

## High-Molecular-Weight Protein 2 of *Yersinia enterocolitica* Is Homologous to AngR of *Vibrio anguillarum* and Belongs to a Family of Proteins Involved in Nonribosomal Peptide Synthesis

INGRID GUILVOUT,<sup>1</sup> ODILE MERCEREAU-PUJALON,<sup>2</sup> SERGE BONNEFOY,<sup>2</sup>  
ANTHONY P. PUGSLEY,<sup>3</sup> AND ELISABETH CARNIEL<sup>1\*</sup>

Unité de Bactériologie Moléculaire et Médicale,<sup>1</sup> Unité de Parasitologie Expérimentale,<sup>2</sup> and Unité de Génétique Moléculaire,<sup>3</sup> Institut Pasteur, 75724 Paris Cedex 15, France

Received 6 May 1993/Accepted 28 June 1993

The iron-regulated *irp2* gene is specific for the highly pathogenic *Yersinia* species and encodes high-molecular-weight protein 2 (HMWP2). Despite the established correlation between the presence of HMWP2 and virulence, the role of this protein is still unknown. To gain insight into the function of HMWP2, the entire coding sequence and the promoter of *irp2* were sequenced. Two putative  $-35$  and  $-10$  promoter sequences were identified upstream of a large open reading frame, and two potential Fur-binding sites were found overlapping the second  $-35$  box. The large open reading frame is composed of 6,126 nucleotides and may encode a protein of 2,035 amino acids (ca. 228 kDa) with a pI of 5.81. A signal sequence was not present at the N terminus of the protein. Despite the existence of 30 cysteine residues, carboxymethylation prevented the formation of most if not all disulfide bonds that otherwise occurred when the cells were sonicated. The protein was composed of three main domains: a central region of ca. 850 residues, bordered on each side by a repeat of 550 residues. A high degree of identity (44.5%) was found between HMWP2 and the protein AngR of *Vibrio anguillarum*. The central part of HMWP2 (after removal of a loop of 337 residues) also displayed significant homology with proteins belonging to the superfamily of adenylate-forming enzymes and, like them, possessed a putative ATP-binding motif that is also present in AngR. In addition, HMWP2 shared with the group of antibiotic and enterochelin synthetases a potential amino acid-binding site. Six consensus sequences defining the superfamily and four defining the family of synthetases were derived from the multiple alignment of the 30 sequences of proteins or repeated domains. A phylogenetic tree that was constructed showed that HMWP2 and AngR are in a family composed of Lys2, EntF, and the tyrocidine, gramicidin, and  $\beta$ -lactam synthetases. This finding suggests that HMWP2 may participate in the nonribosomal synthesis of small biologically active peptides.

The environmental conditions encountered by pathogenic microorganisms in their hosts are different from those found in vitro. The organisms sense a number of parameters (temperature, pH, osmolarity, ion concentration, etc.) which may vary during the course of infection. Bacteria adapt to environmental changes through a series of induced responses (55, 61). Most if not all of these responses are regulated at the gene level: some genes or operons are coordinately switched on, while some others are turned off.

One of the best-studied bacterial responses induced in the host is the microbial adaptation to iron limitation (5). Laboratory growth media contain a large excess of iron, while in mammals, the low level of free iron ( $10^{-18}$  M) is not sufficient to sustain bacterial growth. To circumvent this iron limitation, most bacteria synthesize small excreted molecules called siderophores which chelate iron bound to specific eukaryotic proteins (e.g., transferrin and lactoferrin) and are then transported back into the microorganism. Some bacteria produce outer membrane proteins able to directly bind transferrin and/or lactoferrin and to capture the iron carried by these molecules (5). Finally, bacteria may also produce hemolysins which lyse erythrocytes and liberate

heme-bound Fe molecules (5). Synthesis of the proteins involved in these responses is induced by iron starvation.

The genus *Yersinia* is composed of 11 species, only 3 of which (*Yersinia pestis*, the agent of plague, and two enteric pathogens, *Y. enterocolitica* and *Y. pseudotuberculosis*) are pathogenic for humans. The role of iron in the virulence of *Y. pestis* was described almost 40 years ago (33), but iron-regulated proteins were identified only recently (11, 54, 60). No precise function has yet been attributed to the iron-regulated polypeptides of *Y. pestis* and *Y. pseudotuberculosis*. In *Y. enterocolitica*, the species which synthesizes the largest number of iron-regulated proteins (11), two of the proteins, FoxA (75.7 kDa) and FcuA (81.7 kDa), are the respective receptors for ferrioxamine and ferrichrome, hydroxamate siderophores utilized but not produced by *Y. enterocolitica* (3, 35). Several proteins are encoded by the *Y. enterocolitica* hemin uptake operon. These include HemR (78 kDa), the hemin receptor, HemS (42 kDa), which is either a cytoplasmic membrane permease or a hemin-degrading enzyme, and HemP (6.5 kDa) and HemT (27 kDa), both of which have unknown functions, although it is known that the latter is not necessary for hemin uptake (62). *Y. enterocolitica* also produces an approximately 80-kDa outer membrane protein which is structurally related to the FepA protein of *Escherichia coli* and therefore may serve as an enterochelin receptor (48). Finally, a 65-kDa iron-regulated

\* Corresponding author.

protein (FyuA) has been shown recently to be the receptor for the siderophore yersiniabactin and probably also for the bacteriocin pesticin (29).

All highly pathogenic *Yersinia* species produce two high-molecular weight iron-regulated proteins (high-molecular-weight proteins 1 and 2 [HMWP1 and HMWP2]) (11, 17). The *irp2* gene encoding HMWP2 is absent from weakly virulent strains of *Yersinia* species. The link between the presence of these proteins and virulence was strengthened by two additional observations: (i) highly pathogenic *Y. pestis* strains which spontaneously lost the ability to produce HMWPs were much less virulent when injected subcutaneously into mice (10), and (ii) a mutation in the *irp2* gene of *Y. pseudotuberculosis* caused the disappearance of both HMWPs and a marked decrease in virulence (9). Despite the relationship between pathogenicity and the presence of the HMWPs, the role of these proteins is still unknown.

To gain insight into the role of HMWP2, the *irp2* gene and its promoter region were sequenced. In this work, the nucleotide sequence of *irp2*, the deduced amino acid sequence of HMWP2, and some features of the *irp2* promoter and of the structure of HMWP2 are reported. The existence of potential disulfide bridges in HMWP2 was examined, and the location of the protein within the bacterium is discussed. We show that HMWP2 displays significant homology to AngR of *Vibrio anguillarum* and shares similarities with a superfamily of adenylate-forming enzymes. A role for HMWP2 is proposed and discussed.

## MATERIAL AND METHODS

**Bacterial strains and culture media.** The wild-type strain Ye8081 of *Y. enterocolitica* biotype 1B, serotype O:8, and two strains of *E. coli*, DH5 $\alpha$  and JM101 (1), were used in this study. pIR2 is a pUC18 recombinant with an 8-kb *Cla*I insert containing the entire *irp2* gene from strain Ye8081 (9). Phages M13mp18 and M13mp19 (1) were used for single-stranded DNA sequencing. The *Yersinia* strain was grown at 28°C in peptone broth, and *E. coli* was grown in Luria broth, in 2xYT, or on LB agar plates at 37°C (1). The chemically defined medium used for the extraction of the HMWPs was prepared as described previously (11). Iron repletion or starvation was obtained by addition of 150  $\mu$ M FeCl<sub>3</sub> or 50  $\mu$ M  $\alpha$ - $\alpha'$ -dipyridyl (Sigma Chemical Co.), respectively, to the chemically defined medium. Ampicillin (100  $\mu$ g/ml) was added when needed.

**Cloning procedures and sequencing.** To sequence the *irp2* gene of Ye8081 in pIR2, suitable overlapping inserts were obtained by different strategies. First, several small sub-clones were directly ligated into the replicative form of bacteriophage M13mp18 and M13mp19. Second, large inserts were treated with exonuclease III (Erase-a-Base system; Promega Biotec) to generate a set of smaller fragments (1) which were inserted in M13mp phage DNA. Third, oligonucleotides were synthesized for nonoverlapping regions, using a model 391 DNA synthesizer (Applied Biosystems, Inc., Foster City, Calif.). The nucleotide sequence of the two DNA strands was determined by the dideoxy-chain termination method (1) with use of a modified T7 DNA polymerase (Sequenase; United States Biochemical Corp.). Sequence ambiguities were resolved by using dITP instead of dGTP. The nucleotide and amino acid sequence data were analyzed with various softwares: DNA Strider (42), ClustalV (30), and the FastA, Bestfit, Motifs, and Pileup programs from the version 7 UNIX of the Genetics Computer Group package (18).

**DNA manipulations.** DNA electrophoresis, electroelution, ligation, and electroporation were carried out as previously described (13). Double-stranded plasmids were extracted according to the procedure of Birnboim and Doly (4a), and M13 single-stranded DNA was isolated by the method of Messing as described in reference 1.

**Primer extension.** *Y. enterocolitica* Ye8081 and *E. coli* (pIR2) were grown for 3 days in iron-rich or iron-depleted chemically defined medium, and total RNA was extracted by the hot phenol method of Von Gabain et al. (68), with some modifications (12). Primer extension experiments were performed with reverse transcriptase (Promega Biotec) (1). Three oligonucleotides were used as primers (Fig. 1): P1 (positions 241 to 270), P2 (445 to 469), and P3 (505 to 534). The DNA strands obtained after primer extension were electrophoresed alongside the DNA sequence reaction products obtained with the same primers.

**Cell fractionation and analysis of proteins.** To obtain total proteins, bacteria were washed three times in saline, and the pellet was heated to 100°C in a denaturing solubilization buffer (1). Sonication of the cells and membrane extraction were performed as previously described (8). When needed, the bacteria were suspended in 5 mM MgCl<sub>2</sub>-25 mM K<sub>2</sub>HPO<sub>4</sub> (pH 7) and incubated for 10 min at room temperature in the presence of either 30 mM (for membrane proteins) or 10 mM (for total proteins) iodoacetamide prior to sonication or extraction of total proteins. All samples were diluted in sodium dodecyl sulfate (SDS) solubilization buffer in the presence or absence of  $\beta$ -mercaptoethanol and heated for 5 min at 100°C prior to electrophoresis on an SDS-7.5% polyacrylamide gel. After migration, the gel was stained with Coomassie brilliant blue.

**Nucleotide sequence accession number.** The nucleotide sequence reported in this study has been submitted to GenBank under accession number L18881.

## RESULTS

**Nucleotide sequence of the *irp2* gene.** Sequencing of the *irp2* gene revealed a large open reading frame (ORF) of 6,126 nucleotides (nt) (Fig. 1). The *irp2* ORF was followed by the beginning of another ORF in the same reading frame (data not shown). There was no significant ORF in any other reading frame on the same DNA strand or on the complementary strand. The mol% G+C of the *irp2* gene is 59.86, which is higher than the overall value of 47 to 50 mol% G+C for the *Y. enterocolitica* chromosome (4).

A 10-bp inverted repeat sequence (IR1) is present 23 bp downstream of the TAG termination codon (Fig. 1). This sequence may form a stem-loop structure with a calculated  $\Delta G$  (25°C) of 20.8 kcal (1 kcal = 4.184 kJ) (64). The absence of a poly(T) stretch immediately downstream of the stem-loop structure and its relatively low energy suggest that it is not a typical rho-independent transcription terminator (46). In contrast, two perfect inverted repeats of 14 bp (IR2) and 12 bp (IR3), with calculated energies of -36.2 and -29.4 kcal, respectively, are located upstream of the *irp2* gene and could form tight hairpin structures (Fig. 1). IR2 was followed by a stretch of three thymidine residues and may represent a rho-independent transcription termination signal for the ORF upstream of *irp2*.

***irp2* promoter region.** A potential promoter region composed of a -35 (TTGTTA) and a -10 (TATTAT) sequence separated by 19 bp and quite close to the canonical sequence of *E. coli* (27) was identified upstream of the first ATG codon, between nt 141 and 171 (Fig. 1). This region con-



FIG. 1. Nucleotide and amino acid sequences of the *irp2* gene. -10 and -35, promoter sequences; SD, Shine-Dalgarno ribosome-binding site; IR1 to IR3, nucleotide inverted repeats; Fur, Fur-binding sites; a to e and h, locations of termination of the primer extension products; DR1 to DR5, highly conserved stretches of amino acids repeated twice in HMWP2; H1 and H2, hydrophobic domains which may be membrane associated; LZ, possible leucine zipper; SBS, putative SBS sharing homology to the consensus sequence F(F/Y)XLGG(H/D)S(L/I) which has been shown to bind the activated amino acyladenylate either directly or via the 4'-phosphopantetheine cofactor; P-loop, region containing the consensus sequence SGTGXPKG.

tained two overlapping potential Fur boxes; Fur1 (nt 137 to 155) and Fur2 (nt 143 to 161) exhibit the dyad symmetry classically found in the Fur-binding sequences (45). The sequence AGGAGG 4 bp upstream of the first ATG codon (Fig. 1) might serve as a Shine-Dalgarno ribosome-binding site.

