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,¹ ODILE MERCEREAU-PUIJALON,² SERGE BONNEFOY,² ANTHONY P. PUGSLEY,³ AND ELISABETH CARNIEL^{1*}

Unité de Bactériologie Moléculaire et Médicale,¹ Unité de Parasitologie Expérimentale,² and Unité de Génétique Moléculaire,³ 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 highmolecular-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 β -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 ironregulated proteins were identified only recently (11, 54, 60). No precise function has yet been attributed to the ironregulated 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 highmolecular weight iron-regulated proteins (high-molecularweight 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 HM-WPs 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 α and JM101 (1), were used in this study. pIR2 is a pUC18 recombinant with an 8-kb ClaI insert containing the entire *irp2* gene from strain Ye8081 (9). Phages M13mp18 and M13mp19 (1) were used for singlestranded 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 μ M FeCl₃ or 50 μ M α - α '-dipyridyl (Sigma Chemical Co.), respectively, to the chemically defined medium. Ampicillin (100 μ 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 subclones 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 MgCl₂-25 mM K_2 HPO₄ (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 β -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 ΔG (25°C) of 20.8 kcal (1 kcal = 4.184 kJ) (64). The absence of a poly(T) stretch immediately downstream of the stemloop 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-

1			AA	ACGC	GTGA	TGTA	ACCG	GGGT	CGCG	cccc	CCTA	AATI	cctc	CCTG	ACAG	50		946 243	GTC V	AAT N	ATT I	GAC D	CTG L	CTG L	ATT I	ATG M	GAT D	GCC A	TCC S	AGC S	TTT F	ACG T	CTT L	990 257
51	AGG	CCCG IR2		cccc	ATCC	GGGC	CTCT 2	<u>gt</u> ta	TTTC	cecc	TGTT	-35 ACCA	ACCO	gaaa IR3	<u>cccc</u>	109		991 258	TTC F	TTC F	GAT D	GAG E	CTT L	AAC N	GCC A	CTG L	CTG L	GCC A	GGA G	GAA E	тсс S	CTG L	TCG S	1035 272
110	CAT	AAAT	AAAQ	CCGT	TTCG R3	<u>GGT</u> A de	GCAT	ATAA Fur	TTGT	TAAT	AATT Fur2	ATTA	TTCT	CAAA	-10 TTAT	168	1	.036 273	GCT A	ATC I	GAC D	ACC T	CGC R	TAT Y	GAT D	TTC F	CGC R	TCG S	TAT Y	TTG L	CTG L	CAC H	CAG Q	1080 287
169 1	TAT	тстс	ATAT	GAGA	AATG	CTTT	TCGG	TAAG	ACGT	GCCA	TCAG	gagg SD	АААА	ATG M	ATT I	225 2	:	1081 288	CAG Q	AAG K	ATC I	AAT N	CAA Q	CCA P	CTG L	AGA R	GAC D	GAC D	GCA A	CGC R	GCT A	TAC Y	TGG W	1125 302
226 3	TCT S	GGC G	GCA A	CCA P	S TCT	AAG K	GAT D	TCG S	CTG L	TTA L	CCG P	GAC D	AAC N	CGC R	CAC H	270 17	:	1126 303	CTG L	GCG A	AAA K	GCA A	тсс s	ACG T	CTT L	CCC P	CCC P	GCG		GTC V 2	TTG L	CCG P	CTG	1170 317
271 18	GCG A	GCT A	GAT D	TAC Y	CAA Q	Q Q	L TTA	CGC R	GAG E	CGG R	CTT L	ATA I	CAG Q	GAA E	CTG L	315 32	:	1171 318	GTC V	TGC C	GAA	CCC P	GCC A	ACG T	CTA L	CGT R	GAA E	GTC V	CGT R	AAT N	ACC T	CGG R	CGC R	1215 332
316 33	AAT N	TTA L	ACG T	CCG P	CAG Q	CAG Q	TTA L	CAT H	GAC D	GAA E	AGC S	AAC N	CTG L	ATC	CAG	360 47	:	1216 333	CGC R	ATG M	ATT I	GTC V	CCA P	GCA A	ACA T	CGC R	tgg W	CAC H	GCC A	TTT F	AGC S	AAC N	CGG R	1260 347
361 48	GCC A	GGC G	CTG L SBS	GAT D	TCA S	ата <u>I</u>	AGA R	TTG L	ATG M	AGA R	tgg W	TTA L	CAC H	TGG W	TTT F	405 62	:	1261 348	GCC A	GGC G	GAA E	ТАТ Ү	GGC G	GTG V	ACG T	CCA P	ACA T	ATG M	GCG A	CTG L	GCG A	ACC T	TGT C	1305 362
406 63	CGT R	aaa K	AAT N P2	GGC G	ТАС Ү	CGC R	CTT L	ACC T	CTT L	CGC R	GAG E	CTG L	TAT Y	GCC A	GCC A	450 77	:	1306 363	TTT F	тст S	GCC A	GTG V	CTG L	GCT A	CGC R	TGG W	GGC G	GGC G	CTG L	ACG T	CGT R	CTG L	CTG L	1350 377
451 78	CCC P	ACG T	CTG L	GCG A	GCA A	TGG W	AAC N	CAG Q P3	TTA L	ATG M	CTC L	AGC S	CGG R	TCG S	CCG P	495 92	:	1351 378	CTT L	AAC N	ATC	ACC T	TTA	TTC F	GAC D DR3	CGC R	CAG Q	CCG P	CTG L	CAC H	CCA P	GCG A	GTT V	1395 392
496 93	GAG E	AAC N	GCG A	GAA	GAA E	GAA E	ACG T	CTG L	CCC P	GAC D	GAA E	TCA S	TCC S	TGG W	CCG P	540 107	:	1396 393	GGC G	GCG A	ATG M	CTT L	GCC A	GAC D	TTC F	ACC T	AAT N	ATT I	CTT L	CTG L	CTG L	GAT D	ACC T	1440 407
541 108	AAC N	ATG M	ACC T	GAA E	AGT S	ACC T	CCC P	TTC F	сса Р	TTG L	ACG	CCG P	GTA V	CAG Q	CAC H	585 122	:	1441 408	GCC A	tgc C	GAT D	GGC G	GAT D	ACC T	GTC V	AGC S	AAC N	CTG L	GCG A	CGT R	AAA K	AAC N	CAG Q	1485 422
586 123	GCC A	TAC Y	CTG	ACG T	GGC	CGC R	ATG M	CCG P DR1	GGG G	CAG	ACG T	CTT	GGC G	GGC G	GTG V	630 137	:	1486 423	CTC L	ACG T	TTT F	ACG T	GAG E	GAC D	tgg W	GAG E	CAT H	CGC R	CAC H	tgg W	TCC S	GGC G	GTC V	1530 437
631 138	GGT G	TGC C	CAC	CTG L	тат _ <u>¥</u>	CAG Q	GAG E	TTT F	GAA E	GGC G	САТ Н	tgt C	CTG L	ACG T	GCG A	675 152	:	1531 438	GAA E	TTA L	CTC L	CGT R	GAA E	CTC L	AAA K	CGC R	CAG Q	CAG Q	CGC R	TAC Y	CCC P	CAC H	GGC G	1575 452
676 153	TCG S	CAA Q	CTG L	GAG E	CAG Q	GCC A	ATC I	ACG T	ACC T	TTG L	CTG L	CAA Q	CGC R	CAC H	CCA P	720 167	:	1576 453	GCC A	CCG P	GTG V	GTA	TTT F D	ACC T R4	AGC S	AAT N	CTG L	GGG G	CGT R	тсс s	CTC L	ТАС Ү	AGC S	1620 467
721 168	ATG M	CTG L	САТ Н	ATC I	GCC A	TTT F	CGC R	CCC P	GAC D	GGG G	CAG Q	CAG Q	GTC V	tgg W	CTA L	765 182	:	1621 468	AGC S	CGC R	GCA A	GAA E	TCG S	CCG P	TTG L	GGC G	GAG E	CCG P	GAA E	TGG W	GGC G	ATC I	TCG S	1665 482
766 183	CCG P	CAA Q	CCT P	TAC Y	tgg W	AAC N	GGC G	GTC V	ACC T	GTT V	CAT H	GAT D	TTA L	CGC R	CAT H	810 197	:	1666 483	CAA Q	ACG T		CAG Q	GTC V	tgg W	ATA I	GAT D	CAT H	CTG L	GCG A	TTC F	GAG E	CAT H	CAC H	1710 497
811 198	AAC N	GAC D	GCT A	GAA E	AGC S	CGC R	CAG Q	GCC A	TAT Y	CTG L	GAC D	GCA A	CTG L	CGC R	CAG Q	855 212	:	1711 498	GGC G	GAG E	GTC V	TGG W	CTG L	CAA Q	TGG W	GAC D	AGC S	AAC N	GAC D	GCG A	CTG L	TTC F	CCT P	1755 512
856 213	CGC R	CTC L	AGC S	CAC H	CGT R	CTT L	TTA L	CGC R	GTG V	GAG E	ATC I	GGC G	GAA E	ACA T	TTT F	900 227	:	1756 513	CCG P	GCG A	TTA L	GTC V	gaa E	ACA T	TTG L	TTC F	GAC D	GCC A	TAC Y	tgc C	CAG Q	TTG L	ATT I	1800 527
901 228	GAT D	TTT F	CAG Q	CTG L	ACG T	CTC L	TTG L	CCG P	GAC D	AAT N	CGC R	CAC H	CGC R	CTC L	CAT H	945 242	:	1801 528	AAC N	CAA Q	CTC L	TGC C	GAT D	GAC D	GAA E	AGC S	GCC A	tgg W	CAA Q	AAG K	CCG P	TTC F	GCA A	1845 542

'n

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 SGTTGXPKG.

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 polymerasebinding 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

2746 CGG GCC TTC CGG CCA CAA GGA CAA TTT ATC GCG ATG GGC GGC GCC 2790

1846	GAT	ATG	ATG	CCC	GCC	AGC	CAG	CGC	GCG	ATA	CGC	GAA	CGG	GTC	AAC	1890
543	D	M	M	P	A	S	Q	R	A	I	R	E	R	V	N	557
1891	GCC	ACC	GGC	GCC	CCC	ATT	CCC	GAA	GGC	TTG	CTG	CAT	GAA	GGC	ATT	1935
558	A	T	G	A	P	I	P	E	G	L	L	H	E	G	I	572
1936	TTC	CGT	ATC	GCT	CTG	CAA	CAG	CCG	CAG	GCG	CTG	GCG	GTA	ACG	GAC	1980
573	F	R	I	A	L	Q	Q	P	Q	A	L	A	V	T	D	587
1981	ATG	CGT	TAT	CAG	tgg	AAT	TAT	CAT	GAG	CTG	ACA	GAC	TAT	GCC	CGC	2025
588	M	R	Y	Q	W	N	Y	H	E	L	T	D	Y	A	R	602
2026	CGT	tgc	GCA	GGC	AGG	TTA	GTC	GAG	tgc	GGG	GTT	CAG	CCC	GGC	GAT	2070
603	R	C	A	G	R	L	V	E	C	G	V	Q	P	G	D	617
2071	AAT	GTG	GCT	ATC	ACG	ATG	TCG	aaa	GGC	GCA	GGA	CAA	CTT	GTT	GCG	2115
618	N	V	A	I	T	M	S	K	G	A	G	Q	L	V	A	632
2116	GTT	CTG	GCC	GTC	CTG	CTG	GCC	GGG	GCA	GTT	TAC	GTT	CCG	GTT	TCG	2160
633	V	L	A	V	L	L	A	G	A	V	Y	V	P	V	S	647
2161 648	TTG L	GAC D	CAG Q	CCT P	GCC A	GCA A	CGG R	HI CGC R	GAG E	AAA K	ATC I	TAC Y	GCT A	GAC D	GCC A	2205 662
2206	AGC	GTC	CGG	CTG	GTG	CTC	ATT	tgc	CAG	CAC	GAC	GCC	AGC	GCC	GGG	2250
663	S	V	R	L	V	L	I	c	Q	H	D	A	S	A	G	677
2251	TCA	GAC	GAT	ATT	CCC	GTC	CTT	GCC	tgg	CAG	CAG	GCC	ATT	GAG	GCG	2295
678	S	D	D	I	P	V	L	A	W	Q	Q	A	I	E	A	692
2296	GAG	CCG	ATC	GTC	AAC	CCG	GTG	GTA	CGC	GCC	CCC	ACG	CAA	CCG	GCC	2340
693	E	P	I	V	N	P	V	V	R	A	P	T	Q	P	A	707
2341 708	TAC Y	ATT I	ATC I	TAC Y	ACT T	tcc s	GGC G	TCT S	ACC	GGC	ACG T	CCG P	AAA K	GGG G	GTA V	2385 722
2386 723	GTC V	ATT I	TCT S	CAC H	CGG R	GGA G	GCG A	CTC L	P- AAC N	ACC T	TGT C	TGC C	GAT D	ATC I	AAT N	2430 737
2431	ACC	CGC	TAT	CAG	GTT	GGC	CCG	CAT	GAC	AGG	GTG	CTG	GCC	CTC	TCC	2475
738	T	R	Y	Q	V	G	P	H	D	R	V	L	A	L	S	752
2476	GCC	CTG	CAT	TTT	GAT	TTA	TCG	GTT	TAC	GAC	ATT	TTT	GGC	GTA	CTG	2520
753	A	L	H	F	D	L	S	V	Y	D	I	F	G	V	L	767
2521	CGT	GCG	GGC	GGC	GCG	CTG	GTG	ATG	GTG	ATG	GAA	AAT	CAA	CGG	CGC	2565
768	R	A	G	G	A	L	V	M	V	M	E	N	Q	R	R	782
2566	GAT	CCT	CAC	GCA	tgg	TGT	GAG	CTG	ATC	CAG	CGC	CAT	CAG	GTC	ACG	2610
783	D	P	H	A	W	C	E	L	I	Q	R	H	Q	V	T	797
2611	CTC	TGG	AAC	AGC	GTC	CCG	GCG	CTG	TTC	GAT	ATG	CTG	CTG	ACC	TGG	2655
798	L	W	N	S	V	P	A	L	F	D	M	L	L	T	W	812
2656	тдт	GAA	GGT	TTC	GCC	GAC	GCC	ACG	CCG	GAA	AAC	CTG	CGC	GCA	GTG	2700
813	С	E	G	F	A	D	A	T	P	E	N	L	R	A	V	827
2701	ATG	CTT	тсс	GGC	GAC	tgg	ATC	GGA	CTT	GAC	CTC	CCC	GCC	CGT	TAT	2745
828	M	L	s	G	D	W	I	G	L	D	L	P	A	R	Y	842
															FIC	G. 1—C

