30. **Letoffe, S., P. Delepelaire, and C. Wandersman.** 1990. Protease secretion by *Erwinia chrysanthemi*: the specific secretion functions are analogous to those of *Escherichia coli* alpha-haemolysin. EMBO J. **9**:1375–1382.
31. **Macrina, F. L., R. P. Evans, J. A. Tobian, D. L. Hartley, D. B. Clewell, and K. R. Jones.** 1983. Novel shuttle plasmid vehicles for *Escherichia-Streptococcus* transgeneric cloning. Gene **25**: 145–150.
32. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262**:10035–10038.
34. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
35. **Mortensen, S. B., and M. Kilian.** 1984. Purification and characterization of an immunoglobulin A1 protease from *Bacteroides melaninogenicus*. Infect. Immun. **45**:550–557.
36. **Mulks, M. H., A. G. Plaut, H. A. Feldman, and B. Frangione.** 1980. IgA proteases of two distinct specificities are released by *Neisseria meningitidis*. J. Exp. Med. **152**:1442–1447.
37. **Nakahama, K., K. Yoshimura, R. Marumoto, M. Kikuchi, I. Sook-Lee, T. Hase, and H. Matsubara.** 1986. Cloning and sequencing of *Serratia* protease gene. Nucleic Acids Res. **14**:5843–5855.
38. **Neu, H. C., and L. A. Heppel.** 1965. The release of enzymes from *E. coli* by osmotic shock and during the formation of spheroplasts. J. Biol. Chem. **240**:3685–3692.
39. **Pancholi, V., and V. A. Fischetti.** 1988. Isolation and characterization of the cell-associated region of group A streptococcal M6 protein. J. Bacteriol. **170**:2618–2624.
40. **Phillips, G. N., P. F. Flicker, C. Cohen, B. N. Manjula, and V. A. Fischetti.** 1981. Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. Proc. Natl. Acad. Sci. USA **78**:4689–4693.
41. **Plaut, A. G., R. J. Genco, and T. B. Tomasi, Jr.** 1974. Isolation of an enzyme from *Streptococcus sanguis* which specifically cleaves IgA. J. Immunol. **113**:289–291.
42. **Plaut, A. G., J. V. Gilbert, M. S. Artenstein, and J. D. Capra.** 1975. *Neisseria gonorrhoeae* and *N. meningitidis*: production of an extracellular enzyme which cleaves human IgA. Science **190**:1103–1105.
43. **Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer.** 1987. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature (London) **325**:458–462.
44. **Poulsen, J., J. Brandt, P. Hjorth, H. C. Thogersen, and M. Kilian.** 1989. Cloning and sequencing of the immunoglobulin A1 protease gene (*iga*) of *Haemophilus influenzae* serotype b. Infect. Immun. **57**:3097–3105.
45. **Reinholdt, J., and M. Kilian.** 1987. Interference of IgA protease with the effect of secretory IgA on adherence of oral streptococci to saliva-coated hydroxyapatite. J. Dent. Res. **66**:492–97.
46. **Reinholdt, J., M. Tomana, S. B. Mortensen, and M. Kilian.** 1990. Molecular aspects of immunoglobulin A1 degradation by oral streptococci. Infect. Immun. **58**:1186–1194.
47. **Rubartelli, A., F. Cozzolino, M. Talio, and R. Sitia.** 1990. A novel secretory pathway for interleukin-1 beta, a protein lacking a signal sequence. EMBO J. **9**:1503–1510.
48. **Russell, M. W., J. Reinholdt, and M. Kilian.** 1989. Anti-inflammatory activity of human IgA antibodies and their Fab-alpha fragments: inhibition of IgG-mediated complement activation. Eur. J. Immunol. **19**:2243–2249.
49. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
50. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5643–5647.
51. **Shiroza, T., S. Ueda, and H. K. Kuramitsu.** 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. J. Bacteriol. **169**:4263–70.
52. **Sjodahl, J.** 1977. Repetitive sequences in protein A from *Staphylococcus aureus*. Eur. J. Biochem. **73**:343–351.
53. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
54. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
55. **Tsuzukida, Y., C. C. Wang, and F. W. Putnam.** 1979. Structure of the A2m(1) allotype of human IgA—a recombinant molecule. Proc. Natl. Acad. Sci. USA **76**:1104–1108.
56. **Ueda, S., T. Shiroza, and H. K. Kuramitsu.** 1988. Sequence analysis of the gtfC gene from *Streptococcus mutans* GS-5. Gene **69**:101–109.
57. **Vallee, B. L., and D. S. Auld.** 1990. Active-site zinc ligands and activated H_2O of zinc enzymes. Proc. Natl. Acad. Sci. USA **87**:220–224.
58. **Wandersman, C.** 1989. Secretion, processing and activation of bacterial extracellular proteases. Mol. Microbiol. **3**:1825–1831.
59. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.