To map the transcriptional start point of the *irp2* gene, we performed a primer extension analysis using three different oligonucleotides (P1, P2, and P3) as primers (Fig. 1) and total RNA of *E. coli*(pIR2) grown under iron-rich or iron-poor conditions as the template. No extended cDNA product was found with primers P2 and P3, perhaps because they were too far from the transcriptional start point. With primer P1, extended products were obtained with RNA extracted from iron-starved *E. coli* only (Fig. 2A). This result is in accordance with previous data showing that the *irp2* transcript is synthesized under iron limitation but not iron repletion

conditions in *Yersinia* spp. (12) and that the *irp2* gene is also iron regulated in *E. coli* (9). At least five different extended products (a to e) ending with an adenine residue were obtained (Fig. 2A). A potential promoter sequence was identified in the region from nt 141 to 171, located 17 bp upstream of extended products d and e, suggesting that this sequence may indeed function as an RNA polymerase-binding site in *E. coli* (Fig. 1). The results also suggested the presence of another RNA polymerase-binding site upstream of the a to c transcriptional start points. A potential promoter region composed of a -35 (TTACCA) and a -10 (TAAACC) hexamer, separated by 18 bp, was identified 19 bp upstream of sites a to c (nt 92 to 121) (Fig. 1). Although the degree of identity between this region and the canonical consensus sequence is low, the sequence seemed to be sufficiently conserved to be recognized by the *E. coli* RNA polymerase. Therefore, two potential promoter sequences were identified

1846	GAT ATG ATG CCC GCC AGC CAG CGC GCG ATA CGC GAA CGG GTC AAC	1890	2746	CGG GCC TTC CGG CCA CAA GGA CAA TTT ATC GCG ATG GGC GGC GCC	2790
543	D M M P A S Q R A I R E R V N	557	843	R A F R P Q G Q F I A M G G A	857
1891	GCC ACC GGC GCC CCC ATT CCC GAA GGC TTG CTG CAT GAA GGC ATT	1935	2791	ACC GAG GCG TCT ATC TGG TCT AAC GCC TGC GAA ATT CAC GAC GTC	2835
558	A T G A P I P E G L L H E G I	572	858	T E A S I W S N A C E I H D V	872
1936	TTC CGT ATC GCT CTG CAA CAG CCG CAG GCG CTG GCG GTA ACG GAC	1980	2836	CCT GCC CAC TGG CGT TCC ATC CCT TAC GGT TTT CCG CTA ACC AAC	2880
573	F R I A L Q P Q A L A V T D	587	873	P A H W R S I P Y G F P L T N	887
1981	ATG CGT TAT CAG TGG AAT TAT CAT GAG CTG ACA GAC TAT GCC CGC	2025	2881	CAA CGC TAC CGG GTG GTG GAT GAA CGG GGC CGG GAC TGC CCT GAC	2925
588	M R Y Q W N Y H E L T D Y A R	602	888	Q R Y R V V D E R G R D C P D	902
2026	CGT TGC GCA GGC AGG TTA GTC GAG TGC GGG GTT CAG CCC GGC GAT	2070	2926	TGG GTG TCG GGT GAA TTA TGG ATT GGC GGC ATC GGG GTC GCG GAA	2970
603	R C A G R L V E C G V Q P G D	617	903	W V S G E L W I G G I G V A E	917
2071	AAT GTG GCT ATC ACG ATG TCG AAA GGC GCA GGA CAA CTT GTT GCG	2115	2971	GGC TAT TTC AAC GAT TCC CTG CGC AGC GAG CAG CAA TTT TTG ACG	3015
618	N V A I T M S K G A G Q L V A	622	918	G Y F N D S L R S E G E I E S	932
2116	GTT CTG GCC GTC CTG CTG GCC GGG GCA GTT TAC GTT CCG GTT TCG	2160	3016	CTC CCG GAC GAG CGC TGG TAT CGC ACC GGC GAT CTC GGC TGC TAC	3060
633	V L A V L L A G A V Y V P V S	647	933	L P D E R W Y R T G D L G C Y	947
2161	TTG GAC CAG CCT GCC GCA CGG CGC GAG AAA ATC TAC GCT GAC GCC	2205	3061	TGG CCA GAC GGC ACA ATC GAG TTC CTC GGT CGT CGC GAC AAG CAG	3105
648	L D Q P A A R R E K I Y A D A	662	948	W P D G T I E F L G R R D K Q	962
2206	AGC GTC CGG CTG GTG CTC ATT TGC CAG CAC GAC GCC AGC GCC GGG	2250	3106	GTC AAA GTC GGA GGA TAT CGC ATC GAG CTG GGC GAA ATC GAA AGC	3150
663	S V R L V L I C Y D H I F G V L	677	963	K V G G Y R I E L G E I E S	977
2251	TCA GAC GAT ATT CCC GTC CTT GCC TGG CAG CAG GCC ATT GAG GCG	2295	3151	GCG CTC AGC CAG TTG GCG GGG GTG AAA CAA GCA ACC GTT CTG GCG	3195
678	S D D I P V L A W Q Q A I E A	692	978	A L S Q L A G V K Q A T V L A	992
2296	GAG CCG ATC GTC AAC CCG GTG GTA CGC GCC CCC ACG CAA CCG GCC	2340	3196	ATC GGC GAA AAA GAA AAA ACG CTG GCG GCA TAC GTG GTT CCT CAG	3240
693	E P I V N P V V R A P T Q P A	707	993	I G E K E K T L A A Y V V P Q	1007
2341	TAC ATT ATC TAC ACT TCC GGC TCT ACC GGC ACG CCG AAA GGG GTA	2385	3241	AGC GAG GCT TTT TGC GTT ACC GAT CAT CGG AAC CCG GCA TTG CCG	3285
708	Y I I Y T S G S T G T P K G V	722	1008	S E A F C V T D H R N P A L P	1022
2386	GTC ATT TCT CAC CGG GGA GCG CTC AAC ACC TGT TGC GAT ATC AAT	2430	3286	AAG GCG TGG CAC ACG CTT GCG GGA ACG TTG CCC TGT TGC GCC ATC	3330
723	V I S H R G A L N T C C D I N	737	1023	K A W H T L A G T L P C C A I	1037
2431	ACC CGC TAT CAG GTT GGC CCG CAT GAC AGG GTG CTG GCC CTC TCC	2475	3331	TCG CCA GAG ATC TCC GCA GAA CAG GTA GCC GAT TTC CTT CAG CAT	3375
738	T R Y Q V G P H D R V L A L S	752	1038	S P E I S A E Q V A D F L Q H	1052
2476	GCC CTG CAT TTT GAT TTA TCG GTT TAC GAC ATT TTT GGC GTA CTG	2520	3376	CGC CTG TTA AAA CTG AAG CCG GGT CAC ACC GCT GGC GCC ATC CCT	3420
753	A L H F D L S V Y D I F G V L	767	1053	R L L K L K P G H T A G A D P	1067
2521	CGT GCG GGC GGC GCG CTG GTG ATG GTG ATG GAA AAT CAA CGG CGC	2565	3421	ATC CCC CTG ATG AAT TCA CTC GCT ATC CAG CCG CGC TGG CAG GCC	3465
768	R A G G A L V M V M E N Q R R	782	1068	I P L M N S L A I Q P R W Q A	1082
2566	GAT CCT CAC GCA TGG TGT GAG CTG ATC CAG CGC CAT CAG GTC ACG	2610	3466	GTG GTG GAA CGC TGG TTA GCA TTT CTG GTG ACG CAA CGG CGA CTG	3510
783	D P H A W C E L I Q R H Q V T	797	1083	V V E R W L A F L V T Q R R L	1097
2611	CTC TGG AAC AGC GTC CCG GCG CTG TTC GAT ATG CTG CTG ACC TGG	2655	3511	AAG CCC GCT GCT GAA GGT TAT CAG GTG TGC GCT GGT GAA GAA CGC	3555
798	L W N S V P A L F D I F G L T W	812	1098	K P A E M G Y V C A G E R	1112
2656	TGT GAA GGT TTC GCC GAC GCC ACG CCG GAA AAC CTG CGC GCA GTG	2700	3556	GAG GAT GAG CAC CCG CAC TTC AGC GGA CAT GAT TTA ACG TTA TCG	3600
813	C E G F A D A T P E N L R A V	827	1113	E D E H P H F S G H D L T L S	1127
2701	ATG CTT TCC GGC GAC TGG ATC GGA CTT GAC CTC CCC GCC CGT TAT	2745	3601	CAA ATT CTT CGC GGT GCC CGT AAC GAA CTG TCG TTA CTG AAC GAC	3645
828	M L S G D W I G L D L P A R Y	842	1128	Q I L R G A R N E L S L L N D	1142

FIG. 1—Continued.

upstream of the *irp2* gene; one is close to the consensus sequence of *E. coli* promoters, while the second is somewhat different.

Since transcriptional signals may differ in *E. coli* and *Y. enterocolitica*, and since transcription of the cloned *irp2* gene may differ from that of its chromosomal counterpart, primer extension was performed with primer P1 and with total RNA from iron-replete or iron-starved *Y. enterocolitica* Ye8081. As expected, cDNA bands were present only when the mRNA of iron-starved bacteria was used as the template (Fig. 2B). As in *E. coli*, several transcriptional start points were observed, but only extended products a to c were common to *E. coli* and *Y. enterocolitica*. This could mean that the nonconsensus promoter sequence from nt 92 to 121 acts as an RNA polymerase-binding site in both *Yersinia* spp. and *E. coli* but the region from nt 141 to 171 is not functional in *Yersinia* spp. However, given the potential for hairpin formation in this region, it is also possible that products d and e represent strong pause sites for the RNA polymerase. Additional longer extended products (f to h) were also observed in *Y. enterocolitica* (Fig. 2B). Product h would correspond to a transcriptional start point at position 41, i.e., immediately upstream of the tight IR3 hairpin

structure (Fig. 1), and product g would correspond to a start point approximately 200 bp upstream of nt 1 in Fig. 1. The end of extended product f was too distant to be correctly located. Whether the f to h transcriptional start points far upstream the *irp2* coding sequence are used to generate a long messenger overlapping *irp2* is not clear. Alternatively but less likely, these extended products may result from the hybridization of P1 to other iron-dependent *Yersinia* mRNAs sharing sequence homologies to P1.

**HMWP2 amino acid sequence.** Assuming that the first ATG codon following the potential Shine-Dalgarno sequence is the translational initiation codon, HMWP2 is a protein of 2,035 amino acids with a calculated molecular mass of 228,566 Da, somewhat higher than the molecular mass of 190,000 Da previously estimated by SDS-polyacrylamide gel electrophoresis (PAGE). Identification of the translation initiation codon could not be performed because the NH<sub>2</sub>-terminal amino acid of the protein was blocked. The deduced isoelectric point of HMWP2 is 5.81, but the carboxy terminus of HMWP2 is very rich in arginine residues (7 of the last 12 amino acids [aa]).

Upon bacterial sonication and ultracentrifugation, the HMWPs are found predominantly in the pellet (membrane)

3646	GCG CAG TGG TCG CCG GAA AGC CTG GCC TTT AAC CAT CCG GCC AGC	3690	4546	ACC CGT CTG ACC GGG CAA CTT CAT CAG GCA GGT TAT GAA GCG CAA	4590
1143	A Q W S P E S L A F N H P A S	1157	1443	T R L T G Q L H Q A G Y E A Q	1457
3691	GCC CCG TAT ATT CAG GAA CTG GCG ACA ATT TGC CAA CAG CTT GCA	3735	4591	TTA AGC GAC CTG TTT AAT CAT CCC CGG CTG GCG GAT TTT GCC GCC	4635
1158	A P Y I Q E L A T I C Q Q L A	1172	1458	L S D L F N H P R L A D F A A	1472
3736	CAG CGC TTA CAG CGC CCG GTA CGC CTG CTT GAG GTG GGA ACC CGC	3780	4636	ACG CTG CGT AAA ATC GAC GTC CCG GTC GAA CAA CCA TTC GTC CAC	4680
1173	Q R L Q R P V R L L E V G T R	1187	1473	T L R K I D V P V E Q P F V H	1487
3781	ACC GGC CGC GCC GCA GAA TCG CTG TTG GCA CAG CTC AAC GCC GGA	3825	4681	TCT CCT GAA GAA CGC TAC CAG CCC TTT GCG CTT ACC GAC GTG CAG	4725
1188	T G R A A E S L L A Q L N A G	1202	1488	S P E E R Y Q P F A L T D V Q	1502
3826	CAG ATT GAG TAT GTC GGG CTT GAG CAG AGC CAG GAG ATG CTA CTG	3870	4726	CAG GCT TAC CTG GTG GGG CGT CAG CCG GGC TTT ACC CTG GGC GGC	4770
1203	Q I E Y V G L E Q S Q E M L L	1217	1503	Q A Y L V G R Q P G F T L G G	1517
3871	AGC GCC CGG CAG AGG CTC GCC TCC TGG CCT GGT GCC CGT CTG TCC	3915	4771	GTC GGC TCA CAT TTC TTT GTT GAA TTT GAA ATT GCC GAT CTG GAC	4815
1218	S A R Q R L A S W P R A R L S	1232	1518	V G S H F F V E I A D L D	1532
3916	CCC TGG AAT GCA GAC ACG CTG GCG GCG CAC GCT CAC TCG GGG GAC	3960	4816	CTC ACC CGG CTG GAG ACG GTC TGG AAC CGA TTA ATC GCC CGC CAC	4860
1233	P W N A D T L A A H A H S G D	1247	1533	L T R L E T V W N R L I A R H	1547
3961	ATT ATC TGG CTT AAT AAC GCC CTG CAT CGT CTG CTG CCG GAA GAT	4005	4861	GAT ATG CTA CGC GCC GTC GTG CTT GAT GGA CAG CAA CAG GTG CTC	4905
1248	I I W L N N A L H R L L P E D	1262	1548	D M L R A V V L D G Q Q Q V L	1562
4006	CCC GGG CTC CTT GCG ACA TTA CAA CAG CTT GCC GTT CCC GGC GCG	4050	4906	GAA CAG ACG CCC CCC TGG GTG ATA CCC ACA CAC ACC CTT CAT ACG	4950
1263	P G L L A T L Q L A V P G A	1277	1563	E Q R P W V I P T H T L H T	1577
4051	CTG CTC TAC GTG ATG GAG TTT GCG CAG TTA ACG CCG TCC GCC CTG	4095	4951	CCT GAA GAG GCG TTG CCG GTA CGC GAA AAA CTG GCG CAT CAG GTA	4995
1278	L L Y V M E F R Q L T P S A L	1292	1578	P E E A L R V R E K L A H Q V	1592
4096	CTC AGC ACG CTC CTG TTA ACC AAT GGG CAG CCG GAG GCC TTG CTG	4140	4996	CTC AAC CCC GAA GTG TGG CCG GTA TTC GAT CTC CAG GTC GGA TAC	5040
1293	L S T L L L T N G Q P E A L L	1307	1593	L N P E V W P V F D L Q V G Y	1607
4141	CAT AAC AGC GCC GAC TGG GCG GCA TTA TTT AGC GCG GCC GCC TTC	4185	5041	GTG GAC GGG ATG CCC GCC CGC CTG TGG CTG TGT CTG GAT AAC CTG	5085
1308	H N S A D W A A L F S A A A F	1322	1608	V D G M P A R L W L C L D N L	1622
4186	AAC TGT CAG CAT AGC GAT GAG GTC GCG GGG TTA CAA CGC TTC CTC	4230	5086	TTG CTT GAC GGC CTG AGC ATG CAG ATC CTG CTG GCG GAG CTG GAG	5130
1323	N C Q H S D E V A G L Q R F L	1337	1623	L L D G L S M Q I L L A E L E	1637
4231	GTA CAA TGT CCT GAC AGG CAG GTG CGC CGC GAT CCC CGT CAA CTT	4275	5131	CAC GGC TAC CGC TAC CCG CAA CAG CTG CTT CCG CCG CTG CCC GTC	5175
1338	V Q C P D R Q V R R D P R Q L	1352	1638	H G Y R Y P Q Q L L P P L P V	1652
4276	CAG GCC GCC CTC GCC GGG CGT CTG CCG GGG TGG ATG GTG CCG CAA	4320	5176	ACC TTC AGG GAT TAT CTG CAA CAG CCC TCG CTA CAG TCG CCC AAT	5220
1353	Q A A L A G R L P G W M V P Q	1367	1653	T F R D Y L Q Q P S L Q S P N	1667
4321	CGG ATC GTC TTC CTC GAC GCC TTA CCG CTG ACG GCT AAC GGG AAA	4365	5221	CCA GAT TCT CTG GCA TGG TGG CAG GCG CAG CTT GAT GAT ATT CCT	5265
1368	R I V F L D A L P L T A N G K	1382	1668	P D S L A W W Q A Q L D D I P	1682
4366	ATT GAC TAC CAG GCG CTG AAG CGT CGT CAT ACC CCT AAA GCG GAA	4410	5266	CCG GCG CCA GCG TTG CCG CTG CGC TGC TTG CCT CAG GAG GTT GAA	5310
1383	I D Y Q A L K R R H T P K A E	1397	1683	P A P A L P L R C L P Q E V E	1697
4411	AAC CAG GCC GAA GCG GAT TTA CCC CAG GGC GAC ATT GAA AAA CAG	4455	5311	ACA CCG CGC TTC GCC CGC CTG AAC GGC GCG CTG GAC AGC CGC	5355
1398	N Q A E A D L P Q G G D I E K Q	1412	1698	T P R F D R L N G A R L D S T R	1712
4456	GTT GCC GCC CTC TGG CAG CAA CTC TTA TCG ACT GGC AAT GTC ACC	4500	5356	TGG CAT CGG CTG AAA AAA CGG GCG GCT GAC GCC CAT CTC ACC CCG	5400
1413	V A A L W Q Q L L S T G N V T	1427	1713	W H R L K K R A A D A H L T P	1727
4501	AGA GAA ACC GAC TTC TTC CAG CAA GGC GGC GAT AGC CTG CTG GCG	4545	5401	TCG GCC GTG CTG TTG TCG GTG TGG TCA ACG GTT CTC TCT GCA TGG	5445
1428	R E T D F F Q Q G G D S L L A	1442	1728	S A V L L S V W S T V L S A W	1742