043	R	A	£	R	r	¥	G	¥	r	1	~	м	G	9	ñ	0.57
2791	ACC	GAG	GCG	TCT	ATC	tgg	tct	AAC	GCC	tgc	GAA	ATT	CAC	GAC	GTC	2835
858	T	E	A	S	I	W	S	N	A	C	E	I	H	D	V	872
2836	CCT	GCC	CAC	tgg	CGT	тсс	ATC	CCT	TAC	GGT	TTT	CCG	CTA	ACC	AAC	2880
873	P	A	H	W	R	s	I	P	Y	G	F	P	L	T	N	887
2881	CAA	CGC	TAC	CGG	gtg	gtg	GAT	GAA	CGG	GGC	CGG	GAC	tgc	CCT	GAC	2925
888	Q	R	Y	R	V	V	D	E	R	G	R	D	C	P	D	902
2926	tgg	GTG	тсс	GGT	GAA	TTA	tgg	ATT	GGC	GGC	ATC	GGG	GTC	GCG	GAA	2970
903	W	V	s	G	E	L	W	I	G	G	I	G	V	A	E	917
2971	GGC	TAT	TTC	AAC	GAT	TCC	CTG	CGC	AGC	GAG	CAG	CAA	TTT	TTG	ACG	3015
918	G	Y	F	N	D	S	L	R	S	E	Q	Q	F	L	T	932
3016	CTC	CCG	GAC	GAG	CGC	tgg	TAT	CGC	ACC	GGC	GAT	CTC	GGC	tgc	TAC	3060
933	L	P	D	E	R	W	Y	R	T	G	D	L	G	C	Y	947
3061	TGG	CCA	GAC	GGC	ACA	ATC	GAG	TTC	CTC	GGT	CGT	CGC	GAC	AAG	CAG	3105
948	W	P	D	G	T	I	E	F	L	G	R	R	D	K	Q	962
3106	GTC	AAA	GTC	GGA	GGA	TAT	CGC	ATC	GAG	CTG	GGC	GAA	ATC	GAA	AGC	3150
963	V	K	V	G	G	Y	R	I	E	L	G	E	I	E	S	977
3151	GCG	CTC	AGC	CAG	TTG	GCG	GGG	gtg	AAA	CAA	GCA	ACC	GTT	CTG	GCG	3195
978	A	L	S	Q	L	A	G	V	K	Q	A	T	V	L	A	992
3196	ATC	GGC	GAA	AAA	GAA	AAA	ACG	CTG	GCG	GCA	TAC	GTG	GTT	CCT	CAG	3240
993	I	G	E	K	E	K	T	L	A	A	Y	V	V	P	Q	1007
3241	AGC	GAG	GCT	TTT	TGC	GTT	ACC	GAT	CAT	CGG	AAC	CCG	GCA	TTG	CCG	3285
1008	S	E	A	F	C	V	T	D	H	R	N	P	A	L	P	1022
3286	AAG	GCG	tgg	CAC	ACG	CTT	GCG	GGA	ACG	TTG	CCC	TGT	тGC	GCC	ATC	3330
1023	K	A	W	H	T	L	A	G	T	L	P	C	C	A	I	1037
3331	TCG	CCA	GAG	ATC	TCC	GCA	GAA	CAG	GTA	GCC	GAT	TTC	CTT	CAG	CAT	3375
1038	S	P	E	I	S	A	E	Q	V	A	D	F	L	Q	H	1052
3376	CGC	CTG	TTA	AAA	CTG	AAG	CCG	GGT	CAC	ACC	GCT	GGC	GCC	GAT	CCT	3420
1053	R	L	L	K	L	K	P	G	H	T	A	G	A	D	P	1067
3421	ATC	CCC	CTG	ATG	AAT	TCA	CTC	GCT	ATC	CAG	CCG	CGC	tgg	CAG	GCC	3465
1068	I	P	L	M	N	S	L	A	I	Q	P	R	W	Q	A	1082
3466	gtg	GTG	GAA	CGC	tgg	TTA	GCA	TTT	CTG	GTG	ACG	CAA	CGG	CGA	CTG	3510
1083	V	V	E	R	W	L	A	F	L	V	T	Q	R	R	L	1097
3511	AAG	CCC	GCT	GCT	GAA	GGT	TAT	CAG	GTC	tgc	GCT	GGT	GAA	GAA	CGC	3555
1098	K	P	A	A	E	G	Y	Q	V	C	A	G	E	E	R	1112
3556	GAG	GAT	GAG	CAC	CCG	CAC	TTC	AGC	GGA	CAT	GAT	TTA	ACG	TTA	тсg	3600
1113	E	D	E	H	P	H	F	S	G	H	D	L	T	L	S	1127
3601	CAA	ATT	CTT	CGC	GGT	GCC	CGT	AAC	GAA	CTG	TCG	TTA	CTG	AAC	GAC	3645
1128	Q	I	L	R	G	A	R	N	E	L	S	L	L	N	D	1142
 A	1															

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 NH₂-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	С АТ	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	Н	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 1188	ACC T	GGC G	CGC R	GCC A	GCA A	GAA E	тсс S	CTG L	TTG L	GCA A	CAG Q	CTC L	AAC N	GCC A	GGA G	3825 1202	4681 1488	TCT S	CCT P	GAA E	GAA E	CGC R	TAC Y	CAG Q	CCC P	TTT F	GCG	CTT L	ACC	GAC D	GTG	CAG Q	4725 1502
3826 1203	CAG Q	ATT I	GAG E	TAT Y	GTC V	GGG G	CTT L	GAG E	CAG Q	AGC S	CAG Q	GAG E	ATG M	CTA L	CTG L	3870 1217	4726 1503	CAG Q	GCT A	TAC	CTG L	GTG V	GGG G	CGT R	CAG	CCG P	GGC G	TTT F	ACC T	CTG L	GGC	GGC G	4770 1517
3871	AGC	GCC	CGG	CAG	AGG	CTC	GCC	тсс	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	G	A	R	L	S	1232	1518	V	G	S	H	F	F	V	E	F	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		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 1263	CCC P	GGG	CTC L	CTT L	GCG	ACA T	тта _ ¹	саа 9	CAG Q	CTT L	GCC A	GTT V	CCC P	GGC G	GCG	4050 1277	4906 1563	GAA E	CAG Q	ACG T	CCC P	CCC P	tgg W	gtg V	ATA I	CCC P	ACA T	CAC H	ACC T	CTT L	CAT H	ACG T	4950 1577
4051	CTG	CTC	TAC	GTG	ATG	GAG	TTT	CGC	CAG	TTA	ACG	CCG	TCC	GCC	CTG	4095	4951	CCT	GAA	GAG	GCG	TTG	CGG	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	САТ	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	Н	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	тсg	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	Ω	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		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	ACG	CGC	5355
1398	N	Q	A	E	A	D	L	P	Q	G	D	I	E	K	Q	1412	1698	T	P	R	F	A	R	L	N	G	A	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	С АТ	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	Н	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
								SI	BS																#2								

FIG. 1-Continued.

fraction (8). Analysis of the predicted NH_2 -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 alky5446 AGT GCG CAG CCT GAG TTC ACG CTT AAC CTT ACG CTT TTC GAC AGG 5490 1743 <u>S A Q P</u> E F T <u>L N L T L F D R</u> 1757 DR3 5535 5491 CGA CCG CTG CAC CCG CAA ATC AAC CAG ATT CTG GGC GAT TTC ACT 1758 R P L H P Q I N Q I L G D F T 1772 5580 1787 5625 5581 AGC GCG CAG TCA CTA CAG CAG CGG CTG AGC CAG AAC CTC AAC CAC 1788 SLQQRL S Q N L 1802 Q

 5626
 CGC
 GAT
 GTC
 GCC
 ATC
 CGC
 GTG
 CAA
 CTG
 CGC
 CAA
 CGG
 CAA
 CGG
 56700

 1803
 R
 D
 V
 S
 A
 I
 R
 V
 M
 R
 Q
 L
 A
 Q
 R
 1817

CAA AAC GTG CCT GCC GTC CCG ATG CCC GTC GTC TTT ACC AGC GCG 5715 Q N V P A V P M <u>P V V F T S A</u> 1832 DR4 5671 1818 0 5716 CTG GGC TTT GAG CAG GAT AAC TTC CTC GCC CGG CGT AAT CTG CTC 1833 \underline{L} \underline{G} F \underline{E} Q D N F \underline{L} A R R N \underline{L} L 5760 AAG CCG GTC TGG GGC ATC TCC CAG ACG CCG CAG GTC TGG CTC GAT PVWGISQT DR5 1848 PQVWLD 1862 5850 1877 5806 1863 5851 TTT GTC GCC GCG CTG TTT CCT GCC GGG CAG GTG GAG CGC CAG TTT 5895 1878 A L FP A G 0 v E R 0 1892 5896 5940 1907 1893 AGC TGG CAA CTG CCG CTC GCC GCG CTG GTG CCT CCC GTA AAA CAC 5985 5941 1908 A 1922 5986 GCA GGG CAA TGC GCA GAG CGC CCA CCG CGC GTA TGC CCT GAG CAC 6030 1923 с A Е R P P 1937 TCT CAG CCA CAC ATT GCG GCG GAC GAG AGC ACC GTC AGC CTG ATT 6075 6031 1952 6076 TGC GAC GCC TTC CGC GAG GTG GTT GGC GAG TCT GTC ACA CCC GCA 6120 1953 F R E G Е s 1967 D 6121 GAA AAC TTC TTT GAG GCG GGT GCA ACA TCG CTG AAT CTG GTG CAA 6165 1968 Е A т s CTG CAC GTT TTG TTA CAA CGT CAC GAA TTT TCC ACC CTG ACG TTG 6210 6166 1983 L L Q н Е F т 1997 R s L 6211 6255 1998 2012 CTG GCC GGC GTC GCC ACG GTG GAG AAA ACA AAA CGT CCT CGC CCT 6300 2013 L A G v Е ĸ т ĸ P 2027 6301 GTT CGC CGT CGT CAG CGG CGG ATA TAG CGCGAAGCAAACTGATTTTCCCCCG 6351 2035 2028 RRQRRI 6352 GAACGCCATCGCGAACGCATGGCGTTCCATTGACTTTGTGAATCTTAGGAAACGGGACCG 6411 6412 ATTATGGATAACTTGCGCTTCTCTTCTGCGCCGACAGCAGATTCCATTGATGCATCG 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 data base 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 aminoproximal 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	SGAPSKDSLLPDNRHAADYQQLRERLIQELNLTPQQLHDESNLIQAGLDS	52
соон	1395	.:. :. ::!:. : . !: . .:::!. . KAENQAEADLPQGDIEKQVAALWQQLLSTGNVTRETD <u>FFQQGGDS</u> SBS	1439
	53	IRLMRWLHWFRKNGYRLTLRELYAAPTLAAWNQLMLSRSPENAEEETLPD	102
	1440	LLATRLTGQLHQAGYEAQLSDLFNHPRLADFAATLRKIDVPV	1481
	103	ESSWPNMTESTPFPLTPVQHAYLTGRMPGQTLGGVGCHLYQEFEGHCL	150
	1482	EQPFVHSPEERYQPFALTDVQQAYLVGRQPGFTLGGVGSHFFVEFEIADL	1531
	151	TASQLEQAITTLLQRHPMLHIAFRPDGQQVWLPQPYWNGVTVHDLRHNDA	200
	1532	DLTRLETVWNRLIARHDMLRAVVLDGQQQVLEQTPPW.VIPTHTLHTPE	1580
	201	ESROAYLDALRORLSHRLLRVEIGETFDFOLTLLPDNRHRLHVNIDLLIM	250
	1581	ALRVREKLAHQVLNPEVWPVFDLQVGYVDGMPARLWLCLDNLLL	1624
	251	DASSFTLFFDELNALLAGESLSAIDTRYDFRSYLLHQQKINQPLRDDARA	300
	1625	DGLSMQILLAELEHGYRYPQQLLPPLPVTFRDYL.QQPSLQSPNPDSLA	1672
	301	YWLAKASTLPPAPVLPLVCEFATLREVRNTRRMIVPATRWHAFSNRAGE	350
	1673	WWQAQLDDIPPAPALPLRCLEQEVETPRFARLNGALDSTRWHRLKKRAAD	1722
	351	YGVTPTMALATCFSAVLARWGGLTRLILNITLFDROPLHPAVGAMLADFT	400
	1723	AHLTPSAVLLSVWSTVLSAWSAQPEFTLNLTLFDRRPLHPDINQILGDFT	1772
	401	NILLLDTACDGDTVSNLARKNQLTFTEDWEHRHWSGVELLRELKRQQRYP	450
	1773	SIMLLSW.HPGESWLHSAQSLQQRLSQNLNHRDVSAIRVMRQLAQRQNVP	1821
	451	HGA.PVVFTSNLGRSLYSSRAESPLGEPEWGISQTPQVWIDHLAFEHHGE	499
	1822	AVPMPVVFTSALGFEQDNFLARRNLLKPVWGISQTPQVWLDHQVYESEGE	1871
	500	VWLQWDSNDALFPPALVETLFDAYCQLINQLCDDESAWQKPFADMMPA	547
	1872	LRFNWDFVAALFPAGQVERQFEQYCALLNRMAEDESSWOLPLAALVPP	1919