FIG. 1—Continued.

fraction (8). Analysis of the predicted NH<sub>2</sub>-terminal amino acid sequence of HMWP2 did not reveal a potential signal sequence, suggesting either that the protein is not exported or that another export signal is used. The computer-derived hydrophobicity profile of HMWP2 (38) displayed a succession of short hydrophobic and hydrophilic fragments (data not shown). Two relatively long hydrophobic regions, H1 (24 aa, located at positions 630 to 653) and H2 (22 aa; positions 1725 to 1746) were found and might correspond to membrane-anchored domains (Fig. 1).

One striking feature of the HMWP2 sequence was the presence of two approximately 550-aa direct repeats (DR) at each extremity of the protein (positions 3 to 547 and 1395 to 1919). These repeats are aligned in Fig. 3. They share 35.6% identical residues and 54.6% similar amino acids. Within these regions, five highly conserved sequences are found: DR1 (23 of 33 aa identical), DR2 (9 of 12), DR3 (11 of 13), DR4 (8 of 9), and DR5 (13 of 15) (Fig. 1 and 3).

**Disulfide bridges.** The deduced amino acid sequence of HMWP2 includes 30 cysteine residues. When the HMWPs from sonicated cells are heated to 100°C in SDS buffer in the absence of a reducing agent, they do not enter the polyacryl-

amide gel, suggesting the existence of intermolecular disulfide bridges (8). However, the possibility that the disulfide bonds observed were artificially formed upon sonication, when the proteins were not in their natural environment, could not be excluded. To clarify this point, iron-starved cells of strain Ye8081 were washed several times in saline, then resuspended in SDS dissociation buffer with or without β-mercaptoethanol, and heated to 100°C. The two samples were electrophoresed on a polyacrylamide gel alongside the corresponding membrane fraction obtained after sonication, centrifugation, and dissociation in SDS buffer with or without β-mercaptoethanol. When the membrane fraction was subjected to SDS-PAGE, the HMWPs in the membrane fraction entered the gel only when heated with β-mercaptoethanol (Fig. 4, lanes 3), confirming the results of previous experiments (8). In contrast, the absence of a reducing agent did not alter the migration of the two polypeptides present in the total bacterial extract (Fig. 4, lanes 1). These results could mean that the disulfide bridges formed only after sonication. To test this hypothesis, the cells were incubated with iodoacetamide prior to sonication or extraction of total proteins. Iodoacetamide penetrates the bacteria and alky-

5446	AGT GCG CAG OCT GAG TTC ACG CTT AAC CTT ACG CTT TTC GAC AGG	5490
1743	<u>S A Q P</u> E F T L N L T L F D R	1757
	DR3	
5491	CGA CCG CTG CAC CCG CAA ATC AAC CAG ATT CTG GGC GAT TTC ACT	5535
1758	<u>R P L H P</u> Q I N Q I L G D F T	1772
5536	TCG CTG ATG TTG CTG AGC TGG CAC CCC GGC GAA AGC TGG CTG CAC	5580
1773	S L M L L S W H P G E S W L H	1787
5581	AGC GCG CAG TCA CTA CAG CAG CGG CTG AGC CAG AAC CTC AAC CAC	5625
1788	S A Q S L Q Q R L S Q N L N H	1802
5626	CGC GAT GTG TCA GCC ATC CGC GTG ATG CGT CAA CTG GCG CAA CGG	5670
1803	R D V S A I R V M R Q L A Q R	1817
5671	CAA AAC GTG CCT GCC GTT CCG ATG CCC GTC GTC TTT ACC AGC GCG	5715
1818	A N V P A V P M P V V F T S A	1832
	DR4	
5716	CTG GGC TTT GAG CAG GAT AAC TTC CTC GCC CGG CGT AAT CTG CTC	5760
1833	<u>L G F E Q D N F L A R R N L L</u>	1847
5761	AAG CCG GTC TGG GGC ATC TCC CAG ACG CCG CAG GTC TGG CTC GAT	5805
1848	<u>K P V W G I S Q T P Q V W L D</u>	1862
	DR5	
5806	CAC CAG GTT TAT GAA TCC GAA GGC GAA CTG CGC TTT AAC TGG GAT	5850
1863	<u>H Q V Y E S E G E L R F N W D</u>	1877
5851	TTT GTC GCC GCG CTG TTT CCT GCC GGG CAG GTG GAG CGC CAG TTT	5895
1878	F V A A L F P A G Q V E R Q F	1892
5896	GAA CAG TAT TGC GCA TTG CTA AAC CGA ATG GCC GAG GAT GAA AGC	5940
1893	E Q Y C A L L N R M A E D E S	1907
5941	AGC TGG CAA CTG CCG CTC GCC GCG CTG GTG CCT CCC GTA AAA CAC	5985
1908	S W Q L P L A A L V P P V K H	1922
5986	GCA GGG CAA TGC GCA GAG CGC CCA CCG CGC GTA TGC CCT GAG CAC	6030
1923	A G Q C A E R P P R V C P E H	1937
6031	TCT CAG CCA CAC ATT GCG GCG GAC GAG AGC ACC GTC AGC CTG ATT	6075
1938	S Q P H I A A D E S T V S L I	1952
6076	TGC GAC GCC TTC CGC GAG GTG GTT GGC GAG TCT GTC ACA CCC GCA	6120
1953	C D A F R E V V G E S V T P A	1967
6121	GAA AAC TTC TTT GAG GCG GGT GCA ACA TCG CTG AAT CTG GTG CAA	6165
1968	E N F F E A G A T S L N L V Q	1982
6166	CTG CAC GTT TTG TTA CAA CGT CAC GAA TTT TCC ACC CTG ACG TTG	6210
1983	L H V L L Q R H E F S T L T L	1997
6211	CTT GAC CTC TTC ACC CAC CCT TCT CCT GTT GCC CTG GCC GAT TAT	6255
1998	L D L L F T H P S P V A L A D Y	2012
6256	CTG GCC GGC GTC GCC ACG GTG GAG AAA ACA AAA CGT CCT CGC CCT	6300
2013	L A G V A T V E K T K R P R P	2027
6301	GTT CGC CGT CGT CAG CGG CGG ATA TAG CGCGAAGCAAACGTATTTCCCGC	6351
2028	V R R R Q R R I *	2035
6352	<u>GAA CGCCATCGGGAACGCATGGCGTTCCATTGACTTTTGTAATCTTAGGAAACGGGACCG</u>	6411
	IRI	
6412	ATTATGGATAACTTGCCTTCTCTCTGCGCGACAGCAGATTCCATTGATGCATCG	6468

FIG. 1—Continued.

lates free cysteine sulfhydryl groups, thus preventing the subsequent formation of artificial disulfide bridges between the free SH radicals upon sonication. Incubation of the bacteria with iodoacetamide prior to sonication permitted the HMWPs to migrate into the SDS-polyacrylamide gel, even in the absence of a reducing agent in the solubilization buffer (Fig. 4, lanes 4). The fact that migration of the HMWPs in membrane extracts of iodoacetamide-treated bacteria or in total protein extracts was not affected when β-mercaptoethanol was omitted from the dissociation buffer suggests that despite their high number, the cysteine sulfhydryl groups are not disulfide bonded in situ.

**Homology of HMWP2 to AngR from *V. anguillarum*.** To determine whether HMWP2 had significant homology with known protein sequences, a search in the Swiss-Prot database was performed with the FastA program (44). The highest percentage of similarity or identity was found with the protein AngR of *V. anguillarum* (24). AngR is a 118-kDa protein which has been reported to activate the iron uptake system of *V. anguillarum* (50). When the overall amino acid sequences of HMWP2 and AngR were compared, a stretch

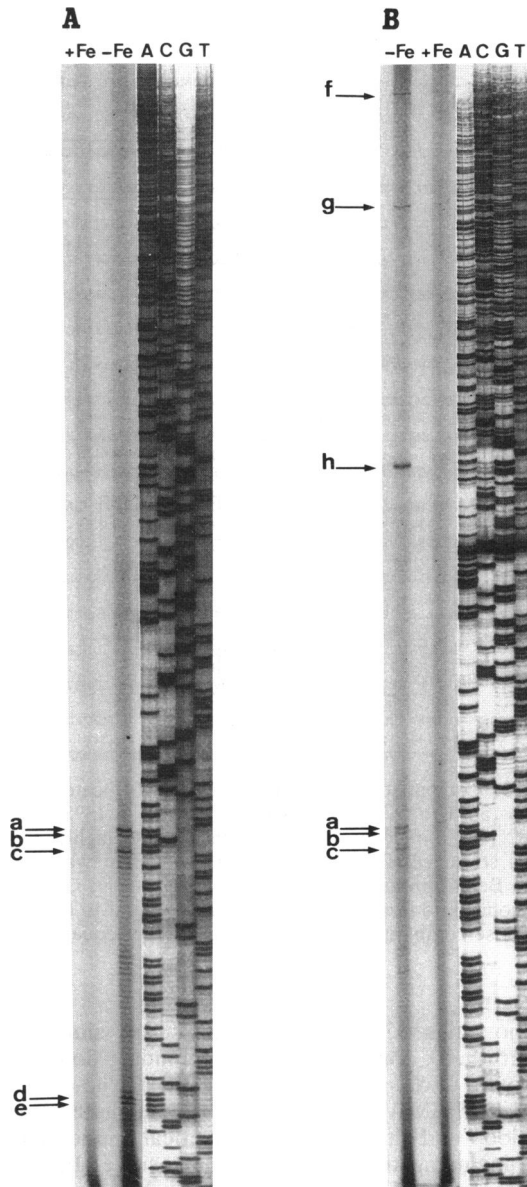


FIG. 2. Primer extension analysis performed with oligonucleotide P1 as the primer and total RNAs isolated from iron-replete (+Fe) and iron-starved (-Fe) *E. coli*(pIR2) (A) and *Y. enterocolitica* Ye8081 (B) as templates. Lanes A, C, G, and T correspond to the labeled nucleotides. a to h are the products of the primer extension experiments.

of 337 aa (1011 to 1347) in HMWP2 was found to be completely absent from AngR. In addition, the 100 amino-proximal and the 550 carboxy-proximal amino acids of HMWP2 are also absent from AngR (see Fig. 6). Otherwise, the complete sequence of AngR could be aligned with a contiguous segment of HMWP2, and the two proteins exhibited a high percentage of similarity or identity (44.5% identical residues and 61.7% identical plus similar amino acids on a length of 1,071 aa; Fig. 5). Some segments of HMWP2 and AngR are more similar than others: the central regions of the two proteins (residues 594 to 1010 in HMWP2) exhibit over 66% identity, while the amino and carboxy flanking regions exhibit only 28 and 34% identity, respectively.

NH2	3	SGAPSKDLSLLPDRNRHAADYQQLRRLRIQELNLTPOQLHDESNIQAGLDS	52
COOH	1395	KAENQAQAEADLPQGDIEKQVAALWQQLLSTGNVTR.....ETDFFQOQGDSS	1439
	53	IRLMRWLHWFRKNGYRLTLRELYAAPTAAANQMLSRSPENAEETLPD	102
	1440	LLATRLTGQLHQAGYEAQLSDLFNHRPLADFAATL.....RKIDVFEV	1481
	103	ESSWPNMTES...TPFFLTPVQHAYLTGRMPGQTLGGVGGCHLYQEFESHCL	150
	1482	EQPFVHSPPEERYQPFALTDVQQAYLVGRQPGFTLGGVGGSHFFVEFEIADL	1531
	151	TASQLEQAITTLQRHPMLHIAFRPDGQVWLPQPYWNGVTVDLDRHDA	200
	1532	DLTRLETVMNRLIARHMDLRAVLDGQQVLEQTPPW.VIPHTLHTPEE	1580
	201	ESRQAYLDALRQLRSHRLLRVEIGETFDQTLTLLPDNRHRLHVNIDLLIM	250
	1581	ALR.....VREKLAHQVLPNPEVWVFDLQVGYVDGMPARLWCLDNLLL	1624
	251	DASSFTLFFDELNALLAGESLSAIDTRYDFRSYLLHQKINQPLRDDARA	300
	1625	DGLSMQIILLAELEHGRYFPQQLLPPLEVTFRDYL..QQPSLQSPNPDLSA	1672
	301	YMLAKASTLFPAPVLPVLCHEATLREVRNTRRRMIVPATRWAFSNRAGE	350
	1673	NWQAQLDDIPFAPALPLRCLLQEVETPRFARLNGALDSTRWHRLKRAAD	1722
	351	YGVTPMALATCFSAVLAIRWGGTLRLINLITLFDROP LHPAVGAMLADEF	400
	1723	AHLTPSAVLLSVSVSVLSAMSQAQPEFLNLTFLDRRLHPDINQLGDET	1772
	401	NILLDTACDGDVTSNLARKNQLTFTEDMEHRHWSGVELLRELKQRQYP	450
	1773	SMLLSW.HPGESWLHSAQSLQRLSQNLRHVDVAIRVMRQLAQRQNPV	1821
	451	HGA.PVVFTSMNLRSLYSRAESPLGPEWGISQTPQVWIDHLAFEHGGE	499
	1822	AVPMPEVFTSALGFEQDNFLARRNLLKPVWGISQTPQVWLDHGVYESEGE	1871
	500	VWLQWDSNDALFPFALVETLFDAYCQLINQLCDESAWQKPFADMMPA	547
	1872	LRFNWDFVAALFPAGQVERQFEQYCALLNRMAEDESSWQLFLAALVFP	1919

FIG. 3. Comparison of the amino acid repeats located at the NH<sub>2</sub> terminus (aa 3 to 547) and COOH terminus (aa 1395 to 1919) of HMWP2. The boxed sequences represent five highly conserved repeated stretches (DR1 to DR5). The alignment was performed with the Bestfit program. |, identical amino acids; :, amino acids whose comparison value is greater than or equal to 0.5; ., amino acids whose comparison value is greater than or equal to 0.1.

Amino acids 873 to 894 of AngR are similar to the helix-turn-helix domain of the DNA-binding domain of the Cro protein of *Salmonella* phage P22 (24). The corresponding region in HMWP2 (984 to 1005) is immediately upstream of the loop and shares 12 of 22 identical amino acids with AngR, including the central glycine residue (Fig. 5). The helix-turn-helix motif of HMWP2 has a basic calculated pI (9.37), as does the corresponding region in AngR and DNA-binding domains in prokaryotic regulatory proteins. However, the predicted score of 31 determined by the method of Dodd and Egan (21) for this segment of HMWP2 is much lower than that for helix-turn-helix motifs known to be involved in DNA binding.

**Homology between HMWP2 and a group of antibiotic synthetases.** A group of proteins involved in the nonribosomal biosynthesis of antibiotics also displayed significant homology to HMWP2. This group includes tyrocidine synthetase 1 (TY1) (69), gramicidin S synthetase 1 (GS1) (31, 37) and GS2 (32, 67) from *Bacillus brevis*, and the  $\delta$ -(L- $\alpha$ -aminoadipyl)-L-cysteinyl-D-valine synthetases (ACVSs) from *Penicillium chrysogenum* (20, 58), *Aspergillus nidulans* (41), *Cephalosporium acremonium* (26), and *Nocardia lactamdurans* (15). All of these proteins belong to the same superfamily of adenylate-forming enzymes (67).

The entire sequences of TY1 (69) and GS1 (37) were similar to the segment of HMWP2 from aa 500 through the 5' carboxy terminus, except for the 337-aa loop (Fig. 6). The entire sequence of GS2 (67) is composed of four repeated

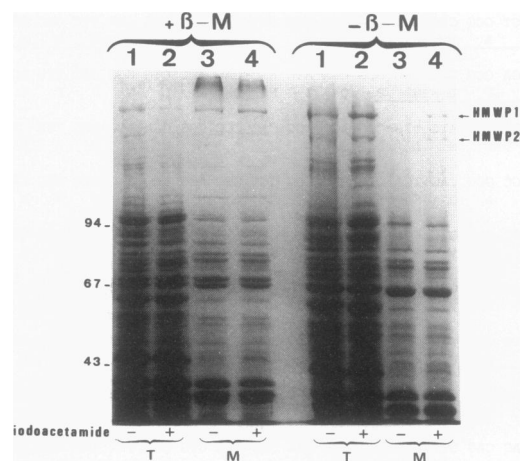


FIG. 4. SDS-PAGE of total proteins (T) and membrane proteins (M) from iron-starved *Y. enterocolitica* Ye8081 obtained after sonication and centrifugation. The cultures were separated in two parts; one part did not receive any treatment (-), and the other was incubated with iodoacetamide (+) prior to sonication or extraction of total proteins. Before electrophoresis, proteins were boiled in Laemmli solubilization buffer in the presence (+  $\beta$ -M) or absence (-  $\beta$ -M) of  $\beta$ -mercaptoethanol. Standard molecular weights (in thousands) are shown on the left.

domains of about 600 aa, separated by nonhomologous sequences of approximately 500 aa. The homology between HMWP2 (without the 337-aa loop) and GS2 extends throughout the HMWP2 sequence (Fig. 6). However, when each of the four repeated domains of GS2 (dA to dD) was aligned separately with HMWP2, the highest matches were found in each case with the central part of HMWP2 (approximately between residues 540 and 1470), and the percentage of amino acid identity was of the same order of magnitude (30.1, 33, 33.9, and 34.7% for dA to dD, respectively; Fig. 6). Strikingly, this region corresponds mainly to the sequence between the two repeats in HMWP2, the region with the highest degree of identity to AngR (Fig. 6).

The ACVSs are very large enzymes (>400 kDa) involved in the biosynthesis of  $\beta$ -lactam antibiotics. All of the sequenced ACVSs contain three functional repeated domains of approximately 570 aa exhibiting significant homology to TY1, GS1, and GS2. The homology with the different ACVSs was found throughout the entire length of HMWP2 (Fig. 6). However, when the three repeats (dA to dC) from the ACVS sequences were separately aligned with the HMWP2 sequence, they, like the four GS2 repeated domains, matched the central part of HMWP2 (Fig. 6).

The consensus sequence SGTGXPKG, which is highly conserved in this group of antibiotic synthetases, is present in each repeat of GS2 and the ACVSs (32). It has been suggested that this conserved stretch of amino acids may be a new class of phosphate-binding loop (P loop) (32). This sequence is also present in different enzymes involved in the catalysis of ATP-P<sub>i</sub> exchange. A highly similar sequence (SGSTGTPKG) was also found in HMWP2 (positions 713 to 721) and in AngR (positions 602 to 610), in the same position relative to the aligned sequences (Fig. 1, 5, and 6).

Recently, it was shown that the sites for covalent substrate amino acid binding (SBS) in each repeat of GS2 (52) share the same consensus sequence: F(F,Y)XLGG(H,D)S(L,I). The serine residue present in this motif has been shown to bind the substrate, which is either the activated

HMWP2 102	DESSWFNMTESTFPFLTPVQAYLHTRMFGQTLGGVGHLYQEFEGHCLT	151
AngR 2	NQNEHFFAPPETKLPFTSNQWQLSTQRQREKKSITNFYQEFDFYENIS	51
152	ASQLEQAITLLQRHFMHLIAFRPDQQVWLPQPYWNGVTVHDLRHRDAE	201
52	RDTLERCLTTIKHHEIFGAKLSDDFYLHFPKTHIETFAVNDLS . . . .	96
202	SRQAYLDALRQLRSHRLLRVEIGET . . . . . FDFQLTLLPDRHRLHVN	244
97	. . . . . NALKQIDKQLADTRSAVTKRSQAIISIMF5SILPKNIIRLHVR	140
245	IDLLIMDASSFTLFFDELNALLAGESLSAIDTRYDFRSYLLHQKINQPL	294
141	FNSVVVDNPSVTLFFEQTLQLLSSGSLFSLNQEQTISAY . . . . . NHKVNNE	187
295	. . . . . RDDARAYWLAKASTLPPAPVLPVCEPATLREVRNTRRRMIVPATRWH	342
188	LSVDLESARWNEYLLWLPSSANLFTICEPEKLDETDITRRCITLSQRKWQ	237
343	AFSNRAGEYGVPTMALATCFSAVLARWGLTRLLNITLFRDQPLHPAV	392
238	QLVTVSKHNVTPEITLASIFSTVLSWGLHMQYKLMRFDTITKINDY . . . .	283
393	GAMLADFTNILLDTACDGDVSNLARKNQLTFDEWEH . . . . . RHWSGVE	438
284	TGIIGQFTEPLLVMGSGPEQSFSLVKNKQKFEAYHYDVKVPVFCQVN	333
439	LLRELKROQYHPGAVVFTSNLGRSLYSSRAESPLGPEWGISQTPQVW	488
334	KLSNISDSHRYP . . . . . ANITFSSELLNTNHSKKA . . . . . VMGCRQSANTW	374
489	IDHLAFEHGVEVWQWSDNALFPALVETLFDAYCQLINQLCDDDESAWQ	538
375	LSLHAVIEQEQLVLQWSDQAIFFPKDMIKDMLHSYTDLLDLSQKDVNWA	424
539	KPFADMPASQAIRERVRNATGA . PIPEGLLHEGIFRIALQQPQALAVTD	587
425	QPLP TLLPKHQESIRKNINQQDLELTKELLHQRFKKNVESTPNALAIH	474
588	MRYQWNYHELTDYARRCAGRLVECGVQPGDNVAITMSKGAQGLVAVLAVL	637
475	GQESLDYITLASYAKSCAGALTEAGVKSQDRVAVTMNKGIGQIVAVLGLI	524
638	LAGAVYVPSLDQPAARREKIYADASVRLVICQHDA . . . . . SAGSDDIPVLA	685
525	YAGAIYVPSLDQPAARREKIYADASVRLVICQHDA . . . . . SAGSDDIPVLA	574
686	WQQAIEAEP IVNVPVVRAPTPAYIIYTSGSTGTPKGVVISHRGALNTCCD	735
575	WQTAIKSEPMRSPQDVAFSQPAYIIYTSGSTGTPKGVVISHQALNTCIA	624
	P-loop	
736	INTRYQVGHDRVRLVLSALHFDLSVYDIFGVLRRAGGALVMVMENQRDPH	785
625	INRRYQIGKNDRLVLSALHFDLSVYDIFGVLRRAGGALVMVMENQRDPH	674
786	AWCELIQRHQVTLWNSVPAALFDMLLTWCEGFADATPENLRVAVMLSGDWIG	835
675	AWCQAIIEHNVTMNSVPAALFDMLLTWCEGFADATPENLRVAVMLSGDWIG	724
836	LDLPARYARFRPQQQFIAMGGATEASIWSNACEIHDVPAHRSIPYGFPL	885
725	LDLPQRYRNYRVDGQFIAMGGATEASIWSNACEIHDVPAHRSIPYGFPL	774
886	TNQRVYVDERGRDCPDWVSGELWIGGIVAGYFNDLSRSEQQLTLFDP	935
775	PRQQYRVVDDLGRDCPDWVSGELWIGGIVAGYFNDLSRSEQQLTLFDP	824
936	ERWYRTGLGCVYWDGTEIFLGRDRKQKVKGVYRIELGEIESALSQLAGV	985
825	HAWYRTGDMGCVYWDGTEIFLGRDRKQKVKGVYRIELGEIEVALNNIPGV	874
	1010 (loop) 1348	
986	KQATVLAIGEKEKTLAAYVVPQSEA . . . . . DPRQLQAALAGRLPGWM	1364
875	QRAVALAVGNKDKTLAAFIWMDSEQAFIVTAPLDAEBEVQLLNKQLPNYM	924
	helix-turn-helix	
1365	VPQRIVFDALPLTANGKIDYQALKRR . . . . . HTPKAENQAEADLPQGDIEKQV	1413
925	VPKRIIFLETFFLTANGKVDHKAALTRMTNREKKTQSINKPIITASEDRV	974
	SBS	
1414	AALWQQLLSTGNVTRTDFQQGGDSLALTRLTGQLHQAGYEAQLSDLFN	1463
975	AKIWNVDLGPTELKYSDDFFLGGDAYNAIEVVKRCKAGYLKLSMLYR	1024
1465	HPLADFAATLRKIDVPEVQP 1484	
1025	YSTIEAFAIMDRCLAPQEE 1045	

FIG. 5. Homology between HMWP2 and AngR performed with the Bestfit program (see the legend to Fig. 3). A loop of 337 aa (1011 to 1347) was withdrawn from the HMWP2 sequence in order to obtain the optimal alignment.

amino acid or more probably the cofactor 4'-phosphopantetheine, which bears a sulfhydryl group that would be involved in thioester formation with the amino acid substrates and in peptidyl transfer (52). This consensus sequence, which is present in GS1, in TY1, and in the three repeated domains of the different ACVSs, was also found at positions 1432 to 1440 in the second direct repeat (DR) of HMWP2 and was well conserved (FFQQGGDSL) (Fig. 1 and 6). A related sequence, LIQAGLDSI, was identified in the first direct repeat at positions 45 to 53 (Fig. 3). While less similar to the consensus sequence, it includes the critical serine residue at the appropriate position. Curiously, the related sequence FFLSGGDAY is present at positions 993 to 1001 in AngR but does not contain the primordial serine residue at position 8 in the consensus sequence (Fig. 5).

**Homology of HMWP2 with EntF from *E. coli*.** In addition to the group of antibiotic synthetases, HMWP2 shares significant homology with another member of the superfamily of adenylate-forming enzymes: the enterobactin synthetase component F (EntF) of *E. coli* (47). EntF is a 142-kDa protein involved in biosynthesis of the siderophore enterobactin. Like the antibiotic synthetases, EntF is involved in the synthesis of peptides through protein template mechanisms and shares sequence homologies with TY1 and GS1 (47). The entire EntF sequence can be aligned with aa 112 to 1803 of HMWP2 (except for the 337-aa loop section) (Fig. 6). The putative binding site for 4'-phosphopantetheine (FFALGGHSL) in EntF (positions 999 to 1007) (47) matches with the corresponding sites in HMWP2 and the different synthetases (Fig. 6). Similarly, the potential P-loop sequence of EntF (SGSTGRPKG) located at positions 604 to 612 aligns with the other P-loop sequences (Fig. 6).