FIG. 3. Comparison of the amino acid repeats located at the NH_2 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 DNAbinding 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 δ -(L- α 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 (+ β -M) or absence (- β -M) of β -mercaptoethanol. Standard molecular weights (in thou-sands) 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 β -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 SGTTGXPKG, 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_i 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	DESSWPNMTESTPFPLTPVQHAYLTGRMPGQTLGGVGCHLYQEFEGHCLT	151
AngR	2	NONEHPFAFPETKLPLTSNONWOLSTORORTEKKSITNFTYOEFDYENIS	51
	152	ASQLEQAITTLLQRHPMLHIAFRPDGQQVWLPQPYWNGVTVHDLRHNDAE	201
	52	. : :::: :: : : : : . RDTLERCLTTIIKHHPIFGAKLSDDFYLHFPSKTHIETFAVNDLS	96
	202	SRQAYLDALRQRLSHRLLRVEIGETFDFQLTLLPDNRHRLHVN	244
	97	: : : . : :.: :.: . . NALKQDIDKQLADTRSAVTKSRSQAIISIMFSILPKNIIRLHVR	140
	245	IDLLIMDASSFTLFFDELNALLAGESLSAIDTRYDFRSYLLHQQKINQPL	294
	141	FNSVVVDNPSVTLFFEQLTQLLSGSPLSFLNQEQTISAYNHKVNNEL	187
	295	. RDDARAYWLAKASTLPPAPVLPLVCEPATLREVRNTRRRMIVPATRWH	342
	188	LSVDLESARWNEYILTLPSSANLPTICEPEKLDETDITRRCITLSQRKWQ	237
	343	AFSNRAGEYGVTPTMALATCFSAVLARWGGLTRLLLNITLFDRQPLHPAV	392
	238	QLVTVSKKHNVTPEITLASIFSTVLSLWGHQKYLMMRFDITKINDY	283
	393	GAMLADFTNILLLDTACDGDTVSNLARKNOLTFTEDWEHRHWSGVE	438
	284	.:::: : :: :: .:. . . .:: . : TGIIGQFTEPLLVGMSGFEQSFLSLVKNNQKKFEEAYHYDVKVPVFQCVN	333
	439	LLRELKRQQRYPHGAPVVFTSNLGRSLYSSRAESPLGEPEWGISQTPQVW	488
	334	.::. : :. .!: .: .: KLSNISDSHRYPANITFSSELLNTNHSKKAVWGCRQSANTW	374
	489	IDHLAFEHHGEVWLQWDSNDALFPPALVETLFDAYCQLINQLCDDESAWQ	538
	375	I. I. IIII IIIIIIIIIIIIIIIIIIIIIIIIIII	424
	539	${\tt KPFADMMPASQRAIRERVNATGA.PIPEGLLHEGIFRIALQQPQALAVTD}$	587
	425	QPLPTLLPKHQESIRNKINQQGDLELTKELLHQRFFKNVESTPNALAIIH	474
	588	MRYQWNYHELTDYARRCAGRLVECGVQPGDNVAITMSKGAGQLVAVLAVL	637
	475	GQESLDYITLASYAKSCAGALTEAGVKSGDRVAVTMNKGIGQIVAVLGIL	524
	638	LAGAVYVPVSLDQPAARREKIYADASVRLVLICQHDASAGSDDIPVLA	685
	525	YAGAIYVPVSLDQPQERRESIYQGAGINVILINESDSKNSPSNDLFFFLD	574
	686	WQQAIEAEPIVNPVVRAPTQPAYIIYTSGSTGTPKGVVISHRGALNTCCD	735
	575	WOTAIKSEPMRSPODVAPSOPAYIIYTSGSTGTFKGVVISHQGALNTCIA P-loop	624
	736	INTRYQVGPHDRVLALSALHFDLSVYDIFGVLRAGGALVMVMENQRRDPH	785
	625	: .:	674
	786	AWCELIORHOVTLWNSVPALFDMLLTWCEGFADATPENLRAVMLSGDWIG	835
	675	<pre> : : . : : . </pre>	724
	836	LDLPARYRAFRPQGQFIAMGGATEASIWSNACEIHDVPAHWRSIPYGFPL	885
	725	. .: .: .: .::: . : LDLPQRYRNYRVDGQFIAMGGATEASIWSNVFDVEKVPMEWRSIPYGYPL	774
	886	TNORYRVVDERGRDCPDWVSGELWIGGIGVAEGYFNDSLRSEQQFLTLPD	935
	775		824
	936	ERWYRTGDLGCYWPDGTIEFLGRRDKQVKVGGYRIELGEIESALSQLAGV	985
	825	. :	874
	986	KQATVLAIGEKEKTLAAYVVPQSEADPRQLQAALAGRLPGWM	1364
	875	<u>QRAVAIAVGNKDKTLAAFIVM</u> DSEQAPIVTAPLDAEEVQLLLNKQLPNYM helix-turn-helix	924
	1365	VPQRIVFLDALPLTANGKIDYQALKRR.HTPKAENQAEADLPQGDIEKQV	1413
	925	VPKRIIFLETFPLTANGKVDHKALTRMTNREKKTSQSINKPIITASEDRV 	974
	1414	AALWQQLLSTGNVTRETDFFQQGGDSLLATRLTGQLHQAGYEAQLSDLFN	1463
	975	AKIWNDVLGPTELYKSSDFFLSGGDAYNAIEVVKRCHKAGYLIKLSMLYR	102
	1465	HPRLADFAATLRKIDVPVEOP 1484	
	1025	YSTIEAFAIIMDRCRLAPQEE 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'-phosphopantheteine, 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'-phosphopantheteine (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 Methanothrix 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 \ddagger (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 (1 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, γ -(L- α -aminoadipyl)-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	mrygwnYhEL	tdyArrcAq	r LvecqVqp	GdnVAI	tMsK	GaggLvaV	LA VL	LAGAVYVP	VsldgPaaRr	(655)
AngR	(470)	LAIIh	ggesLdYitL	asvAkscAq	a LteagVks	GdrVAV	tMnK	GiggIvaV	LG IL	VAGAIYVP	VsldgPgeRr	(542)
TY1	(44)	VAIVf	enrrLSYqEL	NrkANgLAF	a LlekgVqt	DsiVGV	MEK	SienViaI	LA VLI	KAGGAYVP	IDIeYPrdRI	(116)
GS1	(56)	VAIVc	eneqLTYhEL	NvkANqLAR	i fiekgIgk	DtlVGI	mMEK	SidlfigI	LA VLI	KAGGAYVP	IDIeYPkeRI	(128)
GS2.dA	(485)	VAVVf	edekvTYrEL	herSNqLAF	f LrekgVkk	esiIGI	mMER	SvemIvgI	LG ILI	KAGGAÍVP	IDPeYPkeRI	(557)
GS2.dB	(1524)	VAVgw	kdqtLTYrEL	NerANqvAF	v LrqkgVqp	DniVGL	IVER	SpemLvgI	MG ILI	KAGGAY1P	LDPeYPadRI	(1596)
GS2.dC	(2560)	IAVId	erekLSYqEL	NakANqLAF	v LrqkgVqp	NsmVGI	mVDR	SldmIvgm	LG VLI	KAGGAYVP	IDIdYPqeRI	(2632)
GS2.dD	(3607)	IAIVw	egqaLiYhEL	NikANqLAF	v LrekgVtp	NhpVAI	mtER	SlemIvqI	Fs IL	KAGGAYVP	IDPaYPqeRI	(3679)
ACVS.asp.dC	(2509)	aAVVq	gdksLSYtEL	NKTANQLAF	y Iq.svahLrp	DdkV1L	1LDK	Sidmiidi	LA IW	KTGSAYVP	LDPsYPkeRV	(2582)
ACVT.pen.dC	(2485)	IAVVQ	gdraLSYaDL	NGQANQLAR	y iq.svsciga	DagIAL	MLEK	Sidtlidi.	LA IWI	KAGAAYVP	LDPtYPpgRV	(2558)
ACVS.cep.dC	(2452)	TATAG	GUISLOISEL advel TV-FI	NerangLVH	1 11.SSASIVA	DariaL		Sidmvial	LA VWI	KAGAAIVP	LDPTIPSQRT	(2525)
ACVS.noc.dc	(2410)	VAVVI	gavilitel egrel TVral	NerANTLAR	n Li.Svaepia a Ik ediepko	NeiTAL	WUDK	Selutival	LA VWI	KTCCA VUD	IDPSIPAARI	(2489)
ACVT pen dB	(1402)	TAVVY	equality a	Neranrar	q Lr. sdyspap	NevIAL	WMDK	SehmiyuT	LA VWI	KSCCAVUD	IDPerPadRI	(1301)
ACVS.cep.dB	(1378)	VAVVV	edirLTYrEL	NerANaLAf	v Ll.scaaloo	NETTE	IMDK	Sehmitsi	LA VW	KTGGAYVP	IDPryPdoRT	(1451)
ACVS.noc.dB	(1362)	IAVVV	renrLTYrEL	NerANrLAB	v Lr.svveLrp	DdlVAL	VLDK	SelmItal	i A aW	KTGAAYVP	TDsgYPddBT	(1435)
ACVS.asp.dA	(336)	VAIVV	krrgLTYgEL	NagANcFAB	v Lr.,sigIlp	eglVAL	fLEK	Senlivui		KSGAAYVP	IDPLYPdeRV	(408)
ACVT.pen.dA	(309)	IAVVc	dereLTYqEL	NaggNsLAF	y LrsigIlp	eqlVAL	fLDK	Seklivul	LG VW	KSGAAYVP	IDPtYPdeRV	(381)
ACVS.cep.dA	(275)	VALIC	gdkrITYeEL	NamANrLAH	h LvssgIqt	eqlVGL	fLDK	telmIatI	LG IW	KSGAAhVP	IDPqYPdeRV	(349)
ACVS.noc.dA	(267)	eAVVc	gdvrLTYrEV	NerANqFAH	w LiqgpvrVrp	GalIGL	yLDK	SdlgVvat	FG IW	KSGAAYVP	IDPaYPaeRI	(341)
EntF	(472)	pALAd	arylFSYrEm	reqVvaLAr	l LrergVkp	GdsVAV	aLpR	Svfltlal	hA Ive	eAGAAwlP	LDtgYPddRL	(544)
Lys2	(269)	• • • • •	sFTYrDi	NrtSNivAB	y LiktgIkr	GdvVmI	yssR	GvdlmvdV	MG VL	KAGAtfsv	IDPaYPpaRq	(333)
EntE	(48)	• • • • •	qLSYrEL	NqaAdnLAc	s LrrqgIkp	GetalV	qLgn	vaelyitf	FA 1LI	KlGVApVl	alfshqrseL	(112)
Luci	(50)	••••	nITYaEy	femSvrLAe	a mkrygLnt	NhrIvV	csEn	SlqffmpV	LG aL	fiGVAvaP	aNdiYnereL	(114)
4C11	(53)	••••	tFTYsqV	ellSrkvAs	g Lnk.lgIqq	GdtImL	lLpn	Speyffaf	LG asy	yrGAistm	aNPfftsaeV	(117)
4C1	(64)	••••	vLTYaDV	drlSrrLAa	a LrraplgLrr	GgvVms	lLrn	SpefVlsf	FA as	rvGAAvtt	aNPmstpheI	(130)
MenE	(19)	••••	tvTfaEL	faaSkrMAe	q LaahsVrk	GdtaAI	lLqn	raemVyaV	hA cf.	11GVkaV1	LNtklsther	(83)
Acua.met	(133)	••••	KITYGDL	ykevNkFAn	g Lk.sigLkk	GdrVsI	yMpm	ipqlpiam	LA cal	KIGVshiv	VFsgfsskgL	(197)
Acua.neu	(00)	• • • • •	NVIIGEL	1 TevskLAn	s sptwvirk	GOLVAI	yrbbw	ipealvam.	LA CU	CIGAINSV	VFagissdsL	(152)
Acua.asp	(127)	**	CITIGED	LIEVSIVA.	v DKqrgvkk	GULVAI	угърш	+	6 ++	+66 +	+ + +	(191)
Consensus		VAVV-	LTY-EL	NAN-LAR	- TV	GVAL	-1.0K	* STT	57.77 I.A. VWI	+ J J + H	TDP-YPRT	
compensas		ттт	TS DV	SMH			MFR		RG TI.	S C	LNT V	
			F E		* *			G V V		<u>-</u>		
		<u> </u>	<u> </u>	<u>x</u> <u>r</u>							V C 14	
					<u> </u>	R X		ر بر ا		· ±		
					4	N Y		- L		 vii		
					4	K Y		- -		VII		
HMWP2		(707)	AYI i.	 	TSGSTGtPK GV	N Y	. ¥ (728)	- -		VII amgGaTI	La (860)	
HMWP2 AngR		(707) (596)	AYI i. AYI i.	 	TSGSTGtPK GV TSGSTGtPK GV	visHrG visHqG	(728) (617)	-	(853) (742)	VII amgGaTI amgGaTI	Ea (860) Ea (749)	
HMWP2 AngR TY1		(707) (596) (173)	AYI i. AYI i. AYV i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGKPK GL	visHrG visHqG nleHkG	(728) (617) (194)	-	(853) (742) (308)	VII amgGaTl amgGaTl NaYGpTl	Ca (860) Ca (749) Ct (315)	
HMWP2 AngR TY1 GS1		(707) (596) (173) (185)	AYI i. AYI i. AYV i. AYV i.		TSGSTGtPK GV TSGSTGtPK GV TSGTTGkPK Gt TSGTTGnPK Gt	visHrG visHqG nleHkG nleHkG	(728) (617) (194) (206)		(853) (742) (308) (321)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI	Ea (860) Ea (749) Et (315) Et (328)	
HMWP2 AngR TY1 GS1 GS2.dA		(707) (596) (173) (185) (609)	AYI i. AYI i. AYV i. AYV i. fYI i.		TSGSTGtPK GV TSGSTGtPK GV TSGTTGkPK Gt TSGTTGnPK Gt TSGTTGkPK GV	visHrG visHqG nleHkG nleHkG nleHkN	(728) (617) (194) (206) (630)		(853) (742) (308) (321) (754)	VII amgGaTl amgGaTl NaYGpTl NaYGpTl NhYGpsl	Ea (860) Ea (749) Et (315) Et (328) Et (761)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB		(707) (596) (173) (185) (609) (1650)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI m.		TSGSTGtPK GV TSGSTGtPK GV TSGTTGkPK Gt TSGTTGnPK Gt TSGTTGkPK GV TSGSTGkPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN	(728) (617) (194) (206) (630) (1671	L)	(853) (742) (308) (321) (754) (1791)	VII amgGaTl amgGaTl NaYGpTl NaYGpTl NhYGpsJ) NgYGpTl	Ea (860) Ea (749) Et (315) Et (328) Et (761) En (1798)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC		(707) (596) (173) (185) (609) (1650) (2686)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI m. AYI i.	 У У У У У У У У У У У У У У У У У	TSGSTGLPK GV TSGSTGLPK GV TSGTTGKPK GT TSGTTGNPK GT TSGTTGNPK GV TSGSTGKPK GV TSGSTGKPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS	(728) (617) (194) (206) (630) (1671 (2707		(853) (742) (308) (321) (754) (1791) (2835)	VII amgGaTl amgGaTl NaYGpTl NaYGpTl NhYGpsi) NgYGpTl) NsYGvTl	Ea (860) Ea (749) Et (315) Et (328) Et (761) En (1798) Ea (2842)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD		(707) (596) (173) (185) (609) (1650) (2686) (3733)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i.	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS	(728) (617) (194) (206) (630) (1671 (2707 (3754	L))))	(853) (742) (308) (321) (754) (1791) (2835) (3876)	VII amgGaTl amgGaTl NaYGpTl NaYGpTl NhYGpsl NgYGpTl NsYGvTl NsYGvTl	Ca (860) Ca (749) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2612)	AYI i. AYI i. AYV i. AYV i. AYI i. AYI i. AYI i. AYI i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQgG	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (3876) (2781)	VII amgGaTi amgGaTi NaYGpTi NhYGpsi NhYGpsi NSYGvTi NeYGpTi NaYGiTi	Ca (860) Ca (749) Ct (315) Ct (315) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2788)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC ACVT.pen.dC		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2596)	AYI i. AYV i. AYV i. AYV i. AYI i. AYI i. AYI i. AYI i. AYI i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTSGNPK GV TSGTSGNPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQG lveQka	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2662 (2662	L) L) 2) 2) 2) 2)	(853) (742) (308) (321) (754) (1791) (2835) (3876) (3876) (2731) (2757)	VII amgGaTi amgGaTi NaYGpTi NaYGpTi NhYGpsi NgYGpTi NsYGvTi NaYGTI NaYGTI NaYGTI	Ca (860) Ca (749) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2788) Ct (2764)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC ACVS.cep.dC ACVS.cep.dC		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2641) (2586) (2545)	AYI i. AYV i. AYV i. AYV i. AYI i. AYI i. AYI i. AYI i. AYI i. AYV i. AYV i. AYA i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTSCNPK GV TSGTTGNPK GV	visHrG visHqG nleHkG nleHkM nleHkN nieHqS nieHqS lveQgG lveQga lveHqS lveHqS	(728) (617) (194) (206) (1671 (2707 (3754 (2662 (2662 (2662 (2665	L) 	(853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2785) (2725) (2685)	VII amgGaTH amgGaTH NaYGpTH NhYGpsH NhYGpTH NAYGPTH NAYGTTH NAYGTTH NAYGTTH NAYGTTH NAYGTTH	Line Line Ca (860) Ca (749) Cit (315) Cit (328) Cit (761) Cit (761) Cit (2842) Cit (3883) Cit (2788) Cit (2764) Cit (2732) Cit (2732)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB		(707) (596) (173) (1850) (2686) (3733) (2641) (2617) (2586) (2545) (1561)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYZ i. AYZ i. AYZ i.	Y	TSGSTGtPK GV TSGSTGtPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTSGKPK GV TSGTTGKPK GV TSGTTGKPK GV	visHrG visHqG nleHkG nleHkG nleHkN nieHqS nieHqS lveQgG lveQka lveHqS lvsHqS nveHkG	(728) (617) (194) (206) (1671 (2707 (3754 (2662 (2662 (2662 (2666 (2566 (2566	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2781) (2785) (2785) (2685) (2685)	VII amgGaTl amgGaTl NaYGpTl NhYGpsl NhYGpsl NsYGvTl NaYGTl NaYGTTl NaYGTTl NaYGTTl NaYGTTl NaYGTTl	Line Line Ca (860) Ca (749) Cit (315) Cit (315) Cit (328) Cit (761) Cin (1798) Cit (3883) Cit (2788) Cit (2764) Cit (2709)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.pen.dB		(707) (596) (173) (185) (609) (1650) (2686) (2733) (2641) (2617) (2586) (2545) (1561) (1535)	AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYZ i. AYZ i. AYI i.		TSGSTGtPK GV TSGSTGtPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTSGRPK GV TSGTTGkPK GV TSGTTGkPK GV TSGTTGkPK GV	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQgG lveQka lveHqS nveHkG veHkG	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2638 (2662 (2638 (2662 (1558)	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2781) (2785) (2785) (2725) (2685) (1702) (1676)	VII amgGaTl amgGaTl NaYGpTl NhYGpsl NhYGpsl NsYGvTl NaYGvTl NaYGTl NaYGTTl NaYGTTl NaYGTTl NgYGpTl NgYGpTl	Ea (860) Ea (749) Et (315) Et (315) Et (761) En (1798) Ea (2842) En (3883) Et (2788) Et (2788) Et (2788) Et (2732) Et (2692) Et (1709) Et (1709)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.asp.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dC ACVS.asp.dB ACVT.pen.dB ACVS.cep.dB		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2617) (2641) (2546) (2545) (1561) (1535)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYA i. AYI i. AYI i. AYI i. AYI m.		