**Other homologies and multiple alignment.** As expected, HMWP2 also displayed homology with other members of the superfamily of adenylate-forming enzymes, including the amino adipate-semialdehyde dehydrogenase (Lys2) of *Saccharomyces cerevisiae* (43), the acetyl coenzyme A (acetyl-CoA) ligases of *A. nidulans* (14), *Neurospora crassa* (14), and *Methanoxthrix soehngenii* (23), the 4-coumarate-CoA ligases of *Petroselinum crispum* (40) and *Oryza sativa* (71), the *O*-succinylbenzoic acid-CoA ligase (MenE) of *Bacillus subtilis* (22), the luciferase of *Photinus pyralis* (19), and EntE of *E. coli* (59). Multiple alignment of the amino acid sequences of these enzymes, of HMWP2, AngR, TY1, GS1, and EntF, of each repeated domain of GS2, and of ACVS was performed with the Pileup program, using a gap weight of 3. Two main groups of proteins (A and B) were identified. Family A included all of the antibiotic synthetases, EntF, Lys2, AngR, and HMWP2, while family B was composed of luciferase, EntE, and the various ligases (Fig. 7). Some stretches of amino acids were highly conserved among the 30 sequences studied, while others were specific for family A (Fig. 7). Among the most conserved regions, we identified at least six consensus sequences (CS I to VI) defining the whole superfamily (Fig. 8). CS I corresponded to the new class of P loop which, by our criteria (see the legend to Fig. 7), differed slightly from that previously described. CS II to VI were absent from MenE because the protein sequence is short and ends before CS II. The TGD motif found within CS III is present in a wide variety of ATPases, in which it may function as a nucleotide-binding fold (63). Although the exact function of most of these amino acid stretches is still unknown, their high degree of conservation among the entire superfamily suggests that their presence is of prime importance. Within the superfamily, four other CSs (CS VII to X) are clearly specific for family A (Fig. 8). CS X, the known



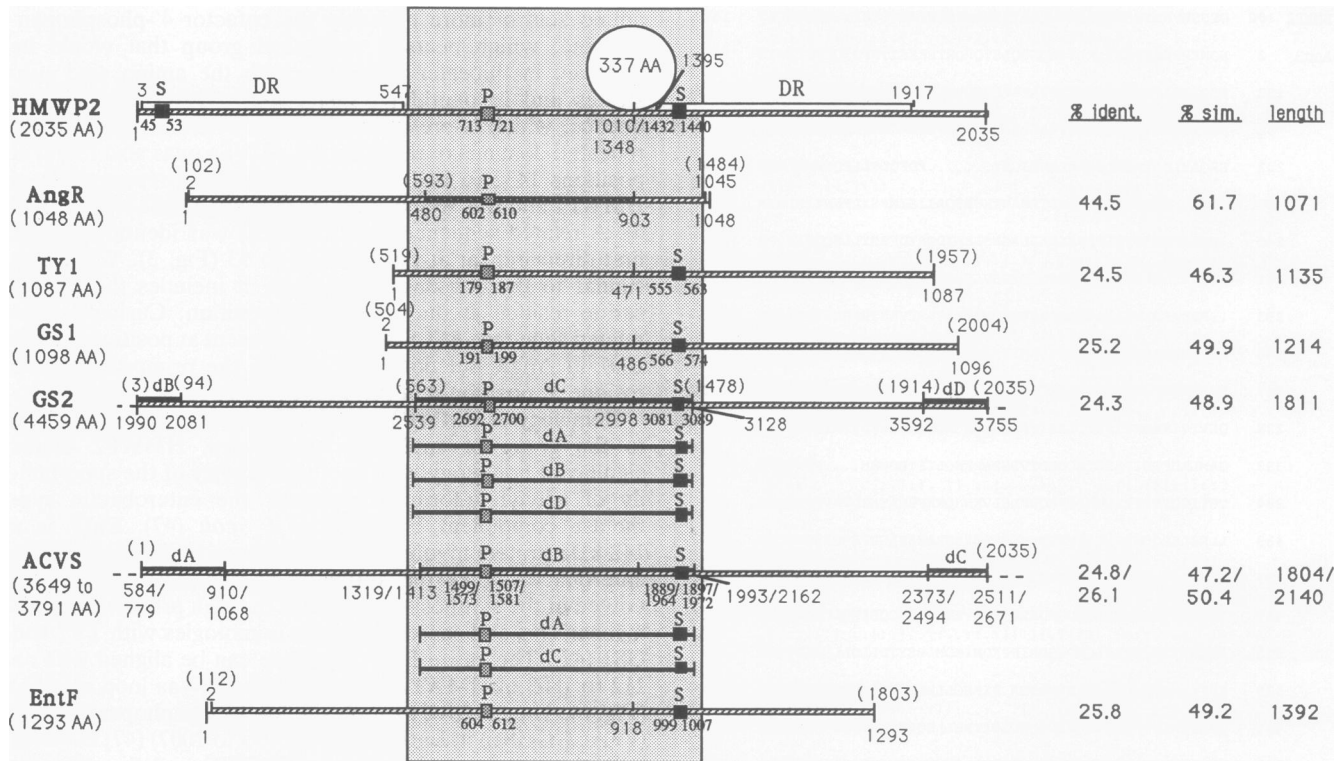


FIG. 6. Schematic alignment of HMWP2, AngR, TY1, GS1, GS2, EntF, ACVS, and repeated domains of ACVS, based on results of analysis with the Bestfit program. The representation of ACVS is a compilation of four ACVS sequences. DR, stretch of amino acids repeated twice in HMWP2; S, putative SBS; P, P loop. Numbers in parentheses above the lines indicate the corresponding positions in HMWP2. The circle corresponds to the loop of 337 aa that is completely absent from the other proteins. dA, dB, dC, and dD are the amino acid repeats found in GS2 or ACVS. The darker region in the representation of AngR corresponds to the region of greatest homology to HMWP2 (66.2% identical amino acids and 77.9% similar plus identical amino acids on a length of 419 residues). % ident., percentage of identical amino acids; % sim., percentage of identical plus similar amino acids. The shaded box delimits the region which was common to all proteins.

SBS, is absent from Lys2, which ended upstream of it. Interestingly, CS VII is highly conserved in all group A proteins except AngR and HMWP2 (Fig. 7). Although CS IX overlapped CS IV, it contained a significant number of residues which were not found in group B. Analysis of the multiple alignment thus suggested that HMWP2 and AngR belong to family A and share some common features which are not present among the other members of this family.

**Phylogenetic tree.** To determine the distance between the different proteins within the superfamily, a phylogenetic tree was constructed by using the neighbor-joining (distance) method of Saitou and Nei (49) from the ClustalV program. The octapeptide repeat antigen sequence (ORA) of *Plasmo-*

*dium falciparum* (25) was used as an outgroup to root the tree. This sequence was chosen because it possessed the characteristic new class of P loop but was the most distantly related to the other members of the superfamily. As shown on Fig. 9, the existence of two families suggested by the multiple alignment was clearly seen on the phylogenetic tree. Although the size, number, and nature of the sequences used for the multiple alignment, and the program used for constructing our tree, were different from those reported by Turgay et al. (67), the two trees are very similar, suggesting that they are not fortuitous. With an exception of Lys2, the proteins or protein domains constituting family A are much larger than those in family B. This accounts for the absence

FIG. 7. The 29 protein sequences sharing the highest percentage of similarity or identity to HMWP2, aligned on 705 aa with the Pileup program. Only the regions displaying the most conserved sequences are shown. A space separates the two main protein subclasses. Uppercase letters correspond to positions where the sum of 1, 2, or 3 different amino acids represents at least 18 of 22 aa in the first subclass or at least 24 of 30 aa among the whole sequences. Numbers in parentheses represent positions of the first and last amino acids in each stretch. A period indicates a gap in the sequence. A consensus sequence displaying a maximum of three possibilities at each position was deduced from the alignment. In this consensus sequence, underlined letters represents amino acids found at least 24 times among the 30 sequences, while those which are not underlined indicate amino acids found at least 18 times among the 22 sequences composing the first subclass. Highly conserved positions in family A are represented above the consensus sequence by ‡ (association of a maximum of 3 aa present at least 21 of 22 aa) and in the superfamily by § (association of a maximum of 3 aa present at least 28 of 30 aa). Roman numerals below dotted lines indicate positions of the consensus sequences defining the superfamily (I to VI) and family A (VII to X). Protein sequences which did not extend to certain conserved segments were removed from the list. Abbreviations for proteins: ACVT,  $\gamma$ -(L- $\alpha$ -amino adipyl)-L-cysteinyl-D-valine synthetase; Luci, luciferase of *Photinus pyralis*; 4C11 and 4C1, 4-coumarate-CoA ligases of *Petroselinum crispum* and *O. sativa*, respectively; Acua, acetyl-CoA ligase. Abbreviations for organisms: asp, *A. nidulans*; pen, *P. chrysogenum*; cep, *C. acremonium*; noc, *N. lactamdurans*; met, *M. soehngenii*; neu, *Neurospora crassa*.

HMWP2	(583)	LAVtd	mryqwnYhEL	tdyArrcAgr	Lve..cgVqp	GdnVAItMsK	GagqLvaVLA	VLLAGAVYVP	VsldqPaarR	(655)
AngR	(470)	LAIth	ggesLdYitL	asyAkscAga	Lte..agVks	GdrVAVtMnK	GigqIvaVLG	ILYAGAlYVP	VsldqPqeRd	(542)
TY1	(44)	VAlVf	enrRLSYqEL	NrkANqLARA	Llek..gVqt	DslVGVmMEK	SienViaILA	VLKAGGAYVP	IDTeYPrdRI	(116)
GS1	(56)	VAlVc	eneqLTYhEL	NvkANqLARI	fiek..gIqk	DtlVGIImEK	SidIficILA	VLKAGGAYVP	IDTeYPkeRI	(128)
GS2.dA	(485)	VAVVf	edekvTYrEL	herSNqLARf	Lrek..gVkk	esiIGImMER	SvemIvcILG	ILKAGGAFVP	IDPeYPkeRI	(557)
GS2.dB	(1524)	VAVgw	kdqtLTYrEL	NerANqvARv	Lrqk..gVqp	DniVGLIVER	SpemLvgIMG	ILKAGGAY1P	LDPeYPadRI	(1596)
GS2.dC	(2560)	IAlVd	erekLSYqEL	NakANqLARv	Lrqk..gVqp	NsmVGIImVDR	SldmIvcmLG	VLKAGGAYVP	IDIdYPqeRI	(2632)
GS2.dD	(3607)	IAlVw	egqaLiYhEL	NikANqLARv	Lrek..gVtp	NhpVAImtER	SlemIvcIFs	ILKAGGAYVP	IDPaYPqeRI	(3679)
ACVS.asp.dC	(2509)	aAVVq	gdksLSYtEL	NkrANqLARY	Iq.svahLrp	DdkVLLlLdk	SldmIicILA	IWKtGsAYVP	LDPsYPkeRV	(2582)
ACVT.pen.dC	(2485)	IAlVq	gdralSYaDL	NqgANqLARY	Iq.svscIga	DdgIALmLEK	SidTIdicILA	IWKAGAAAYVP	LDPtYPpgRV	(2558)
ACVS.cep.dC	(2452)	IAlAd	gtrLSYSyEL	NerANqLVhI	Ii.ssasIva	DdrIALlLdk	SldmViaILA	VWKAGAAAYVP	LDPtYPsqRt	(2525)
ACVS.noc.dC	(2416)	VAVVh	gdvrLTYrEL	NerANrLAHh	Lr.svaepra	DelIALvLdk	SeltLvaILA	VWKAGAAyMp	IDPsYPdRI	(2489)
ACVS.asp.dB	(1428)	VAVVy	eqrsLTYrQL	NerANrMAHq	Lk.sdispkp	NsiIALvVdk	SehmIatILA	VWKtGGAYVP	IDPeYPdRI	(1501)
ACVT.pen.dB	(1402)	IAlVv	eetsLTYrEL	NerANrMAHq	Lr.sdvspnp	NevalVMDK	SehmIvrILA	VWKSGGAYVP	IDPgYPndRI	(1475)
ACVS.cep.dB	(1378)	VAVVy	edirLTYrEL	NsrANAfY	Ll.sqaalcp	NklVGLImDK	SehmItsILA	VWKtGGAYVP	IDPrYPdqRI	(1451)
ACVS.noc.dB	(1362)	IAlVv	renrLTYrEL	NerANrLAHh	Lr.svveLrp	DdlVALvLdk	SelmItsIia	aWKtGAAYVP	IDsgYPdRI	(1435)
ACVS.asp.dA	(336)	VAlVy	krqLTYgEL	NaqANcFAHy	Lr..sigilp	eqLVALfLEK	SenlIvtILG	IWKSGAAAYVP	IDPtYPdeRV	(408)
ACVT.pen.dA	(309)	IAlVc	dereLTYgEL	NaqGNSLARY	Lr..sigilp	eqLVALfLdk	SeklIvtILG	VWKSGAAAYVP	IDPtYPdeRV	(381)
ACVS.cep.dA	(275)	VALic	gdkrLTYeEL	NamANrLAHh	Lv..ssqIqt	eqLVALfLdk	telmIatILG	IWKSGAAHVP	IDPgYPdeRV	(349)
ACVS.noc.dA	(267)	eAVVc	gdvrLTYrEV	NerANqFAHw	LiqgpvrVrp	GalIGLyLdk	SdlgVvatFC	IWKSGAAAYVP	IDPaYPaerI	(341)
EntF	(472)	pAlAd	arylFSYrEm	reqVvaLAnl	Lrer..gVkp	GdsVAVaLpR	SvfltlalHh	IveAGAAwLP	LdtgYPdRI	(544)
Lys2	(269)	.....	...sFTYrDi	NrtSNivAHy	Lik..tgIkr	GdvVmIySSR	GvdlmVcVMG	VLKAGAtfsv	IDPaYPaerI	(333)

EntE	(48)	.....	...qLSYrEL	NqaAdnLacs	Lrrq..gIkp	GetalVqLgn	vaelYitfFA	lLkLGVApVl	alfshqrseL	(112)
Luci	(50)	.....	...nITYaEY	femSvrLaea	mkr..yglnt	NhrIvVcsEn	SlqffmpVLG	aLfiGVAvaP	aNdiYnereL	(114)
4Cl1	(53)	.....	...tFTYsqV	ellSrkvAsg	Lnk..lgIqq	GdtImLlLpn	SpeyffafLG	asyrGAlst	aNpfftsaeV	(117)
4Cl	(64)	.....	...vLTYaDV	drlSrrLAaa	LrraplglRr	GgvVmslLrn	SpefVlSfFA	asrvGAavtT	aNPmstpheI	(130)
MenE	(19)	.....	...tVtfaEL	faaSkRMAeq	La..ahsVrk	GdtaAIlLqn	raemVyaVhA	cflLGVkaVl	LntKlsthR	(83)
Acua.met	(133)	.....	...kITYgDL	ykeVnkFang	Lk..slgIkk	GdrVsIyMpm	ipqlpIamLA	caKlGVshiv	VfagfsskgL	(97)
Acua.neu	(88)	.....	...nvTYgEL	lreVskLams	sp..tswIrk	GdtVAIYlpm	ipeaIvamLA	ctriGAIhsv	VfagfssdsL	(152)
Acua.asp	(127)	.....	...tITYgEL	lreVsrVawv	Lk..qrgVkk	GdtVAIYlpm	ipeaIiafLA	csriGAVhsv	VfagfssdsL	(191)

Consensus VAVV- ---LTY-EL N--AN-LAR- L-----V-- G--VAL-LDK S---I---ILA VWKAGAAAYVP IDP-YP--RI  
 I II IS DV S M H I I D IGI MER G V VFG IL S G LNI Y  
 L LA F V F L N V V L M T V VF L