TSGSTGtPK GV TSGSTGtPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTSGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV	visHrG visHrG nleHkG nleHkG nleHkN nveHrN nieHqS lveQgG lveQgG lveHqS lvsHgS nveHhG cveHhG nveHhG	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2638 (2662 (1556 (1553)	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2781) (2781) (2785) (2785) (2725) (2685) (1702) (1676) (1656)	VII amgGaTI amgGaTI NaYGpTI NhYGpSJ NgYGpTI NsYGvTI NaYGITI NaYGITI NaYGTI NgYGpTI NgYGpTI	Ea (860) Ea (749) Et (315) Et (315) Et (328) Et (761) En (1798) Ea (2842) En (3883) Et (2788) Et (2764) Et (2732) Et (2692) Et (1709) Et (1683) Ev (1687)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.asp.dC ACVS.noc.dC ACVS.noc.dC ACVS.ep.dB ACVS.cep.dB ACVS.noc.dB		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2617) (2617) (2586) (2545) (1561) (1553) (1509)	AYI i. AYV i. AYV i. AYV i. AYI i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGKPK GT TSGTTGNPK GT TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTSGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV	visHrG visHrG nleHkG nleHkG nleHkN nveHrN nieHqS nveHqS lveQgG lveQgG lveHqS lveHqS nveHhG veHhG nveHhG	(728) (617) (194) (206) (630) (1671) (2707 (2662 (2662 (2662 (2662 (2566 (1552 (1556 (1554	L))))))))))))))	(853) (742) (308) (321) (754) (1791) (2835) (2757) (2725) (2785) (2785) (2785) (2785) (1792) (1676) (1654)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NSYGVTI NSYGVTI NAYGTI NAYGITI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI	La (860) Ca (749) Ca (749) Ct (315) Ct (315) Ct (328) Ct (315) Ct (328) Ct (2842) Cn (3883) Ct (2764) Ct (2764) Ct (2692) Ct (1633) Cv<(1683)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACV5.asp.dC ACV5.asp.dC ACV5.noc.dC ACV5.noc.dC ACV5.pen.dB ACV5.pen.dB ACV5.noc.dB ACV5.noc.dB ACV5.noc.dB		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2641) (2617) (2586) (2545) (1561) (1561) (1509) (1493) (1493)	AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI m. AYI m. AYI i. AYX i. AYX i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GT TSGTTGNPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTSGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV	visHrG visHrG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQG lveQta lveQta lveQta lveHhG tveHhG tveHhG LveHTN	(728) (617) (194) (206) (206) (206) (2707 (2662 (2662 (2662 (2662 (2662 (2662 (2566 (1582 (1556 (1556 (1556) (1516) (496)	L))))))))))))))	(853) (742) (308) (321) (754) (1791) (2835) (2757) (2757) (2757) (2757) (2757) (2757) (2757) (2757) (2757) (2757) (2757) (1776) (1676) (1655)	VII amgGaTI amgGaTI NaYGpTI NhYGpJI NSYGPTI NSYGVTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI	Line Line Ca (860) Cine Ca (749) Cine Ca (315) Cine Cine (315) Cine (761) Cine (1798) Cine (2842) Cine (2788) Cine (2788) Cine (2764) Cine (2692) Cine (1683) Cine (1657) Cine (1641) Cine (2622)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC ACVS.noc.dC ACVS.noc.dC ACVS.ep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVT.pen.dA		(707) (596) (173) (185) (2686) (2686) (2617) (2617) (2586) (2545) (1561) (1535) (1509) (1493) (448)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI m. AYI i. AYI t. AYV t.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTSGKPK GV TSGTSGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGFPK GI	visHrG visHqG nleHkG nleHkG nleHkM nveHrN nieHqS nieHqS lveQgG lveQgA lveQgA lveQkA lveHqS lveHhG tveHhG tveHhG LveHrG lkqHtN fkqHtN	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2632 (2632 (1556 (1556 (1556 (1556 (1556) (1514) (469)	L) 7) 3) 5) 5) 5) 1)	(853) (742) (308) (321) (754) (1791) (2835) (3876) (2725) (2725) (2785) (2725) (2685) (1676) (1650) (1634) (615)	VII amgGaTI amgGaTI NaYGpTI NAYGpJI NSYGVTI NSYGVTI NAYGITI NAYGITI NAYGITI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NGYGFTI	La (860) Ca (749) Ca (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2768) Ct (2732) Ct (2692) Ci (1709) Cv (1683) Cv (1657) Ci (1641) Ci (2692) Ci (1641) Ci (595)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB ACVS.cep.dB ACVS.noc.dB ACVS.pen.dA		(707) (596) (173) (185) (609) (1650) (2686) (2617) (2586) (2545) (1561) (1535) (1509) (1493) (1493) (475) (448) (418)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYV i. AYV i. AYI i. AYV t. AYV t.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GT TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GI	visHrG visHqG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQgG lveQka lveQgS lveQka lveHqS lveHqS lveHhG tveHhG LveHrG lkqHtN ykeHtS	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2662 (2666 (1582 (1582 (1582 (1586) (1582 (1594) (469) (439)	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (2781) (2781) (2781) (2785) (1702) (1676) (1650) (1650) (1651) (588) (558)	VII amgGaTI amgGaTI NaYGpTI NAYGpJI NSYGPTI NSYGVTI NAYGITI NAYGITI NAYGITI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NEYGFTI	La (860) Ca (749) Ca (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2732) Ct (2732) Ct (2732) Ct (2692) Ci (1709) Cv (1683) Cv (1657) Ci (1641) Cs (525) Cs (555)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.noc.dB ACVS.noc.dA		(707) (596) (173) (185) (16609) (1650) (2686) (2617) (2586) (2545) (1561) (1535) (1509) (1493) (1475) (448) (418) (405)	AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYV i. AYI i.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGRPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI	visHrG visHrG visHqG nleHkG nleHkG nleHkM nveHrN nieHqS nveQgG lveQgG lveQgG lveQgG lveQgG lveHqS lveHqS lveHhG cveHhG nveHhG kqHtN kqHtN ykeHtS pkyHyS	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2662 (1582 (1582 (1582 (1582 (1584 (496) (439)) (426)	L) 7) 1) 2) 3) 7) 5) 5) 5)	(853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2785) (2785) (1702) (1676) (1650) (1650) (1634) (615) (588) (558) (558)	amgGaTI amgGaTI NaYGpTI NaYGpTI NAYGpJI NSYGPTI NSYGTI NAYGTI NAYGTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGFTI NGYGFTI NEYGTTI NEYGTTI	La (860) Ca (749) Ca (315) Ct (328) Ct (749) Ca (860) Ct (328) Ct (749) Ca (842) Cn (3883) Ct (2788) Ct (2732) Ct (2732) Ct (2692) Ct (1709) Ev (1683) Ev (1657) Ci (1641) Cs (525) Cs (555) Ca (552)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.ssp.dB ACVS.pen.dB ACVS.pen.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA		(707) (596) (173) (185) (1609) (1650) (2686) (3733) (2641) (2586) (1561) (1535) (1561) (1535) (1509) (1493) (475) (448) (418) (405) (598)	AYI i. AYV i. AYV i. AYV i. fyI i. AYI t. AYU t. AYV t. AYV t. AYV t. AYI i.	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GI TSGTTGFPK GI TSGTTGFPK GI	visHrG visHqG nleHkG nleHkG nleHkN nieHqS nveQgG lveQgG lveQgG lveQgA lveHqS lveHqS lveHqS lveHhG tveHhG LveHtS kqHtN fkqHtN fkqHtS pkyHyS nvgQta	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2662 (1582 (1556 (1582 (1556 (1582 (1514 (496)) (426) (439) (426) (619)	L) () () () () () () () () () ((853) (742) (308) (321) (754) (2751) (2751) (2735) (2781) (2785) (2785) (2785) (1634) (1650) (1634) (1658) (558) (558) (558) (545)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpsI NSYGvTI NSYGvTI NAYGTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NEYGFTI NEYGFTI NEYGFTI	La (860) Ca (749) Ca (315) Ct (328) Ct (749) Ct (328) Ct (749) Ca (860) Ca (749) Ct (328) Ct (761) Ca (2842) Ca (3883) Ct (2764) Ct (2732) Ct (2692) Ct (2762) Ct (1683) Cv (1657) Ct (1657) Ct (1641) Cs (595) Cs (552) Ca (552) Ca (552)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVT.pen.dB ACVS.noc.dB ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.sasp.dA ACVS.soc.dA EntF Lys2		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2586) (2545) (1561) (1535) (1535) (1493) (475) (448) (405) (598) (407)	AYI i. AYV i. AYV i. fYI i. AYI i. AYU t. AYV t. AYV t. AYU t.	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV	visHrG visHqG nleHkG nleHkG nleHkN nieHqS nieHqS lveQgG lveQgG lveQka lveHqS lveHqS lveHqS lveHhG LveHtG lkqHtN fkqHtN fkqHtN fkqHtS pkyHyS nvgQta lgrHfS	(728) (617) (194) (206) (630) (1671) (2707 (3754 (2662 (2662 (2566 (1582 (1556 (1582 (1556 (1514 (496)) (426) (426) (426) (437)	L)))))))))))))))))	(853) (742) (308) (321) (754) (2835) (32791) (2757) (2781) (2781) (2785) (2785) (2685) (1634) (1650) (1658) (558) (558)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpSJ NSYGVTI NAYGTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NEYGFTI NEYGFTI NEYGFTI NEYGFTI NIYGPTI	La (860) Ca (749) Ca (749) Ca (749) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2788) Ct (2782) Ct (2692) Ct (2692) Ct (2692) Ct (2692) Ct (2692) Ct (2692) Ct (1683) Cv (1683) Cv (1641) Cs (565) Ca (552) Ca (751) Ct (565)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.sp.dB ACVT.pen.dB ACVT.pen.dB ACVS.ep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2586) (2545) (1501) (1535) (1509) (1493) (475) (448) (418) (405) (598) (407)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI t. AYV t. AYI i.	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPFK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV	visHrG visHrG nleHkG nleHkG nleHkK nveHrN nieHqS lveQgG lveQgG lveHqS lveHqS lveHG LveHG LveHG LveHG LveHrG lkqHtN fkqHtN fkqHtN gkeHtS okyHyS nvgCta lqrHfS	(728) (617) (194) (206) (630) (1671 (2707 (2707 (2707 (2707 (2707 (2566 (1552 (1553 (1554 (1553) (1514 (496)) (437) (437)	L)))))))))))))))	(853) (742) (308) (754) (1791) (2835) (2751) (2751) (2751) (2752) (2685) (1702) (1650) (1654) (1658) (588) (545) (744) (558)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NsYGvTI NsYGvTI NaYGTI NaYGITI NaYGTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NeYGfTI NeYGfTI NeYGfTI NeYGfTI NeYGfTI NiYGpTI	La (860) Ca (749) Ca (315) Ct (315) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2788) Ct (2764) Ct (2692) Ct (2692) Ct (1663) Cv (1657) Ct (1641) Cs (525) Ca (552) Ca (552) Ca (552) Ca (565) Ca (565)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2617) (2586) (2545) (1501) (1535) (1535) (1535) (1493) (475) (448) (407) (192) (192)	AYI i. AYI i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI t. AYV t.	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GI TSGTTGPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTSGPK GV TSGTSGPK GV	visHrG visHrG visHqG nleHkG nleHkK nveHrN nieHqS nveHrS lveQqG lveQqG lveQta lveHqS lveHqS lveHqS lveHG LveHrG lkqHtN fkqHtN fkqHtN ykeHS pkyHyS nvgtha lqrHfS	(728) (617) (194) (206) (630) (1671 (2707 (2707 (2707 (2707 (2707 (2707 (2707 (2707 (2707) (266) (1514 (1514 (1514 (1514 (1514)	L))))))))))))))	(853) (742) (308) (321) (1754) (1754) (1754) (2757) (2757) (2787) (2787) (2787) (2787) (2787) (2787) (2787) (2787) (2787) (2785) (1658) (588) (545) (545) (545) (545) (543) (558)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpsJ NSYGVTI NSYGVTI NAYGITI NAYGITI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NEYGFTI NEYGFTI NHYGTI NHYGTI	La (860) Ca (749) Ca (749) Ca (749) Ct (315) Ct (315) Ct (315) Ct (328) Ct (761) Cn (3883) Ct (2788) Ct (2764) Ct (2764) Ct (2764) Ct (2762) Ct (2692) Ct (2692) Ct (2692) Ct (1683) Cv (1657) Ct (1641) Cs (565) Ca (552) Ca (552) Ca (555) Ca (337) Ct (565)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2617) (2586) (2545) (1561) (1535) (1535) (1535) (1493) (475) (448) (418) (407) (598) (407) (172) (182)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI i. AYI t. AYV t.		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GT TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GI TSGTTGFPK GI	visHrG visHrG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQgG lveQgG lveHqS lveHqS lveHG tveHhG tveHhG tveHhG lkqHtN fkqHtN fkqHtN ykeHtS pkyHyS nvgQta lgrHfS prtHnd alpHrt nltHr	(728) (617) (194) (206) (630) (1671 (2707 (2707 (2707 (2707 (2707 (2707 (2662 (1532 (1556 (1532 (1556 (1534 (1534 (1534) (469)) (437) (437) (206) (214) (206)	L))))))))))))))	(853) (742) (308) (321) (1754) (1791) (2835) (2757) (2757) (2787) (2777)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGp31 NsYGvTI NsYGvTI NaYGTI NaYGTI NaYGTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NeYGfTI NeYGfTI NeYGfTI NhYGTI NhYGTI NmYGtTI QYfGmal qgYGITI	La (860) Ca (749) Ca (15) Ct (315) Ct (315) Ct (315) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2764) Ct (2764) Ct (2692) Ct (1683) Cv (1657) Ct (1641) Ca (552) Ca (552) Ca (751) Ct (565) Ca (337) Ct (345)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.asp.dC ACVS.roc.dC ACVS.roc.dC ACVS.roc.dC ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 4C1		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2641) (2641) (2586) (2545) (1561) (1509) (475) (448) (418) (407) (407) (172) (184) (172) (184) (176)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI i. AYV t. AYV t. AYI i. AYI i. AYV t. AYI i. AYI i. AYV t. AYV t. AYI i. AYI i. AYV t. AYI i. AYI i. AYV t. AYV t. AYI i. AYI i. AYI i. AYV t. AYV t. AYI i. AYI i. AYI i. AYV t. AYV t. AYV t. AYV t. AYI i. AYI i. AYI i. AYV t. AYV AYV	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GT TSGTTGNPK GT TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GV SSGTTG1PK GV	visHrG visHrG visHgG nleHkG nleHkM nveHrN nieHqS nieHqS lveQgG lveQda lveQda lveHhG lveHhG lveHhG lveHtS nveHhG lkqHtN ykeHtS pkyHyS nvgQta lgrHfS prtHnd alpHrt nltHcs	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (1532 (1556 (1532 (1556 (1531 (1556 (1531 (1556) (1531 (1556) (1531) (439) (439) (439) (437) (206) (214) (206) (225)	L))))))))))))))	(853) (742) (308) (321) (1754) (1791) (2835) (2835) (2725) (2725) (2725) (2725) (2725) (2725) (2725) (2725) (2725) (2758) (1702) (1656) (558) (558) (558) (558) (558) (558) (330) (338) (332) (351)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpJI NSYGVTI NSYGVTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NGYGFTI NEYGFTI NEYGFTI NHYGFTI NHYGTI NHYGTI QYGGMAJ GGYGMTI	Li Li Ca (860) Ca (749) Ca (15) Ct (315) Ct (315) Ct (328) Ct (761) Cn (1798) Ca (2842) Cn (3883) Ct (2764) Ct (2764) Ct (2764) Ct (2692) Ct (16709) Ct (2692) Ct (1677) Ct (1641) Cs (555) Ca (552) Ca (555) Ca (565) Ca (337) Ct (339) Ca (339) Ca (359)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.asp.dC ACVS.noc.dC ACVS.noc.dC ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 4C1 MenE		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2641) (2647) (2586) (2545) (1561) (1509) (1493) (475) (448) (405) (598) (407) (172) (184) (177) (196)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI m. AYI i. AYI m. AYV t. AYV t. AYV t. AYV vgp edftatp spa esf dro kvf dpo	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGRPK GT TSGTTGRPK GT TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTSGRPK GV TSGTSGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGSTGFPK GV TSGSTGFPK GV SSGTTGPK GV SSGTTG1PK GV SSGTTG1PK GV	visHrG visHrG visHqG nleHkG nleHkM nveHrN nieHqS nieHqS lveQga lveQga lveQta lveQta lveHhG tveHhG tveHhG tveHhG lkqHtN ykeHtS pkyByS nvgta lgrHfs ortHnd alpHrt nltHrS motfn%	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (1574 (2662 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582) (426) (437) (437) (437) (206) (214) (206) (225) (160)	L) 7) 3) 5) 5) 5)	(853) (742) (308) (321) (1791) (2835) (2835) (3876) (2737) (2757)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpJI NSYGvTI NSYGvTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NEYGFTI NEYGFTI NEYGFTI NHYGFTI NHYGTI NHYGTI NHYGTI NHYGTI QYFGMAI QGYGMTI QYGMTI	La (860) Ca (749) Ct (315) Ct (328) Ct (328) Ct (328) Ct (383) Ca (2784) Ct (2782) Ct (2783) Ct (2764) Ct (2692) Ct (16732) Ct (2692) Ct (1677) Ct (1657) Ct (1657) Ct (555) Ca (555) Ca (565) Ca (337) Ct (345) Ca (358)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.cep.dB ACVS.cep.dB ACVS.noc.dC ACVS.noc.dA ACVS.noc.dA ACVS.noc.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 MenE Acua.met		(707) (596) (173) (185) (2686) (2686) (2641) (2617) (2546) (2545) (1561) (1561) (1509) (1493) (475) (448) (418) (407) (1984) (407) (172) (184) (177) (184) (177) (1961) (140) (270)	AYI i. AYI i. AYV i. AYV i. fYI i. AYI i. AYV t. AYV AYV t. AYV AYV t. AYV AYV	Y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTSGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGPK GV TSGTTGLPK GV SSGTTGLPK GV TSGTTGRPK GV	visHrG visHrG visHqG nleHkG nleHkG nleHkM nveHrN nieHqS nieHqS lveQgG lveQgA lveQgA lveHqS lveHgS lveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG treHhG nthHS algHHT	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (1574 (2662 (1582 (1556 (1582 (1556 (1582 (1556) (1594) (469) (469) (437) (437) (206) (214) (206) (225) (160) (304)	L) (L) (L) (L) (L) (L) (L) (L) (L	(853) (742) (308) (321) (754) (1791) (2835) (3876) (2787)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NsYGpTI NsYGvTI NaYGTI NaYGTI NaYGTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NeYGfTI NeYGfTI NeYGfTI NeYGfTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI NaYGTI	La (860) Ca (749) Ca (315) Ct (328) Ct (328) Ct (328) Ct (383) Ca (2842) Cn (3883) Ct (2788) Ct (2788) Ct (2692) Ci (1709) Cv (1683) Cv (1657) Ci (1641) Cs (555) Ca (552) Ca (555) Ca (337) Ct (345) Ca (358) Ct (345)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.asp.dB ACVS.noc.dB ACVS.noc.dB ACVS.noc.dA BACVS.noc.dA EntF Lys2 EntE Luci 4C11 MenE Acua.met Acua.neu		(707) (596) (173) (185) (2686) (2686) (2545) (1561) (1535) (1561) (1535) (1561) (1535) (1493) (1493) (448) (418) (405) (598) (407) (172) (184) (177) (184) (177) (196) (140) (140) (270) (225)	AYI i AYI i AYV i fYI i AYI m. AYI i AYI i AYV t AYV t AYV t AYV t AYI i GVV vgl edftatp spi esf dro vvI nso kvf dpc 	Adevayfq 1 ddvval.p Y her y he	TSGSTGLPK GV TSGSTGLPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGRPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV	visHrG visHrG visHqG nleHkG nleHkG nleHkG nieHqS nieHqS nieHqS lveQgG lveQgG lveQgA lveHqS lveHqS lveHqS lveHhG tveHhG tveHhG tveHhG tveHhG lkqHtN ykeHtS pkqHtN ykeHtS prtHnd alpHrt nltHkG nltHrS qqffgN ahaTqG	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2632) (2632 (2632 (2632 (1553) (1532) (1556) (1532) (1556) (1532) (1556) (1532) (1556) (469) (437) (426) (437) (206) (214) (206) (225) (160) (259)	L) (L) (L) (L) (L) (L) (L) (L) (L	(853) (742) (308) (321) (754) (1791) (2835) (3876) (2781) (2785) (2785) (2785) (1676) (1676) (1650) (1676) (1650) (1676) (1658) (558) (558) (558) (558) (558) (330) (332) (331) (435) (386)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NsYGvTI NsYGvTI NaYGTI NaYGITI NaYGTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NeYGfTI NeYGfTI NeYGfTI NeYGfTI NeYGfTI NgYGpTI NgYGpTI NgYGpTI NeYGfTI NeYGfTI NgYGTI NgYGTI NgYGTI NgYGTI NgYGMI Common State NgYGMI NgYGMTI GyYGMI Common State NgYGMTI Common State NgYGMTI NgYGMTI Common State NgYGMTI NgYGMTI	La (860) Ca (749) Ca (315) Ct (328) Ct (328) Ct (328) Ct (383) Ca (2842) Cn (3883) Ct (2788) Ct (2764) Ct (2732) Ct (2692) Ct (1607) Ct (2692) Ct (1657) Ct (1657) Ct (1641) Ct (555) Ca (555) Ca (555) Ca (337) Ct (345) Ca (339) Ca (358) Ct (442) Ct (393)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.noc.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 4C1 MenE Acua.met Acua.neu Acua.asp		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2586) (2545) (1561) (1561) (1535) (1509) (1493) (475) (448) (448) (405) (598) (407) (172) (184) (177) (196) (140) (225) (225) (2264)	AYI i. AYV i. AYV i. AYV i. fYI i. AYI i. AYU t. AYV t. AYU t. GvV vg eefftatp space evor	y y y y y y y y y y y y y y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGPK GV TSGSTGIPK GV TSGTTGLPK GV TSGTTGLPK GV TSGTTGLPK GV TSGTTGNPK GV	visHrG visHrG visHqG nleHkG nleHkG nleHkG nleHkG nieHqS lveQgG lveQgG lveQgA lveHqS lveHqS lveHqS lveHgS nveHhG cveHhG tveHG kqHtN ykeHtS pkqHtN ykeHtS prtHnd alpHrt nltHkG nltHrS qqffgN ehaTgG ahsTqG nhtTaG	(728) (617) (194) (206) (630) (1671 (2707 (3754 (2662 (2662 (2662 (1582 (1582 (1582 (1582 (1582 (1584) (469) (439) (426) (437) (426) (437) (206) (214) (206) (225) (160) (225) (160) (2298)	L) 7) 1) 2) 3) 7) 5) 5) 2)	(853) (742) (308) (321) (754) (1791) (2835) (2781) (2785) (1792) (2785) (1676) (1676) (1650) (1676) (1658) (558) (558) (558) (558) (558) (558) (558) (330) (338) (332) (335) (386) (428)	amgGaTI amgGaTI amgGaTI NaYGpTI NaYGpTI NSYGVTI NSYGVTI NSYGVTI NAYGITI NAYGITI NAYGITI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NGYGFTI NEYGFTI NEYGFTI NHYGFTI NHYGFTI NHYGFTI NHYGFTI GYYGMAI GYYGMI GYYGMI CHYMTI dtYWqTI	La (860) Ca (749) Ca (749) Ct (315) Ct (315) Ct (315) Ct (328) Ct (761) Cn (1798) Ct (2788) Ct (2788) Ct (2782) Ct (2692) Ct (2692) Ct (2692) Ct (1683) Cv (1683) Cs (565) Ca (552) Ca (371) Ct (345) Ca (339) Ca (358) Ct (442) Ct (435)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dB ACVS.noc.dC ACVS.noc.dB ACVS.noc.dB ACVS.noc.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 4C1 MenE Acua.met Acua.asp		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2586) (2545) (1561) (1535) (1535) (1535) (1493) (475) (448) (407) (172) (184) (177) (184) (177) (196) (140) (225) (264)	AYI i. AYV i. AYV i. AYV i. AYI i. AYU t. AYV t. AYV t. AYU t. GvV vgv eefftatp space eecepvw vdg AYtapds vng	y y y y y y y y y y y y y y	TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGFPK GV TSGTTGPK GV	visHrG visHrG visHqG nleHKG nleHKG nleHKG nleHKG nieHqS lveQgG lveQgG lveQgG lveQgG lveQgG lveHqS lveHqS lveHqS lveHG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG tveHhG nveHhG tveHhG nveHhG tveHf shgHtN shaTgG ahsTgG ahsTgG sh	(728) (617) (194) (206) (630) (1671 (2707 (2707 (2707 (2707 (2662 (1552 (1553 (1514 (496)) (426) (437) (437) (206) (214) (206) (225) (160) (225) (160) (259) (298)	L) () () () () () () () () () ((853) (742) (308) (321) (754) (1791) (2835) (2781) (2785) (2785) (1702) (1676) (1650) (1676) (1658) (558) (558) (558) (558) (558) (558) (330) (332) (335) (335) (386) (428)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpsJ NSYGVTI NSYGVTI NAYGTI NAYGTI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NGYGTI NGYGTI NGYGTI GYYGMI GYYGMTI GYYGMTI dtYWqTI dtYWqTI S SS	La (860) Ca (749) Ca (315) Ct (315) Ct (328) Ct (315) Ct (328) Ct (383) Ct (2788) Ct (2788) Ct (2782) Ct (2692) Ct (2692) Ct (2692) Ct (1657) Ct (1657) Ct (1641) Cs (555) Ca (555) Ca (751) Ct (345) Ca (339) Ca (358) Ct (442) Ct (435)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVT.pen.dB ACVS.noc.dB ACVS.noc.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C11 4C1 MenE Acua.met Acua.asp Consensus		(707) (596) (173) (185) (609) (1650) (2686) (3733) (2641) (2617) (2586) (2545) (1535) (1535) (1535) (1493) (475) (448) (407) (1493) (475) (1484) (177) (184) (177) (184) (177) (196) (140) (225) (264)	AYI i. AYV i. AYV i. AYV i. fyI i. AYI i. AYU t. AYV t. AYV t. AYV t. AYV t. AYV t. AYU t. AYU t. GvV vg eeftatp sp eesf dp ag AYIapds vns ## ag		TSGSTGLPK GV TSGSTGLPK GV TSGTTGNPK GV TSGTTGPK GV TSGTTGPK GV TSGTTGFPK GI TSGTTGFPK GV TSGTTGFPK GV TSGTTGPK GV TSGTTGPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGIPK GV TSGTTGPK GV TSGTTGPK GV TSGTTGPK GV TSGTTGPK GV	visHrG visHrG visHqG nleHkG nleHkKG nleHkK nveHrN nieHqS lveQqG lveQqG lveQtqS lveHqS lveHqS lveHqS lveHqS lveHqS lveHtqS lveH	(728) (617) (194) (206) (630) (1671 (2707 (2707 (2707 (2707 (2707 (2662 (1582 (1582 (1582 (1582 (1582 (1514 (496)) (426)) (426) (437) (449) (437) (206) (225) (206) (225) (160) (225) (298)	L))))))))))))))	(853) (742) (308) (754) (1791) (2835) (2751) (2751) (2751) (2781) (2781) (2781) (2781) (2781) (2781) (2781) (2782) (1650) (16534) (1658) (545) (545) (545) (545) (545) (330) (332) (331) (332) (332) (336) (428)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGpsJ NSYGVTI NSYGVTI NSYGVTI NAYGITI NAYGITI NAYGTI NAYGTI NGYGPTI NGYGPTI NGYGPTI NGYGFTI NGYGFTI NGYGFTI NGYGFTI NGYGFTI NGYGFTI NGYGFTI NGYGFTI GYYGMAJ GYGMTI GYYGMTI dtwWqTI dtYWqTI S SS N-YG-TJ	La (860) Ca (749) Ca (315) Ct (315) Ct (315) Ct (315) Ct (315) Ct (328) Ct (761) Cn (3883) Ct (2788) Ct (2764) Ct (2762) Ct (2692) Ct (2692) Ct (2692) Ct (2692) Ct (2692) Ct (1657) Ct (1657) Ct (1657) Ca (552) Ca (552) Ca (337) Ct (345) Ca (339) Ca (3358) Ct (435) Ct (435)	
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.noc.dB ACVS.pen.dB ACVS.pen.dB ACVS.pen.dA ACVS.cep.dA ACVS.noc.dA EntF Lys2 EntE Luci 4C1 MenE Acua.met Acua.asp Consensus		(707) (596) (173) (185) (609) (1650) (2686) (2641) (2617) (2586) (2545) (1561) (1535) (1535) (1535) (1493) (475) (448) (407) (172) (174) (172) (172) (174) (172) (172) (174) (172) (174) (172) (172) (174) (172) (172) (174) (172) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (172) (174) (AYI i. AYV i. AYV i. AYV i. AYI i. AYU t. AYV t. GvV vgp edftatp spacesf eccepvw vdp ygp AYIapds vns # ayi V v	<pre></pre>	TSGSTGLPK GV TSGSTGLPK GV TSGTTGKPK GV TSGTTGNPK GT TSGTTGNPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGKPK GV TSGTTGFKK GV TSGTTGFFK GI TSGTTGFFK GI TSGTTGFFK GI TSGTTGFFK GV TSGTTGFFK GV TSGTF	visHrG visHrG visHrG nleHkG nleHkG nleHkN nveHrN nieHqS nieHqS lveQgG lveQgG lveHqS lveQgG lveHqS lveHG veHhG nveHhG tveHG kqHtN fkqHtN fkqHtN fkqHtN fkqHtN fkqHtS prtHnd alpHrt nltHkS qqtfgN ahaTgG ahaTgG s <u>H-G</u> <u>S</u>	(728) (617) (194) (206) (206) (1671 (2707 (2707 (2707 (2707 (2707 (2662 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582 (1582) (1582) (226) (225) (226) (225) (298)	L) 7) 1) 2) 3) 7) 5) 5) 5) 5) 5) 1)	(853) (742) (308) (321) (1791) (2835) (2754) (1791) (2835) (2781) (2787) (2781) (2781) (2787) (2781) (2782) (2782) (1658) (1658) (545) (545) (558) (330) (338) (332) (351) (435) (386) (428)	VII amgGaTI amgGaTI NaYGpTI NaYGpTI NhYGp31 NsYGvTI NsYGvTI NaYGTI NaYGTI NaYGTI NaYGTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGpTI NgYGFTI NeYGfTI NeYGfTI NeYGfTI NhYGTI NhYGTI NhYGTI NhYGTI dyYGmai qgYGTI qgYGmTI dtWWqTI dtYWqTI dtYWqTI M	La (860) Ca (749) Ca (315) Ct (328) Ct (2842) Ca (2842) Ca (2764) Ct (2764) Ct (2764) Ct (2692) Ct (1683) Cv (1657) Ct (1641) Cs (565) Ca (552) Ca (552) Ca (375) Ct (345) Ca (358) Ct (435) Ct (435) Ct (435)	