VII

HMWP2	(707)	AYI....	i.....	YTSGSTGTPK	GvvisHrG	(728)	(853)	amgGaTEa	(860)
AngR	(596)	AYI....	i.....	YTSGSTGTPK	GvvisHqG	(617)	(742)	amgGaTEa	(749)
TY1	(173)	AYV....	i.....	YTSGTTGPK	GtmleHkG	(194)	(308)	NaYGpTEt	(315)
GS1	(185)	AYV....	i.....	YTSGTTGPK	GtmleHkG	(206)	(321)	NaYGpTEt	(328)
GS2.dA	(609)	fYI....	i.....	YTSGTTGPK	GvnieHkN	(630)	(754)	NhYGpsEt	(761)
GS2.dB	(1650)	AYI....	m.....	YTSGSTGPK	GvnhvHrN	(1671)	(1791)	NgYGpTEt	(1798)
GS2.dC	(2686)	AYI....	i.....	YTSGTTGPK	GvnieHqS	(2707)	(2835)	NsYGvTEa	(2842)
GS2.dD	(3733)	AYI....	i.....	YTSGTTGPK	GvnieHqS	(3754)	(3876)	NeYGpTEt	(3883)
ACVS.asp.dC	(2641)	AYI....	i.....	FTSGTSGPK	GvlveOgG	(2662)	(2781)	NaYGiTEt	(2788)
ACVT.pen.dC	(2617)	AYI....	i.....	FTSGTSGPK	GvlveOka	(2638)	(2757)	NaYGtTEt	(2764)
ACVS.cep.dC	(2586)	AYV....	i.....	FTSGTTGPK	GvlveHqS	(2607)	(2725)	NaYGiTEt	(2732)
ACVS.noc.dC	(2545)	AYa....	i.....	YTSGTTGPK	avlvshqS	(2566)	(2685)	NaYgtTEt	(2692)
ACVS.asp.dB	(1561)	AYI....	i.....	YTSGTTGPK	GvnhvHhG	(1582)	(1702)	NgYGpTEt	(1709)
ACVT.pen.dB	(1535)	AYI....	i.....	YTSGTTGPK	GvveHhG	(1556)	(1676)	NgYGpTEv	(1683)
ACVS.cep.dB	(1509)	AYI....	m.....	YTSGTTGPK	GvnhvHhG	(1530)	(1650)	NgYGpTEv	(1657)
ACVS.noc.dB	(1493)	AYa....	i.....	YTSGTTGPK	avLveHrG	(1514)	(1634)	NgYGpTEt	(1641)
ACVS.asp.dA	(475)	AYV....	t.....	YTSGTTGPK	GIkqhTn	(496)	(615)	NeYgfTEs	(622)
ACVT.pen.dA	(448)	AYV....	t.....	YTSGTTGPK	GIkqhTn	(469)	(588)	NeYgfTEs	(595)
ACVS.cep.dA	(418)	AYV....	t.....	YTSGTTGPK	GIkqhTn	(439)	(558)	NeYgfTEs	(565)
ACVS.noc.dA	(405)	AYV....	t.....	YTSGTTGPK	GvpkyHys	(426)	(545)	NeYafTEa	(552)
EntF	(598)	AYI....	i.....	FTSGSTGPK	GvnhvQta	(619)	(744)	NLYGpTEa	(751)
Lys2	(407)	qvV....	vqpsnptls	FTSGSeGPK	GvlgrHfs	(437)	(558)	NmYgtTEt	(565)

Consensus AYI-----YTSGTTG-PK GV---H-G N-XG-TE-  
 V ES SS I Q S N

I = P-loop II

Table with 3 columns: Accession ID, Gene Name, and Sequence. Includes entries like HMWP2 (880) PyGfp ltNqryRVV, AngR (769) PyGyp lprgqyRVV, TY1 (335) PIGkp iqNthiYIV, etc.

Table with 3 columns: Accession ID, Gene Name, and Sequence. Includes entries like EntE (357) ypmcp ddEvvwaecr rkstaarev, Luci (363) gkvvp ffEakvvdld tGktlgnvqr, 4C11 (362) gtvvr naEmkivdpe tNaslprnqr, etc.

Consensus: PIG-- --N---YVL- -N---Q--P- ---GELYIGG --VARGY-N- -ELT--RFI- NPF
SL E RIV D K CLA ISK G DV K L
A G G HV LTE L V

VIII

Table with 3 columns: Accession ID, Gene Name, and Sequence. Includes entries like HMWP2 (938) WYRTGDLGcy w....pDGtI, AngR (827) WYRTGDMGcy w....pDGtI, TY1 (396) WYRTGDLAKw l....tDGtI, etc.

Table with 3 columns: Accession ID, Gene Name, and Sequence. Includes entries like EntE (410) fYcSGDLIsi ....dpEGyI, Luci (417) WlHSGDIAyw ....deDehf, 4C11 (416) WlHTGDIGfI ....ddDdeL, etc.

Consensus: LYRTGDLAR- ---DG-I EYLGR-D-QV KIRG-RIELG EIE--L---- -V--A-VVA- ----LVAY
M L K S Y K E V E D K I NV QVAPA V D I S S IV CGE
H GC N L I L L L S L L S A

III

IX

V

FIG. 7—Continued.

HMWP2	(1356)	La	grLPgwMVPq	rI...VfLda	LPLTANGKI	DyqALK	(1389)
AngR	(916)	Ln	kqLPnYMVPk	rI...IfLet	FPLTANGKV	DhKALT	(949)
TY1	(479)	aa	qkLPaYMLPs	yF...VkLd.	KMPLTpNGKI	DrKALP	(512)
GS1	(491)	ss	eeLPtYMIPs	yF...IqLd.	KMPLTsNGKI	DrKqLP	(524)
GS2.dA	(924)	Lg	kaLPdYMIPs	fF...VpLd.	hvrLhLNGKI	DrKsLP	(957)
GS2.dB	(1960)	ia	keLPvYMVPa	yF...VqIe.	qMPLTqNGKV	nrsALP	(1993)
GS2.dC	(3007)	La	ndrAa.MIPs	yF...VsLe.	amPLTANGKI	DkRsLP	(3039)
GS2.dD	(4044)	La	kILPaYMIPn	yF...IqLd.	sIPLTpNGKV	DrKALP	(4077)
ACVS.asp.dC	(2965)	Mk	skLPaYMVPk	yL...crLeg	qLPVTINGKL	DvRKLp	(2999)
ACVT.pen.dC	(2944)	Mk	arLPtYMVPs	hL...ccLeg	aLPVTINGKL	DvRRLP	(2978)
ACVS.cep.dC	(2909)	Lh	anLPpYMVPs	qI...hqLeg	sLPVTvNGKL	DlnRLs	(2943)
ACVS.noc.dC	(2862)	Lr	aqLmpSMVPs	lL...VrLdr	pLPmTiTGKL	DvdALP	(2896)
ACVS.asp.dB	(1883)	Mq	srLPgYMIPs	sF...IpI.s	sLPVTpSGKL	DtKALP	(1916)
ACVT.pen.dB	(1857)	Mq	srLPgYMVPs	rL...Ilv.s	kfPVTpSGKL	DtKALP	(1890)
ACVS.cep.dB	(1834)	Ml	tsLPdYMVPa	qL...VpI.a	kfPVTvSGKL	DaKALP	(1867)
ACVS.noc.dB	(1812)	Mr	kkLPeSvVPa	rv...lrI.t	dIPVTpSGKL	DaRRLP	(1845)
ACVS.asp.dA	(799)	Le	lkLPPrYMIPT	rL...Vrv.s	qIPVTvNGKa	DlRALP	(832)
ACVT.pen.dA	(772)	Le	kkLPPrYMIPT	rL...VqL.s	qIPVNVNGKa	DlRALP	(805)
ACVS.cep.dA	(742)	Le	kkLPPrYMPt	rL...VqL.a	qIPtNNGKa	DlRALP	(775)
ACVS.noc.dA	(726)	Le	qrLirIMVPa	rm...VrL.t	sIPVNVNGKV	DWRALP	(759)
EntF	(926)	Lr	etLPphMVPv	vL...lqL.p	qLPLSaNGKL	DrKALP	(959)
Lys2	(789)	Lk	krLAsYamPs	lI...Vvmd.	KLPLNpNGKV	DkpKLq	(822)
EntE	(493)	Lr	eqg...iaef	klpdrVecvd	sLPLTavGKV	DkKqLR	(527)
Luci	(502)	va	sqv...ttak	kLrggVvFvd	evPKgITGKL	DaRKlR	(536)
4Cl1	(501)	vs	qkv...vfyk	rIf.rVfVd	aIPKSpSGKI	LrKdLR	(534)
4Cl	(517)	va	kev...iyyk	kIr.eVfVd	kIPKapsGKI	LrKeLR	(550)
Acua.met	(609)	ia	f.vrktLgPv	aapteVhFvn	dLPKTrSGKI	MrRvVK	(645)
Acua.neu	(556)	ps	lqvrrSigPf	aapkaIyIvp	dLPKTLSGKI	MrRlLR	(593)
Acua.asp	(598)	Li	lqvrrSigPf	aapkaVfvvd	dLPKTrSGKI	MrRlLR	(635)
			++ ++ †		SS† \$\$\$ \$ \$		
Consensus		L-	--LP-YMVP-	-L---Y-L--	-LPLT-NGKI	D-BALP	
		M	A S I	E I I	I V N S L M K K R		
				I E	M K S T V L R K		

VI

HMWP2	(1410)	EkqVa	aLWqqLLs..	tgntvtrEDF	FqggGDSLlA	trIltqQL.h	qagyeaqlsd	LFnhPrL	(1467)
AngR	(971)	EdrVa	kIWndVLG..	ptelyksSDF	FlsGGDaynA	ievVkrC..h	kagyliklsm	LYrySTI	(1028)
TY1	(533)	EsiLv	sIWqnVLGi..	.ekiGirDNF	YsLGGDSIqA	iqvvarL.hs	.yqlkLetkd	LLnyPTI	(590)
GS1	(544)	BetLv	tIWqdVLGi..	.ekiGikDNF	YaLGGDSIka	iqvaarL.hs	.yqlkLetkd	LLkyPTI	(601)
GS2.dA	(977)	EekLa	kIWeeVLGi..	.sqiGicDNF	FsLGGHSLka	itIIsrm.nk	ecnvdiPlrl	LFeaPTI	(1035)
GS2.dB	(2013)	EmkLa	eIWnnVLGv..	.nkiGvLdnf	FsLGGHSLrA	mtmIsqV.hk	efdvLplkv	LFetPTI	(2071)
GS2.dC	(3059)	EgkLe	eIWkdVLG1..	.qrvGihDDF	FtiGGHSLka	maVIsqV.hk	ecqteVplrv	LFetPTI	(3117)
GS2.dD	(4097)	EaqLv	lIWqeVLGi..	.eliGitDNF	FeLGGHSLka	tllvakI.ye	ymqIemPlnv	VFkhSTI	(4155)
ACVS.asp.dC	(3021)	EadLc	rIWasaLGT..	.ercGicDDL	FrLGGDSItA	lhlaaqI.hh	qigrkVtvrd	IFdhPT.	(3078)
ACVT.pen.dC	(2997)	Eakmc	rIWesaLGM..	.ercGicDDL	FrLGGDSItS	lhIvaqI.hn	qvgckItvrd	IFehRTA	(3055)
ACVS.cep.dC	(2961)	BetLc	qIWasaLLGv..	.dhcGicDDL	FarGGDSIss	lrlvgdI.yr	algrkVtvkd	IYlhRsv	(3019)
ACVS.noc.dC	(2915)	EarLc	hIWasaLpg..	.gtvGicDDF	FrcGGDSIsA	lhlasqV.qr	eierkVsvky	LFdhPTV	(2973)
ACVS.asp.dB	(1934)	EsiLc	gIIsagLLdis	aqtiGsdSDF	FtLGGDSLks	tklIsfkI.he	vfgtIsvsa	LFrhRT.	(1993)
ACVT.pen.dB	(1908)	ErsLc	dIWaeLLEmh	peeiGiySDF	FsLGGDSLks	tklIsfmI.he	sfnraVavsa	LFchRTV	(1968)
ACVS.cep.dB	(1885)	EriLa	gIWseLLElp	vdriGiySDF	FsLGGDSLks	tklIsfaa.tr	algvaVsvrn	LFshPTI	(1945)
ACVS.noc.dB	(1865)	ElkLc	gIWaqVLEla	pdriGvhDDF	FaLGGDSIrA	malaaqI.tt	gfgqgLvgt	VLqhtTL	(1925)
ACVS.asp.dA	(851)	Eialg	kIWadVLGah	hIsiSrkdNF	FrLGGHSItC	iqIharI.rq	qlgviIsied	VFssRT.	(910)
ACVT.pen.dA	(824)	Eialg	eIWadVLGar	qrsvSrnDNF	FrLGGHSItC	iqIharI.rq	rlsvsIsved	VFatRTL	(884)
ACVS.cep.dA	(796)	EsdLa	aIWgnlLsvp	aqdiGseSNF	FrLGGHSIaC	iqIharV.rq	qlgqgItlee	VFqtKTL	(856)
ACVS.noc.dA	(789)	teqLr	aIWseVLGvp	qnriGerDDF	FrLGGqSIsC	illIharV.rq	rlsIsLgved	VFalRTL	(849)
EntF	(977)	Etiia	aafssLLGcd	vq..dadadDF	FaLGGHSLlA	mkIlaaqLs.r	qvarqVtpgq	VmvaSTV	(1035)
			† † †		## † ##### † †			## †	
Consensus		E--L-	-IW--VLG--	----G--DDF	F-LGGDSI-A	--L--I--	-----I--	--LF--PTI	
		V	L L E	S	SNL Y H L S	V V	V	VL R L	
					C	L	L	IY S V	

X = SBS

FIG. 7—Continued.

of the important SBS (CS X) in the latter group. However, length is not the only trait which distinguishes the two groups of proteins. Indeed, CS VII to IX, which are a hallmark of family A, are located in a region that is present in all of the proteins (except MenE for CS VIII and IX) but differs drastically between the two families. Analysis of the phylogenetic tree confirmed that Lys2 is the most distantly

related sequence in family A. It also confirmed that HMWP2 and AngR are very close to each other and belong to family A but, interestingly, constitute a specific branch which sets them apart from the antibiotic synthetases and from EntF (Fig. 9).

**Other features of HMWP2.** The 337-aa loop segment is almost completely absent from all of the proteins homolo-

## CONSENSUS SEQUENCES DEFINING:

## the superfamily of adenylate forming enzymes:

- I (Y,F)(T,S)SQ(T,S)(T,S)GXPKG(V,I)  
 II Y(G,W)XTE  
 III (L,M,W)(Y,L)X(T,S)GDL(A,V,G)  
 IV (G,D)RX(D,K)XQ(V,I,L)(K,N)(I,V,L)XGX(R,Q)(I,V,L)(E,A)(L,P)(G,A,S)E(I,V,L)(E,D)  
 V (V,C,A)(A,G)(Y,F)  
 VI (L,I,M)P(L,V,K)(T,N,S)X(N,S,T)GK(I,L,V)(D,M,L)X(R,K)XL(P,R,K)

## the family A:

- VII (I,V)(L,F,M)(A,G)(V,I)(W,L)K(A,S,T)G(A,G)AYVP(I,L,V)D(P,I)XYPXXR(I,V,L)  
 VIII GEL(Y,C,H)(I,L,V)(G,A)QXX(V,I,L)(A,S,T)(R,K,E)QY  
 IX (I,L,V)E(Y,F)LRGXDXQ(V,I,L)K(I,V,L)RQXR(I,V)E(L,P)(G,S)E(I,V)  
 X (D,S)(D,N)(F,L)(F,Y)XLGG(D,H)S(I,L)X(A,S,C)XX(L,V)

FIG. 8. Consensus sequences derived from the alignment of HMWP2 with the 29 sequences that are most similar. Positions of these sequences are shown in Fig. 9. Letters in parentheses represent positions where conserved amino acids were alternatively found. Boldface letters indicate amino acids or association of amino acids highly conserved among the members of the superfamily ( $\geq 28$  of 30) or of family A ( $\geq 21$  of 22). CS I corresponds to the position of the new type of P loop, and CS X corresponds to the position of the activated SBS.

gous to HMWP2. A search in the protein data base did not reveal any important identity. The last 35 aa of the loop share some homology with the ACVSSs, indicating that the loop may be of only 300 aa when aligned with this family of synthetases. The characteristic feature of this loop is the presence of a long leucine zipper-like sequence (Lx6Lx6Lx6Lx6L) at positions 1251 to 1279 (Fig. 1). This motif has been shown to facilitate dimerization of many gene-regulatory proteins (6, 39). However, since proline residues, which are usually absent from leucine zippers, are present within the leucine-rich segment, and since leucine zipper-like sequences have been identified in a wide variety of proteins in which their role in dimerization has not been demonstrated, the presence of a leucine zipper-like sequence in HMWP2 should be interpreted with caution.

A search in the Prosite dictionary of A. Bairoch (University of Geneva), using the Motifs program from the Genetics Computer Group package, was performed to determine whether additional motifs of interest were present in HMWP2. Two motifs deserve attention: a sugar transport protein signature and an adipokinetic hormone family signature. The first, (L,I,V,M,S,T)(D,E)X(L,I,V,M,F,A)GR(R,K)X(4,6)G, is a signature sequence found in members of a family of integral membrane proteins involved in sugar transport (56) which possesses the consensus pattern. A perfectly conserved pattern (IEFLGRRDKQVKVG) was found in HMWP2 at positions 953 to 966 and also in AngR. Interestingly, with the exception of the expected R or K residue at position 7 of the sugar transport protein consensus sequence, this pattern is well conserved in family A, being part of CS IX described above (Fig. 7 and 8). The significance of this similarity is still unclear.

The second interesting motif is the consensus sequence Q(L,V)(N,T)(F,Y)(S,T)XXW of the adipokinetic hormone

family (51). These small hormones (8 to 10 aa) are produced by arthropods and cause the release of diglycerides from the fat body. A typical sequence (QLTFTE~~W~~) was present at positions 422 to 429 in HMWP2. The relationship between these very small hormones produced by insects and the large iron-regulated protein of *Yersinia* spp. is unknown, but the fact that no other protein sequence bearing this motif has been detected hitherto in the Swiss-Prot library should be noted.

## DISCUSSION

The *irp2* gene of *Y. enterocolitica* coding for the iron-regulated protein HMWP2 is an ORF of 6,126 nt encoding a protein of approximately 228,000 Da. A typical ribosome-binding site and two potential -35 and -10 promoter regions were identified upstream of this ORF. Two Fur-binding sites overlap the second -35 region, in line with our previous observation that *irp2* is under the control of the Fur repressor in *E. coli* (9). The Fur2 box was the most highly conserved, with 16 of 19 nt (84.2%) identical to the *E. coli* consensus sequence (70). Fur boxes have also been identified in the promoter regions of the *foxA* gene coding for the ferrioxamine receptor (3), the *fcuA* gene encoding the ferrichrome receptor (35), and the hemin uptake operon (*hem*) of *Y. enterocolitica* (62) and of the *fur* gene of *Y. pestis* (60). The level of nucleotide conservation between Fur2 and the three 19-bp-long Fur boxes of *foxA*, *fcuA*, and *hemP* is high (15 of 19, 14 of 19, and 12 of 19 conserved nucleotides, respectively), indicating that the Fur recognition signal is well conserved among different iron-regulated genes of *Yersinia* spp.

The beginning of a second ORF was found immediately downstream of and in the same reading frame as the HMWP2 coding sequence. Since previous results suggested that the genes encoding HMWP1 (*irp1*) and HMWP2 form an operon, with *irp2* located upstream of *irp1*, the ORF downstream of *irp2* may correspond to the *irp1* gene.

The HMWPs were previously found in the particulate (membrane) fraction of sonicated cell extracts (8). We show here, however, that the HMWPs may form large aggregates with intermolecular disulfide bonds when cells are sonicated. This observation, together with the absence of a typical signal peptide from the N-terminal region of HMWP2, raises doubts about previous conclusions that the HMWPs are located in the outer membrane (8). However, HMWP2 does have two regions of relatively high hydrophobicity that might span the cytoplasmic membrane. The subcellular location of HMWP2 is currently under investigation.

A characteristic feature of HMWP2 is the presence of two 550-aa-long repeated domains at each extremity of the protein, suggesting a duplication of a fragment of *irp2* during evolution. The existence of five highly conserved stretches of amino acids (DR1 to DR5) in each repeat may indicate that they are important for the function, conformation, or location of HMWP2. Curiously, the region between these two large repeats is the segment that is the most highly conserved among related proteins and corresponds to the domains that are repeated three times in the ACVSSs and four times in GS2. The presence of two repeats flanking the central conserved region and of a loop of more than 300 aa located within the conserved region but absent from the other family members are characteristic features unique to HMWP2. This special organization might confer some specific properties upon HMWP2.

A very high degree of similarity or identity was found

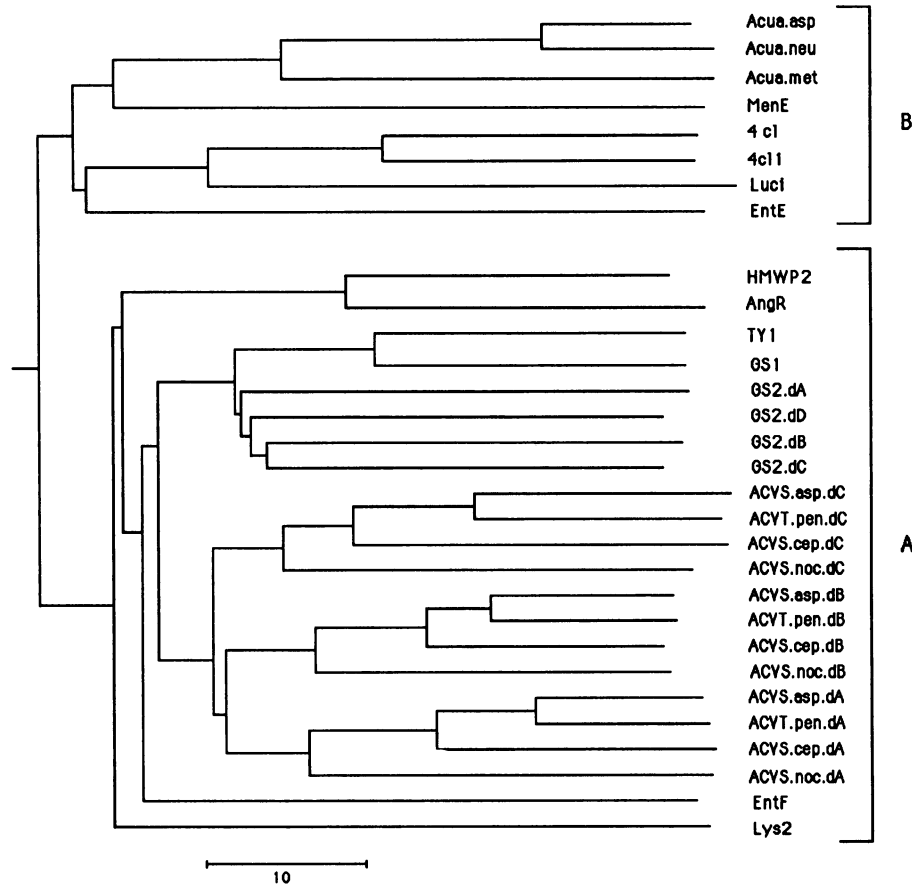


FIG. 9. Phylogenetic tree of the 30 aligned sequences obtained with the ClustalV program (distance method). The scale bar indicates the relative evolutionary distance. Abbreviations are as defined in the legend to Fig. 7.

between the entire lengths of AngR and HMWP2. In addition to their sequence homology, HMWP2 and AngR share other properties: (i) their synthesis is stimulated under iron deprivation, and Fur boxes are present upstream of their genes (9, 24); (ii) AngR and HMWP2 are both acidic (pI 6.34 and 5.81, respectively); and (iii) mutations in *angR* or *irp2* reduce virulence (9, 57). The high degree of similarity or identity found between HMWP2 and AngR might be considered surprising since *Y. enterocolitica* and *V. anguillarum* are distantly related. However, it was shown recently that the iron transport system mediated by plasmid pJM1 of *V. anguillarum* shares significant homology with those found in members of the family *Enterobacteriaceae* (36). Furthermore, Koebnik et al. (35) recently reported that the ferrichrome receptor (FcuA) of *Y. enterocolitica* shares sequence similarity with several other siderophore receptors and especially with the anguibactin receptor (FatA). Therefore, it appears that the iron-regulated genes from these two organisms are much closer than expected from their phylogenetic distance.

AngR has been reported to activate expression of genes coding for the iron uptake system of *V. anguillarum* (50). The genes thought to be under the control of AngR code for the siderophore anguibactin, for the anguibactin receptor, and for a protein (p40) involved in the transport of iron into the cell (16). A second activator (Taf) is also reported to activate transcription of these genes (65). The loci coding for AngR (*angR*), for Taf (*taf*), and for the iron uptake system are located on the 65-kb plasmid pJM1 (50). However, the

large size of AngR, the existence of the new class of P loop, and the presence of a gene coding for a putative *S*-acyl fatty acid thioesterase immediately downstream of *angR* (24) are inconsistent with a gene-regulatory function and are more consistent with a role in anguibactin biosynthesis (2). This hypothesis was strengthened by a recent report indicating that the cloned *angR* gene is able to complement an *E. coli* strain mutated in *entE* (66). Interestingly, our results and those of Turgay et al. (67) indicate that AngR and EntE belong to the same superfamily.

Moreover, the results of the multiple alignment and the phylogenetic tree show that HMWP2 and AngR may be included in family A of the superfamily of adenylate-forming enzymes (67). Except for Lys2, which is the most distantly related protein and which is too short to contain the SBS, family A is composed of different synthetases which can be divided into three groups. The first group includes the synthetases TY1, GS1, and GS2 from *B. brevis*. TY1 (115 kDa) and GS1 (126 kDa) both activate the first phenylalanine amino acid, racemize it, and form an amino acid-adenylate-enzyme complex. The aminoacyl moiety is transferred and covalently bound to another site (the SBS) on the same enzyme (34) and subsequently transferred to the second thioester-linked amino acid covalently bound to another SBS on GS2 (530 kDa). Binding of the last three amino acids also takes place on GS2. The mechanism of tyrocidine elongation is similar to that of gramicidin S except that two enzymes (TY2 and TY3) are involved. The second group of synthetases is composed of different ACVSs found in eukaryotic

filamentous fungi and bacteria. They are involved in the first step in the biosynthesis of  $\beta$ -lactam antibiotics, formation of the tripeptide  $\delta$ -(L- $\alpha$ -aminoadipyl)-L-cysteinyl-D-valine. The third group contains only EntF, an *E. coli* protein involved in biosynthesis of the siderophore enterochelin, a macrocyclic trimer of 2,3-dihydroxybenzoylserine. EntF is associated with three other proteins (EntE, EntG, and EntH) to form a large complex which carries out a series of reactions with enzyme-bound intermediates (55). EntF itself activates L-serine via an L-seryl-AMP intermediate (47) and contains a covalently bound 4'-phosphopantetheine cofactor (47). All of these enzymes, and even individual repeated domains within each enzyme, possess the new class of P loop, the SBS, and the additional eight consensus sequences.

The existence in HMWP2 and AngR of the typical new type of P loop and of the four additional consensus sequences defining the superfamily of adenylate-forming enzymes strongly suggests that they activate their substrate via an ATP-dependent process, leading to the formation of an acyladenylate-enzyme complex. Although AngR belongs to family A, the atypical sequence of the potential SBS defining this family suggests that it might not be functional and may explain why AngR is functionally close to EntE, which belongs to family B. In contrast, the presence of a perfectly conserved SBS in HMWP2 suggests that the protein may catalyze the transfer of the aminoacyl moiety from the adenylate site to the SBS, yielding an activated amino acid covalently bound either directly to the protein itself or more likely to the thiol group of 4'-phosphopantetheine cofactor.

Although the idea is entirely speculative for the moment, we propose that HMWP2 may belong to a multienzyme complex involved in the nonribosomal synthesis of a peptide. HMWP1, whose gene (*irp1*) probably forms an operon with *irp2* (9), would combine with HMWP2 to form an enzymatic complex in a manner similar to that for GS2 with GS1, TY2 and TY3 with TY1, or EntE, EntG, and EntH with EntF. The resulting product could be either an antibiotic, a siderophore, or a new class of peptide. The fact that production of HMWP1 and HMWP2 is iron regulated would be consistent with their involvement in siderophore synthesis. A new class of siderophore (yersiniabactin) which is probably a catechol-like iron chelator has been reported in *Yersinia* spp. (28, 29). Like HMWP1 and HMWP2, this molecule is chromosomally encoded and is synthesized only by highly pathogenic *Yersinia* spp. An attractive hypothesis would be that the HMWPs are enzymes involved in yersiniabactin biosynthesis. Preliminary attempts to distinguish between the parental *Y. pseudotuberculosis* strain IP2790 and its derivative mutated in *irp2* (9) on siderophore indicator medium (53) were inconclusive (7). Moreover, the strain of *Y. pseudotuberculosis* serotype II studied by Heesemann et al. was shown to produce yersiniabactin (29), while we found that none of five strains of this serotype harbored the *irp2* gene (17). These results do not eliminate the possibility that the HMWPs are involved in yersiniabactin synthesis, but the relatively distant relationship of HMWP2 to EntE, EntF, and the antibiotic synthetases in the phylogenetic tree and its unique molecular organization may also indicate that it has a novel activity that is different from those of the well-characterized enzymes in the superfamily.

#### ACKNOWLEDGMENTS

We thank members from the computer service of Institut Pasteur (M. Keller, B. Caudron, F. Chauveau, and F. Tekaiia) for advice during the computer analysis and J. D'Alayer (microsequencing laboratory) for his attempt to perform gas-phase sequencing of

HMWP2. We also thank Annie Guiyoule for helping in reading the sequencing gels.