I = P-loop

-----II

FIG. 7—Continu	ied.
----------------	------

			Ś	L E		RIV	D	K		CLA	IS <u>K</u>	G	DV	K	L		
			A	<u> </u>			G			HV	LT <u>E</u>		L		v		
												-					
										V	III						
UNKID 2	(020)	WYDRODI	7		-F	PPTE		017	NU-CUDIELC	2701		a Vier				kakt 1334	(1003)
AMWP2	(936)	WIRIGUI	ΤY.	wpDG			D~Dr	QV OV	KVGGYRIELG	1	sanadia	gvk	IAL VLA.	1		Kekt TAAT	(1003)
MIGK TV1	(027)	WYPTCDI	J.	wpbg			KI DK		KINCHAIEIG	5	valinip	5741	AVAIA	· · ·		A CHAR	(032)
CE1	(390)	IVERCO		1			RIDI	QV 0V	KINGHRIELG		svillane	NICE	t aVaV	L	kdbaa	g.qylcAl	(403)
	(900)	MYPECDI		1			N - Dh		KINGHKVEDE		STRIKIN	yist	N.UTA			q.pyicai	(975)
GSZ.CA	(0%1) (1077)	MINIGUI	T.	1pDG		EVIL	RaDii	20	KINGHRIELG	546	IT-PPP	yvxe hur	Seri -U	 	aduky	g.kylchi	(306)
	(10//)	MIRIGUI		1pDG	Ч.		RIDG		KIRGYRIEFG	5 .	TTAKUK	AVA	Seutas	• • •	edning	q.kalcAl	(1999)
	(2924)	LIKIGDI		mpDG			RIDI	QV	KIRGIRIELG	545	adrikuq	SIKe	MUVIA:	L	eanmx	e.kylchi	(2991)
	(3901)	MINIGDI	T.	rsug	Π.	EIVP	RVDe	QV	KVRGYRIELG	516	salleye	KIK6		S	entas	e.qmLCAI	(4028)
ACVS.asp.dC	(28/5)	LIRTGDI	TRC	riarqnqp	ľ.		RADI	QV	KIRGYRIEPS	E VQ	nvLascp	gvre		k ye	encoaysri	a.kruvGr	(2901)
ACVT.pen.dC	(2851)	LIKTGDI	JAKS	riniddd	J.	EILG	RGDT	QI	KMRGYRIE1S		NULTSSD	gvre	gavva	k ye	ennacysrc	a.nsuvGI	(2927)
ACVS.cep.dC	(2818)	LIKTGDI	JAKI .	rgpn		ET LG	RKDQ	QV	KLKGIRIELS	EVE	aviais	avke		k ya		V.SALVCI	(2891)
ACVS.noc.dC	(2775)	LINTGD	T .	rgpDG		quip	RnDa	QV	KINGIRIEPG	EVE	alages	gvr	ICAVVA	g ao	apqaperkr	- herver	(2894)
ACVS.asp.dB	(1/9/)	LINTGDI	L'URW	1pgsNG	Τ.	EILG	RNDI	QV	KIRGIRIELG		avmssnp		ISAATA S	K • ·	sgkega	q. KILIVGI	(100/)
ACVT.pen.dB	(1//1)	LIKTGDI	. YRW	1pgssG	<u> </u>	EILG	RDDI	QV	KIRGIRIELG	ETE	ailssyn	gike gike	ISVVIA	K	dcrega	q.KILVGI	(1841)
ACVS.cep.dB	(1/45)	LIKTGDI	THE REAL	innangDG	Щ.	EIL	KNDI	QV	KIRGQRIELG		avLssyp	GIK	ISAATU	к	arknag	q.kyuvGI	(1817)
ACVS.noc.dB	(1/29)	LIKTGDI	-VRW		T.	EILG	KCD1	QV	KIRGQRVELG	EVE	aalssyp	gvvi	SIVVA:	F • •	e.navg	q. KyllyGr	(1/96)
ACVS.asp.dA	(710)	MIKTGDI			T .	EILG	RaDI	QI	KLRG1R1EPG	E HE	stLagyp	gvru	SIVUS	K I.	LINGEKELL	n.enuvGI	(782)
ACVT.pen.dA	(683)	MYKTGDI	ARW	1pNG	T.	ETH	RaDI	01 01	KLRG1RIEPG	ETE	mLamyp	IVIT	STAR	K K.	Irngpeett	n.enuvGI	(755)
ACVS.Cep.dA	(653)	MYKTGDI	ARW	1psG	T.	ETUG	RADI	Q1	KLRG1R1EPG	ETE	stLamyp	gira	151775	K K.	Lisqgqeti	q.anuvGI	(725)
ACVS.noc.dA	(640)	LIRTGDI	ARV	1ING	T.	ErmG	RaDI	QL QL	KLNGVRVEPG	ETE	aqaterp	gvk)	CVVVA	κ.	enarg	a.rnivGi	(708)
Entr	(835)	MYRTGD	TARW.	1dNG	av.	ETHE	RSDO	QL	KIRGQRIELG	ΞŢ	rvmqalp	avec	IAVENA	c v:	ingaaatgg	darquvGI	(908)
Lys2	(677)	LYRTGDI	GRY	1pNG	jqc	ECCG	RaDo	QV	KIRGIRIELG	EID	chisqhp	IVre	enit LV:	r .	. Knadnept	litimvpr	(/48)
B-+B	(410)	EVACOD		4-70	·		Deve	.	NecokTles	PTP		-174 -		_	odo I		(477)
LUCE	(410)	HI LCSGDI		dpEG	γy⊥	EVQ.	DIN	17	NIGGER IAde	ETF	di laba	avij	ALVS ALVS	n		mgekschi	(477)
LUCI	(417)	WLDSGDJ		debe			RINS	17	KYKGYQVAPA	212	strtdub	+ 1	INGVAG.	L .	pada	agerphav	(404)
4011	(410)	WINTCD		ddDd	ieL ioT	fr.	DINO	4.7	KYRGIQVAPA	ELE DID	allatha	CISC oTac	i a sviva	n		fooldVAF	(403)
ACT Det	(433)	WLILLOU	Ľ	dubu					AVACEDTARA	ELE	armichp	b Tak		L		LGGI WAE	(501)
Acua.met	(324)	UV FTCD		dkDG	>¥		.DD.	V L	NUCCEDI ALA	PTP	sauvanp	ALC:		.	puev	t control	(539)
Acua neu	(9/2)	UVETCD		dbEC	91 Y		Dund	• •	NVoChDI At A	EIE	autenn	allar ava	aliva	•••·	adel	togavnAF	(591)
Acua.asp	(314)	6+666	6+ 	unee	•¥¥	+ 5	6 6	+6	S+ 6 666+6	66	aantenb	E C	sna v vy.	- • •	auer	CUUN ALA	(301)
Consensus		LVDTCDI	3+ 	D	т 1-т	T J	9 9 9-0-	+ 3 .0V	3+ 3 333+3 KTDC-DIFIC	33 PTP.		-V	->-\//>			I.VAV	
consensus		MINE	1/12	<u>D.</u> .	1-1 V	ET DO	K-R-	<u>-</u>	NU OVADA		T	5	e TV			CGE	
		W N	<u>.</u>		T.	- <u>-</u>		T.		T.	-	5	<u>и</u> <u>т</u> е			2002	
			20	n	ч	*		**					20			A	
						-			TV								
		111														v	
								1A.									