This work was supported in part by grant CRE 920604 from the Institut National de la Santé et de la Recherche Médicale.

#### REFERENCES

1. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.). 1989. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York.
2. Bairoch, A. Unpublished results.
3. Baumler, A. J., and K. Hantke. 1992. Ferrioxamine uptake in *Yersinia enterocolitica*—characterization of the receptor protein FoxA. *Mol. Microbiol.* 6:1309–1321.
4. Bercovier, H., and H. H. Mollaret. 1984. The family *Enterobacteriaceae*, genus *Yersinia*, p. 498–506. In N. R. Krieg and J. M. Holt (ed.), *Bergey's manual of systematic bacteriology*. The Williams & Wilkins Co., Baltimore.
- 4a. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513–1523.
5. Bullen, J. J., and E. Griffiths. 1987. Iron and infection. John Wiley & Sons, New York.
6. Busch, S. J., and P. Sassone-Corsi. 1990. Dimers, leucine zippers and DNA-binding domains. *Trends Genet.* 6:36–40.
7. Carniel, E. Unpublished results.
8. Carniel, E., J. C. Antoine, A. Guiyoule, N. Guiso, and H. H. Mollaret. 1989. Purification, location, and immunological characterization of the iron-regulated high-molecular-weight proteins of the highly pathogenic yersiniae. *Infect. Immun.* 57:540–545.
9. Carniel, E., A. Guiyoule, I. Guilvout, and O. Mercereau-Puijalón. 1992. Molecular cloning, iron-regulation and mutagenesis of the *irp2* gene encoding HMWP2, a protein specific for the highly pathogenic *Yersinia*. *Mol. Microbiol.* 6:379–388.
10. Carniel, E., A. Guiyoule, O. Mercereau-Puijalón, and H. H. Mollaret. 1991. Chromosomal marker for the high pathogenicity phenotype in *Yersinia*. *Contrib. Microbiol. Immunol.* 12:192–197.
11. Carniel, E., D. Mazigh, and H. H. Mollaret. 1987. Expression of iron-regulated proteins in *Yersinia* species and their relation to virulence. *Infect. Immun.* 55:277–280.
12. Carniel, E., O. Mercereau-Puijalón, and S. Bonnefoy. 1989. The gene coding for the 190,000-dalton iron-regulated protein of *Yersinia* species is present only in the highly pathogenic strains. *Infect. Immun.* 57:1211–1217.
13. Conchas, R. F., and E. Carniel. 1990. A highly efficient electroporation system for transformation of *Yersinia*. *Gene* 87:133–137.
14. Connerton, I. F., J. R. S. Finchman, R. A. Sandeman, and M. J. Hynes. 1990. Comparison and cross-species expression of the acetyl-CoA synthetase genes of the ascomycete fungi, *Aspergillus nidulans* and *Neurospora crassa*. *Mol. Microbiol.* 4:451–460.
15. Coque, J. J. R., J. F. Martin, J. G. Calzada, and P. Liras. 1991. The cephamycin biosynthetic genes *pcbAB*, encoding a large multidomain peptide synthetase, and *pcbC* of *Nocardia lactam-durans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. *Mol. Microbiol.* 5:1125–1133.
16. Crosa, J. H. 1989. Genetics and molecular biology of siderophore-mediated iron transport in bacteria. *Microbiol. Rev.* 53:517–530.
17. de Almeida, A. M. P., A. Guiyoule, I. Guilvout, I. Itean, G. Baranton, and E. Carniel. 1993. Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. *Microb. Pathog.* 14:9–21.
18. Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.
19. De Wet, J. R., K. V. Wood, M. De Luca, D. R. Helsinki, and S. Subrami. 1987. Firefly luciferase gene, structure and expression in mammalian cells. *Mol. Cell. Biol.* 7:725–737.

20. Diez, B., S. Gutierrez, J. L. Barredo, P. van Solingen, L. H. M. van der Voor, and J. F. Martin. 1990. The cluster of penicillin biosynthesis genes. Identification and characterization of the *pcbAB* gene encoding the  $\alpha$ -aminoadipyl-cysteiny-valine synthetase and linkage to the *pcbC* and *penDE* genes. *J. Biol. Chem.* **265**:16358–16365.
21. Dodd, I. B., and J. B. Egan. 1987. Systematic method for the detection of potential 1 Cro-like DNA-binding regions in proteins. *J. Mol. Biol.* **194**:557–564.
22. Driscoll, J. R., and H. W. Taber. 1992. Sequence organization and regulation of the *Bacillus subtilis* *menBE* operon. *J. Bacteriol.* **174**:5063–5071.
23. Eggen, R. I. L., A. C. M. Geerling, A. B. P. Boshoven, and W. M. De Vos. 1991. Cloning, sequence analysis, and functional expression of the acetyl coenzyme A synthetase gene from *Methanotrix soehngenii* in *Escherichia coli*. *J. Bacteriol.* **173**:6383–6389.
24. Farrell, D. H., P. Mikesell, L. A. Actis, and J. H. Crosa. 1990. A regulatory gene, *angR*, of the iron uptake system of *Vibrio anguillarum*: similarity with pMhage P22 *cro* and regulation by iron. *Gene* **86**:45–51.
25. Favalaro, J. M., V. M. Marshall, P. E. Crewther, R. L. Coppel, D. J. Kemp, and R. F. Anders. 1989. cDNA sequence predicting an octapeptide repeat antigen of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **32**:297–299.
26. Gutierrez, S., B. Diez, E. Montenegro, and J. F. Martin. 1991. Characterization of the *Cephalosporium acremonium* *pcbAB* gene encoding  $\alpha$ -aminoadipyl-cysteiny-valine synthetase, a large multidomain peptide synthetase: linkage to the *pcbC* gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. *J. Bacteriol.* **173**:2354–2365.
27. Harley, C. B., and R. P. Reynolds. 1987. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* **15**:2343–2361.
28. Heesemann, J. 1987. Chromosomal-encoded siderophores are required for mouse virulence of enteropathogenic *Yersinia* species. *FEMS Microbiol. Lett.* **48**:229–233.
29. Heesemann, J., K. Hantke, T. Vocke, E. Saken, A. Rakin, I. Stojiljkovic, and R. Berner. 1993. Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65000 Da and pesticin sensitivity. *Mol. Microbiol.* **8**:397–408.
30. Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**:237–244.
31. Hori, K., Y. Yamamoto, T. Minetoki, T. Kurotsu, M. Kanda, S. Miura, K. Okamura, J. Furuyama, and Y. Saito. 1989. Molecular cloning and nucleotide sequence of the gramicidin S synthetase 1 gene. *J. Biochem.* **106**:639–645.
32. Hori, K., Y. Yamamoto, K. Tokita, F. Saito, T. Kurotsu, M. Kanda, K. Okamura, J. Furuyama, and Y. Saito. 1991. The nucleotide sequence for a proline-activating domain of gramicidin S synthetase 2 gene from *Bacillus brevis*. *J. Biochem.* **110**:111–119.
33. Jackson, S., and T. W. Burrows. 1956. The virulence enhancing effect of iron on non-pigmented mutants of virulent strains of *Pasteurella pestis*. *Br. J. Exp. Pathol.* **37**:577–583.
34. Kleinkauf, H., and H. von Döhren. 1990. Nonribosomal biosynthesis of peptide antibiotics. *Eur. J. Biochem.* **192**:1–15.
35. Koebnik, R., K. Hantke, and V. Braun. 1993. The TonB-dependent ferrichrome receptor FcuA of *Yersinia enterocolitica*: evidence against a strict coevolution of receptor structure and substrate specificity. *Mol. Microbiol.* **7**:383–393.
36. Köster, W. L., L. A. Actis, L. S. Waldbeser, M. E. Tolmasky, and J. H. Crosa. 1991. Molecular characterization of the iron transport system mediated by the pJM1 plasmid in *Vibrio anguillarum* 775. *J. Biol. Chem.* **266**:23829–23833.
37. Krätzschmar, J., M. Krause, and M. A. Marahiel. 1989. Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. *J. Bacteriol.* **171**:5422–5429.
38. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**:105–132.
39. Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**:1759–1764.
40. Loyza, E., H. Hoffman, C. Douglas, W. Schulz, D. Scheel, and K. Halbrock. 1988. Primary structure and catalytic properties of isoenzymes encoded by the two 4-coumarate CoA ligase genes in parsley. *Eur. J. Biochem.* **176**:661–667.
41. MacCabe, A. P., H. van Liempt, H. Palissa, S. E. Unkles, M. B. R. Riach, E. Pfeifer, H. von Döhren, and J. R. Kinghorn. 1991.  $\gamma$ -(L- $\alpha$ -aminoadipyl)-L-cysteiny-D-valine synthetase from *Aspergillus nidulans*. Molecular characterization of the *acvA* gene encoding the first enzyme of the penicillin biosynthetic pathway. *J. Biol. Chem.* **266**:12646–12654.
42. Marck, C. 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* **16**:1829–1836.
43. Morris, M. E., and S. Jinks-Robertson. 1991. Nucleotide sequence of the *LYS2* gene of *Saccharomyces cerevisiae*: homology to *Bacillus brevis* tyrocidine synthetase 1. *Gene* **98**:141–145.
44. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**:2444–2448.
45. Pressler, U., H. Staudenmaier, L. Zimmerman, and V. Braun. 1988. Genetics of the iron dicitrate transport system of *Escherichia coli*. *J. Bacteriol.* **170**:2716–2724.
46. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**:319–353.
47. Rusnak, F., M. Sakaitani, D. Drueckhammer, J. Reichert, and C. T. Walsh. 1991. Biosynthesis of the *Escherichia coli* siderophore enterobactin: sequence of the *entF* gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. *Biochemistry* **30**:2916–2927.
48. Rutz, J. M., T. Abdullah, S. P. Singh, V. I. Kalve, and P. E. Klebba. 1991. Evolution of the ferric enterobactin receptor in gram-negative bacteria. *J. Bacteriol.* **173**:5964–5974.
49. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
50. Salinas, P. C., M. E. Tolmasky, and J. H. Crosa. 1989. Regulation of the iron uptake system in *Vibrio anguillarum*: evidence for a cooperative effect between two transcriptional activators. *Proc. Natl. Acad. Sci. USA* **86**:3529–3533.
51. Schaffer, M. H., B. E. Noyes, C. A. Slaughter, G. C. Thorne, and S. J. Gaskell. 1990. The fruitfly *Drosophila melanogaster* contains a novel cloned adipokinetic-hormone-family peptide. *Biochem. J.* **269**:315–320.
52. Schlumbohm, W., T. Stein, C. Ullrich, J. Vater, M. Krause, M. A. Marahiel, V. Kruff, and B. Wittmann-Liebold. 1991. An active serine is involved in covalent substrate amino acid binding at each reaction center of gramicidin S synthetase. *J. Biol. Chem.* **266**:23135–23141.
53. Schwyn, B., and J. B. Neillands. 1987. Universal chemical assay for the detection and determination of siderophores. *Anal. Biochem.* **160**:47–56.
54. Sikkema, D. J., and R. R. Brubaker. 1989. Outer membrane peptides of *Yersinia pestis* mediating siderophore-independent assimilation of iron. *Biol. Metals* **2**:174–184.
55. Silver, S., and M. Walderhaug. 1992. Gene regulation of plasmid- and chromosome-determined inorganic ion transport in bacteria. *Microbiol. Rev.* **56**:195–228.
56. Silverman, M. 1991. Structure and function of hexose transporters. *Annu. Rev. Biochem.* **60**:757–794.
57. Singer, J. T., K. A. Schmidt, and P. W. Reno. 1991. Polypeptides p40, pOM2, and pAngR are required for iron uptake and for virulence of the marine fish pathogen *Vibrio anguillarum* 775. *J. Bacteriol.* **173**:1347–1352.
58. Smith, D. J., A. J. Earl, and G. Turner. 1990. The multifunctional peptide synthetase performing the first step of penicillin biosynthesis in *Penicillium chrysogenum* is a 421 073 dalton protein similar to *Bacillus brevis* peptides antibiotic syn-



- thetases. *EMBO J.* **9**:2743–2750.
59. **Staab, J. F., M. Elkins, and C. F. Earhart.** 1989. Nucleotide sequence of the *Escherichia coli entE* gene. *FEMS Microbiol. Lett.* **59**:15–20.
60. **Staggs, T. M., and R. D. Perry.** 1992. Fur regulation in *Yersinia* species. *Mol. Microbiol.* **6**:2507–2516.
61. **Stock, J. B., A. J. Ninfa, and A. M. Stock.** 1989. Protein phosphorylation and regulation of adaptative responses in bacteria. *Microbiol. Rev.* **53**:450–490.
62. **Stojiljkovic, I., and K. Hantke.** 1992. Hemin uptake system of *Yersinia enterocolitica*: similarities with other TonB-dependent systems in Gram-negative bacteria. *EMBO J.* **11**:4359–4367.
63. **Taylor, W. R., and N. M. Green.** 1989. The predicted secondary structure of the nucleotide-binding sites of six cation-transporting ATPases lead to a probable tertiary fold. *Eur. J. Biochem.* **179**:241–248.
64. **Tinocco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. *Nature (London) New Biol.* **246**:40–41.
65. **Tolmasky, M. E., L. A. Actis, and J. H. Crosa.** 1988. Genetic analysis of the iron uptake region of the *Vibrio anguillarum* plasmid pJM1: molecular cloning of genetic determinants encoding a novel *trans*-activator of siderophore biosynthesis. *J. Bacteriol.* **170**:1913–1919.
66. **Tolmasky, M. E., L. A. Actis, and J. H. Crosa.** 1992. Functional characterization of the AngR protein, Abstr. B-115. Abstr. 92nd Gen. Meet. Am. Soc. Microbiol. 1992.
67. **Turgay, K., M. Krause, and M. A. Marahiel.** 1992. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol. Microbiol.* **6**:529–546.
68. **Von Gabain, A., J. G. Belasco, J. L. Schottel, A. C. Y. Chang, and S. Cohen.** 1983. Decay of mRNA in *Escherichia coli*: investigation of the fate of specific segments of transcripts. *Proc. Natl. Acad. Sci. USA* **80**:653–657.
69. **Weckerman, R., R. Fürbab, and M. A. Marahiel.** 1988. Complete nucleotide sequence of the *tycA* gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. *Nucleic Acids Res.* **16**:11841.
70. **Wee, S., J. B. Neilands, M. L. Bittner, B. C. Hemming, B. L. Haymore, and R. Seetharam.** 1988. Expression, isolation and properties of Fur (ferric uptake regulation) protein of *Escherichia coli* K12. *Biol. Metals* **1**:62–68.
71. **Zhao, Y., S. D. Kung, and S. K. Dube.** 1990. Nucleotide sequence of rice 4-coumarate:CoA ligase gene, 4-CL.1. *Nucleic Acids Res.* **18**:6144.