HMWP 2	(880)	PyGfp	ltNgryRVV.	.DergrdcPd	WVSGELWIGG	igVAEGYENd	sLrseqqFLt	1P.	(934)
AngR	(769)	PyGyp	lprqqyRVV.	.DdlgrdcPd	wvaGELwIGG	dgIAlGYEdd	eLkTqaqFLh	id.	(823)
TY1	(335)	PIGkp	iqNthiYIV.	.NedlQllPt	adeGELCIGG	vgLARGYWNr	pDLTaeKFVd	NPF	(390)
GS1	(347)	PIGap	iqNtqiYIV.	.DenlQlksv	geaGELCIGG	egLARGYwkr	pELTsqKFVd	NPF	(402)
GS2.dA	(780)	PIGkp	isNtwiYI1.	.DqeqQlqPq	givGELYIsG	anVgRGYLNn	qELTaeKFfa	dPF	(835)
GS2.dB	(1816)	PIGka	isNstvYIm.	.DrygQlqPv	gvpGELCVGG	dgVARGYnNq	palTeeKFVp	NPF	(1871)
GS2.dC	(2863)	PIGkp	yaNmkmYIm.	.NqylQiqPv	gviGELCIGG	agVARGYLNr	pDLTaeKFVp	NPF	(2918)
GS2.dD	(3900)	tIGrp	lsNvdvYIV.	.NcnhQlqPv	gvvGELCIGG	qgLARGYLNk	pELTadKFVv	NPF	(3955)
ACVS.asp.dC	(2806)	ALrel	lpGtraYlL.	.NhatQpvPm	navGELYLAG	dcVARGYLNq	pvLTgdRFIq	NPF	(2861)
ACVT.pen.dC	(2782)	ALrev	lpGtraYVL.	.NaalQpvPf	davGELYLAG	dsVTRGYLNq	pLLTdqRFIp	NPF	(2837)
ACVS.cep.dC	(2750)	ALchg	ipGshvYVL.	.NdrlQrvPf	navGELYLGG	dcLARGYLNq	daLTneRFIp	NPF	(2805)
ACVS.noc.dC	(2710)	tLGap	lgNtrlYVL.	.GdgmKllPt	gavGELYLAG	dcVTEGYLhr	pELTreRFLp	NPF	(2765)
ACVS.asp.dB	(1728)	SIGqq	igNstsYVL.	.NadmKrvPi	gavGELYLGG	egVARGYnNr	pEVTaeRFLr	NPF	(1783)
ACVT.pen.dB	(1702)	SIGqq	vhNstsYVL.	.NedmKrtPi	gavGELYLGG	egVvRGYhNr	aDVTaeRFIp	NPF	(1757)
ACVS.cep.dB	(1676)	SIGcq	ldNstsYVL.	.NddmKrvPi	GavGELYLGG	dgVARGYnNr	pDLTadRFpa	NPF	(1731)
ACVS.noc.dB	(1660)	SIGfp	vaNtkchVL.	.NkamKpvPv	ggiGELYIGG	igVTRGYLNr	eDLTadRFVe	NPF	(1715)
ACVS.asp.dA	(641)	SLGrp	vrNvkcYIL.	.NkslKrvPi	gatGELHIGG	lgISKGYINr	pDLTpqRFIp	NPF	(696)
ACVT.pen.dA	(614)	SLGrp	vrNvkcYIL.	.NpslKrvPi	gatGELHIGG	lgISKGYINr	pELTphRFIp	NPF	(669)
ACVS.cep.dA	(584)	SLGrp	vrNvkcYIL.	.DanlKrvPi	gvtGELHIGG	lgISRGYmNr	eELTrqKFLp	NPy	(639)
ACVS.noc.dA	(572)	SIGrp	lrNvkwYVL.	.sqglKqlPi	gaiGELYIGG	CGVApGYLNr	dDLTaeRFta	NPF	(627)
EntF	(774)	PIGyp	vwNtglRIL.	.DammhpvPp	gvaGdLYLtG	iqLAqGYLGr	pDLTasRFIa	dPF	(829)
Lys2	(592)	PaGkg	mlNvqllVVn	rNdrtQicgi	geiGEiYVra	ggLAEGYrGl	pELnkeKFVn	NwF	(649)
					•				
EntE	(357)	ypmcp	ddEvwvaecr	rkstaarev.	ggrlmtrg	.pytfRGYyks	pqhnasaF	• • •	(405)
Luci	(3େ)	gkvvp	ffEakvvdld	tGktlgvnqr	gelcvrg	pmImsGYvNn	pEaTnali	• • •	(412)
4C11	(362)	gtvvr	naEmkivdpe	tNaslprnqr	geicirg	dqImKGY1Nd	pEsTrtti	• • •	(411)
4C1.	(381)	gtvvr	naElkiidpd	tGkslgrnlr	geicirg	qqImKGY1Nn	pEaTknti	• • •	(430)
Acua.met	(461)	tfplp	gydisilde.	eGnevplgsg	gnivaLk.py	psmlRafwGd	kErfmkeywq	fyw	(516)
Acua.neu	(413)	Sfpff	giEpalvdpv	tGeeirgndv	egvlafkqpw	psmARtvwGa	hkrymetyLh	vy.	(469)
Acua.asp	(455)	SLpff	giEpaiidpv	sGeeisgndv	egvlafkqpw	psmARtvwGa	hkrymdtyLq	vy.	(511)
		‡	` S ‡		+++ + +	‡ ‡ ‡	‡		
Consensus		PIG	<u>N</u> YVL-	- <u>N</u> QP-	GELYIGG	VA <u>RGY-N</u> -	- <u>E</u> LTRFI-	NPF	
		SL	<u>e</u> riv	<u>D</u> K	CLA	IS <u>K G</u>	<u>D</u> V KL		
		A	G	G	HV	LT <u>E</u>	T A		

HMWP2	(1356	5) La	grLPgwMVPq	rIVfLda	LPLTaNGKI	DygALK (1389)		
AngR	(916)	Ln	kqLPnYMVPk	rIIfLet	.fPLTaNGKV	DhKALt (949)		
TYÍ	(479)	aa	gkLPaYMlPs	yFVkLd.	MPLTpNdKI	DrKALP (512)		
GS1	(491)	SS	eeLPtYMIPs	vFlaLd.	MPLTSNGKI	DrKgLP (524)		
GS2.dA	(924)	Lα	kaLPdYMIPs	fFVpLd.	hvrLhlNGKI	DrKsLP (957)		
GS2.dB	(1960)) ia	keLPvYMVPa	vFVale.	MPLTONGKV	nrsALP (1993)		
GS2.dC	(3007) La	ndrAa.MIPs	vFVsLe.	MPLTANGKI	DkRsLP (3039)		
GS2.dD	(4044) La	klLPsYMIPn	vFIgLd.	SIPLTONGKV	DrKALP (4077)		
ACVS.asp.dC	(2965	5) Mk	skLPaYMVPk	vLcrLeg	DEPUTINGKL	DVRKLP (2999)		
ACVT.pen.dC	(2944) Mie	ar I.Pt YMVPs	hLuccleg	ALPVTINGKI.	DVRRLP (2978)		
ACVS.cep.dC	(2909) T.h	anLPpYMVPs	alhalea	SLPVTVNGKL	DinRLs (2943)		
ACVS.noc.dC	(2862) Lr	agLmpSMVPs	1L. VrLdr	DEPMTITGKL	DVdALP (2896)		
ACVS.asp.dB	(1883) Ma	srLPgYMIPs	sFIpI.s	SLPVTpSGKL	DEKALP (1916)		
ACVT.pen.dB	(1857) Ma	srLPqYMVPs	rLIlv.s	KEPVTpSGKL	DtKALP (1890)		
ACVS.cep.dB	(1834	і) мі	tsLPdYMVPa	qLVpI.a	REPVTVSGKL	DaKALP (1867)		
ACVS.noc.dB	(1812) Mr	kkLPeSvVPa	rvlrI.t	dIPVTpSGKL	DaRRLP (1845)		
ACVS.asp.dA	(799)	Le	lkLPrYMIPt	rLVrv.s	GIPVTVNGKa	DIRALP (832)		
ACVT.pen.dA	(772)	Le	kkLPrYMIPt	rLVqL.s	GIPVNVNGK a	DIRALP (805)		
ACVS.cep.dA	(742)	Le	kkLPrYMVPt	rLVqL.a	DIPTNINGKa	DIRALP (775)		
ACVS.noc.dA	(726)	Le	qrLiriMVPa	rmVrL.t	SIPVNVNGKV	DWRALP (759)		
EntF	(926)	Lr	etLPphMVPv	vLlqL.p	GLPLSaNGKL	DrKALP (959)		
Lys2	(789)	Lk	krLAsYamPs	lIVvmd.	KLPLNpNGKV	DkpKLq (822)		
-						- 1			
EntE	(493)	Lr	eqgiaef	kLpdrVecvd	sLPLTavGKV	DkKqLR (527)		
Luci	(502)	va	sqvttak	kLrggVvFvd	evPKglTGKL	DaRKIR (536)		
4C11	(501)	vs	kqvvfyk	rIf.rVfFvd	aIPKSpSGKI	LrKdLR (534)		
4C1	(517)	va	keviyyk	kIr.eVfFvd	KIPKapSGKI	LrKeLR (550)		
Acua.met	(609)	ia	f.vrktLgPv	aapteVhFvn	dLPKTrSGKI	MrRvvK (645)		
Acua.neu	(556)	DS	lovrrSigPf	aapkaIvIvp	dLPKT1SGKT	MrRiLR (593)		
Acua.asp	(598)	Li	lavrkSigPf	aapkaVfvvd	dLPKTrSGKI	MrRilR (635)		
•	••		** **	+	551 555	5 5			
Consensus		L-	LP-YMVP-	-LY-L	-LPLT-NGKI	D-RALP			
		м	ASI	FII	IVNSL	MKKR			
				I E	MKSTV	LRK			
					VI				
HMWP2	(1410) E	2kqVa	alWqqLLs	tgnvtretDF	FqqGGDSL1A	trltgqL.	.h qagyeaqlsd	LFnhPrL	(1467)
HMWP2 Angr	(1410) E (971) E	kqVa drVa	aLWqqLLs kIWndVLG	tgnvtretDF ptelyksSDF	FqqGGDSL1A F1sGGDaynA	trltgqL. ieVvkrc.	.h qagyeaqlsd .h kagylIklsm	LFnhPrL LYrySTI	(1467) (1028)
HMWP2 AngR TY1	(1410) E (971) E (533) E	lkqVa LdrVa LsiLv	aLWqqLLs kIWndVLG sIWqnVLGi.	tgnvtretDF ptelyksSDF .ekiGirDNF	FqqGGDSL1A FlsGGDaynA YsLGGDSIqA	trltgqL. ieVvkrc. iqVvarL.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd	LFnhPrL LYrySTI LLnyPTI	(1467) (1028) (590)
HMWP2 AngR TY1 GS1	(1410) E (971) E (533) E (544) E	EkqVa EdrVa EsiLv EstLv	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi.	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF	FqqGGDSL1A FlsGGDaynA YsLGGDSIqA YaLGGDSIkA	trltgqL. ieVvkrc. iqVvarL. iqVaarL.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd	LFnhPrL LYrySTI LLnyPTI LLkyPTI	(1467) (1028) (590) (601)
HMWP2 AngR TY1 GS1 GS2.dA	(1410) E (971) E (533) E (544) E (977) E	CkqVa CdrVa CsiLv CetLv CetLv CekLa	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi.	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl	LFnhPrL LYrySTI LLnyPTI LLkyPTI LFeaPTI	(1467) (1028) (590) (601) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB	(1410) E (971) E (533) E (544) E (977) E (2013) E	CkqVa CdrVa CsiLv CetLv CetLa CmkLa	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv.	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .nkiGv1DNF	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FeLGGHSLrA	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm. mtmisqV.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv	LFnhPrL LYrySTI LLnyPTI LLkyPTI LFeaPTI LFetPTI	(1467) (1028) (590) (601) (1035) (2071)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E	CkqVa CdrVa CsiLv CetLv CekLa CmkLa CgkLe	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1.	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGv1DNF .qrvGihDDF	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FeLGGHSLkA FtiGGHSLkA	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itLisrm. mtmisqV. maVisqV.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv	LFnhPrL LYrySTI LLnyPTI LLkyPTI LFeaPTI LFetPTI LFetPTI	(1467) (1028) (590) (601) (1035) (2071) (3117)
HMWF2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC	(1410) E (971) E (533) E (544) E (977) E (2013) E (2013) E (3059) E (4097) E	CkqVa CdrVa CsiLv CetLv CetLv CekLa CmkLa CgkLe CaqLv	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhVLGv. eIWkdVLG1. lIWqeVLGi.	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .eliGitDNF	FqqGGDSL1A FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLCGHSLkA FeLGGHSLkA FeLGGHSLkA	trltgqL. ieVvkrc. iqVvarL. itlisrm. itlisrm. mtmisqV. maVisqV. tllvakI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI LFetPTI VFkhSTI	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC GS2.dD ACVS.asp.dC	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (3021) E	CkqVa CdrVa SsiLv CetLv CetLa CmkLa CgkLe CaqLv CaqLv CadLc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWhdVLG1. lIWqeVLGi. rLWasaLGt.	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .eliGitDNF .ercGidDDL	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FeLGGHSLkA FeLGGHSLkA FrLGGDSItA	trltgqL. ieVvkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhLaaqI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd	LFnhPrL LYrySTI LLayPTI LLkyPTI LFetPTI LFetPTI VFkhSTI IFdhPT.	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC	(1410) E (971) E (533) E (544) E (977) E (3059) E (4097) E (3021) E (2997) E	CkqVa CdrVa CsiLv CetLv CekLa CmkLa CgkLe CaqLv CadLc Cakmc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWhVLGv. eIWkdVLG1. 1IWqeVLG1. rLWasaLGt. rLWesaLGm.	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .sqiGiqDNF .qrvGihODF .eliGitDNF .ercGidDL .ercGidDL	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FtIGGHSLkA FtIGGHSLkA FrLGGDSItA FkLGGDSItS	trltgqL. ieVwkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhlvaqI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd hn qvgckItvrd	LFnhPrL LYrySTI LLnyPTI LLkyPTI LFeaPTI LFetPTI LFetPTI IFdhPT. IFehRTa	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC ACVT.pen.dC	(1410) E (971) E (533) E (544) E (977) E (3059) E (4097) E (2997) E (2997) E (2961) E	CkqVa CdrVa CsiLv CetLv CekLa CmkLa CgkLe CaqLv CadLc CatLc CetLc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. rLWasLGt. rLWesaLGm. qLWasLLGv.	tgnvtretDF ptelyksSDF .ekiGirDNF .griGigDNF .nkiGv1DNF .eliGitDNF .ercGidDDL .ercGidDDL .dhcGidDDL	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FeLGGHSLkA FrLGGHSLkA FrLGGDSItA FkLGGDSItS FarGGDSISS	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhIvaqI. lrlvgdI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd nn qugckItvrd yr algrkVtvkd	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI LFetPTI VFkhSTI IFdhPT. IFehRTa IYlhRsV	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (3021) E (2997) E (2991) E	CkqVa CdrVa SsiLv CetLv CekLa CmkLa CgkLe CagkLe CagkLe CagkLe CagkLe CagkLe CagkLe CagkLe CagkLe	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLGi. rLWasaLGt. rLWasaLGt. nLWasaLLGv. hLWsaqLpg.	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .eliGitDNF .ercGidDDL .ercGidDDL .dhcGidDL .gtvGidDDF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YaLGGDSIKA FsLGGHSLKA FeLGGHSLKA FeLGGHSLKA FrLGGDSILA FrLGGDSILS FarGGDSISS FrCGGDSISA	trltgqL. ieVvkrc. iqVvarL. itIisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhlvaqI. lhIvqGI. lhlasqV.	.h qagyeaqlsd .h kagyllklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky	LFnhPrL LYrySTI LLkyPTI LFeaPTI LFetPTI LFetPTI IFdhPT. IFchRTa IYlhRsV LFdhPTV	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019) (2973)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.pen.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (4097) E (2997) E (2997) E (2915) E (1934) E	CkqVa CdrVa SsiLv CetLv CekLa CmkLa CgkLe CagLv CadLc CatLc CatLc CatLc SsiLc	aLWqqLLs kIWndVIG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1. 1IWqeVLG1. 1IWqeVLG1. rLWasaLGt. rLWasaLGt. nLWsaqLpg. gIsagLLdis	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .ercGidDDL .ercGidDDL .dhcGidDDL .gtvGidDDF aqtiGsdSDF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YsLGGDSIkA FsLGGHSLkA FeLGGHSLkA FrLGGHSLkA FrLGGDSItA FrLGGDSItS FrCGGDSISA FtLGGDSLkS	trltgqL. ieVvkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhIvaqI. lhIvaqI. lhIvaqI. lhIsqV. tkIsfkI.	.h qagyeaqlsd .h kagyllklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvkd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI LFetPTI VFKhSTI IFdhPT. IFehRT. IYlhRsV LFdhPTV LFrhRT.	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3055) (3019) (2973) (1993)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACV5.asp.dC ACV5.cep.dC ACV5.cep.dC ACV5.noc.dC ACV5.asp.dB ACV5.asp.dB	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (4097) E (3021) E (297) E (2961) E (2915) E (1934) E (1908) E	CkqVa CdrVa SsiLv CetLv CekLa CgkLe CaghC	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. lIWqeVLG1. rLWasaLGt. rLWasaLGt. qLWasLLGw. hLWsaqLpg. gIsagLLdis dIWaeLLEmh	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .nkiGvlDNF .grvGihDDF .ercGidDDL .ercGidDDL .dhcGidDDL .gtvGidDDF aqtiGsdSDF peeiGiySDF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YaLGGDSIKA FsLGGHSLKA FeLGGHSLKA FrLGGHSLKA FrLGGDSItS FarGGDSISS FrCGDSISS FrLGGDSLKS FsLGGDSLKS	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhLaaqI. lhLaaqI. lhLaaqI. lhIasqV. tkIsfkI. tkIsfmI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymq1emplnv hh qigrkVtvrd hn qvgckItvrd yr algrkVtvkd qr e1erkVsvky he vfgrtIsvsa he sfnraVsvsa	LFnhPrL LYrySTI LLryPTI LFeePTI LFetPTI IFdhPT. IFdhPT. IFdhRTA IYlhRsV LFdhPTV LFchRTV	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3059) (3019) (2973) (1993) (1968)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.asp.dB ACVS.pen.dB ACVT.pen.dB	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (3059) E (3021) E (2997) E (2997) E (2997) E (2915) E (1934) E (1908) E (1985) E	CkqVa CdrVa SsiLv CetLv CetLv CetLa CgkLe CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLv CaqLs CaqLa CaqLa CaqLa CagLa	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLGi. rLWasaLGt. rLWasaLGt. gIsagLLdis dIWaeLLEmh gIWasLLEip	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .eliGitDNF .ercGidDDL .dhcGidDDL .gtvGidDDF aqtiGsdSDF peeiGiySDF	FqqGGDSL1A FlsGGDSL1A YsLGGDSIQA YaLGGDSIQA FsLGGHSLKA FeLGGHSLKA FrLGGHSLKA FrLGGDSItA FrLGGDSISS FrCGGDSISA FtLGGDSLKS FsLGGDSLKS	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhlaaqI. lhlaaqI. lhlaaqV. tkIsfkI. tkIsfmI. tkIsfaa.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa tr algvaVsvrn	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI VFkhSTI IFdhPT. IFdhPT. IFdhRTa IYlhRsV LFdhPTV LFrhRT. LFchRTV LFshPTI	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3055) (3019) (2973) (1993) (1993) (1968) (1945)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.roc.dC ACVS.noc.dC ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.noc.dB	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (4097) E (3021) E (2997) E (2961) E (2915) E (1908) E (1885) E (1865) E	CkqVa CdrVa SsiLv CetLv CetLv CetLa CgkLe CaqLv CadLc CaqLv CadLc CatLc CatLc CatLc CatLc CatLc CatLc CatLc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGV. eIWkdVLG1. lIWqeVLG1. rLWasaLG1. rLWasaLG2. hLWsaqL9. gIsagLLdis dIWaeLLE1p gIWaqVLE1a	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .grvGihDDF .eliGitDNF .ercGidDL .ercGidDL .gtvGidDDL .gtvGidDDL aqtiGsdSDF peeiGiySDF vdriSiySDF	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YaLGGDSIkA FsLGGHSLkA FtIGGHSLkA FrLGGHSLkA FrLGGDSItS FrCGDSISS FrCGDSISS FrLGGDSLkS FsLGGDSLkS FsLGGDSLKS	trltgqL. ieVwkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhIaaqI. lhIvaqI. lhIasqV. tkIsfmI. tkIsfmI. tkIsfaa. malaqaI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv y mqiemplnv hh qigrkVtvrd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa tr algvaVsvrn tt gfqqgLgvat	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI VFkhSTI IFdhPT. IFehRTa IY1hRsV LFdhPTV LFchRTV LFchRTV LFchRTV	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3055) (3019) (2973) (1993) (1998) (1945) (1945)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.roc.dC ACVS.roc.dC ACVS.roe.dB ACVS.cep.dB ACVS.roc.dB ACVS.noc.dB ACVS.noc.dB ACVS.noc.dB ACVS.noc.dB	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (3059) E (3059) E (3059) E (2997) E (2997) E (2915) E (1908) E (1908) E (1865) E (851) E (851) E	CkqVa CarVa SetLv CetLv CetLv CetLv CetLv CetLo CakLa CgkLe CaqLv CakLa CgkLe CatLc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. rLWasaLGt. rLWasaLGt. gIsagLLGs. dIWasLLGy. hLWsaqLpg. gIsagLLdis dIWaeLLEmh gIWseLLEip gIWaqVLE1a kIWadVLGA.	tgnvtretDF ptelyksSDF .ekiGirDNF .griGigDNF .nkiGv1DNF .eliGitDNF .ercGidDDL .ercGidDDL .gtvGidDDL .gtvGidDDF aqtiGsdSDF peeiGiySDF vdriSiySDF pdriGvhDDF hlsiSrkDNF	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YsLGGDSIkA FsLGGHSLkA FeLGGHSLkA FrLGGDSILA FrLGGDSILS FrLGGDSILS FrLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKA	trltgqL. ieVvkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhIaqI. lhIvaqI. lhIvaqI. lhIasqV. tkIsfkI. tkIsfmI. tkIsfaa. maIaqaI. iqIiarI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd nn qvgckItvrd yr algrkVtvkd qr elerkVsvky he vfgrtIsvsa tr algvaVsvn tt gfgqgLgvat rq qlgviIsied	LFnhPrL LYrySTI LLnyPTI LFetPTI LFetPTI VFkhSTI IFdhPT. IFehRTa IY1hRsV LFdhPTV LFchRTV LFchRTV LFchRTV LFshPTI VLqhtTL VFssRT.	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019) (2973) (1968) (1945) (1945) (1925) (920)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB ACVT.pen.dB ACVS.cep.dB ACVS.cep.dB ACVS.noc.dB ACVS.asp.dA ACVS.asp.dA ACVS.asp.dA	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (3021) E (297) E (2961) E (2961) E (1934) E (1934) E (1968) E (1865) E (851) E (851) E (824) E	CkqVa ckqVa cklv ckla	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLGi. 1IWqeVLGi. rLWasaLGt. rLWasaLGv. hLWsaqLpg. gIsagLLdis dIWaeLLEip gIWaqVLEia kIWadVLGah eIWadVLGah	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .nkiGv1DNF .eliGitDNF .ercGidDDL .ercGidDDL .dhcGidDDL .gtvGidDDF peeiGiySDF pdriGvhDDF hlsiSrkDNF grsvSrnDNF	FqqGGDSLlA FlsGGDaynA YsLGGDSIQA YsLGGDSIQA FsLGGHSLkA FeLGGHSLkA FrLGGDSILS FrLGGDSILS FarGGDSISS FrCGDSLSS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS	trltgqL. ieVvkrc. iqVvarL. iqVaarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhIaqI. lhIvaqI. lhIvaqI. lhIasqV. tkIsfkI. tkIsfkI. tkIsfaa. maIaqaI. iqIiarI. iqIiarI.	.h qagyeaqlsd h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd n qugckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvrn tt gfgqgLgvat rq qlgviIsied rq rlsvsIsved	LFnhPrL LYrySTI LLkyPTI LFeaPTI LFetPTI IFehPT. IFehPT. IFehRTa IYlhRsV LFdhPTV LFshPTI VLGhTTL VFssRT. VFatRTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019) (2973) (1993) (1993) (1993) (1945) (1925) (910) (884)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.pen.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB ACVS.cep.dB ACVS.cep.dB ACVS.asp.dA ACVS.asp.dA ACVS.asp.dA ACVS.cep.dA	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (4097) E (2961) E (2961) E (2961) E (1934) E (1908) E (1885) E (1885) E (851) E (824) E (796) E	CkqVa SdrVa SsiLv SetLv SetLv SetLa SatLc SatLc SatLc SiLc SiLc SiLc SiLc SiLc SiLc SiLc Si	aLWqqLLs kIWndVIG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. lIWqeVLG1. lIWqeVLG1. rLWasaLGv. hLWsaqLpg. gIsagLLdis dIWaeLLEmp gIWaeLLEmp gIWaeVLE1a kIWadVLGan eIWadVLGan aIWgniLsvp	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .nkiGvlDNF .ercGidDDL .ercGidDDL .dhcGidDL .gtvGidDDF aqtiGsdSDF periGiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF aqdiGseNF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YaLGGDSIKA FsLGGHSLKA FeLGGHSLKA FeLGGHSLKA FrLGGDSITA FrLGGDSITS FrCGGDSISS FrCGGDSISS FsLGGDSLKS FsLGGDSLKS FaLGGDSITA FrLGGHSITC FrLGGHSITC FrLGGHSIC	trltgqL. ieVvkrc. iqVvarL. itIisrm. mtmisqV. maVisqV. tllvakI. lhlvaqI. lhlvaqI. lhlasqV. tkIsfkI. tkIsfkI. tkIsfaa. maIaqaI. iqIiarI. iqIiarI.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvd nn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvn tt gfqgLgvat q qlgviIsied rq rlsvsIsved rq qlgqItlee	LFnhPrL LYrySTI LLkyPTI LFeaPTI LFetPTI IFehPTI IFdhPT. IFchRTA IYlhRSV LFdhPTV LFshPTI VLghtTL VFssRT. VFatRTL VFatRTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019) (2973) (1993) (1993) (1993) (1994) (1945) (1925) (910) (884) (856)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.pen.dC ACVS.cep.dC ACVS.noc.dC ACVS.asp.dB ACVS.cep.dB ACVS.cep.dB ACVS.asp.dA ACVS.cep.dB ACVS.asp.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (4097) E (2907) E (2907) E (2907) E (2915) E (1934) E (1908) E (1865) E (851) E (851) E (824) E (796) E (789) t (977) E	kqVa kdrVa satLv actLv actLv actLc a	aLWqqLLs kIWndVIG sIWqnVIGi. tIWqdVIGi. kIWeeVIGi. eIWhnVIGv. eIWkdVIG1. 1IWqeVIG1. 1IWqeVIG1. rLWasaLGt. rLWasaLGt. rLWasaLGt. gIwasLLGy. hLWsaqLpg. gIsagLLdis dIWaeLLEmp gIWaqVLEia kIWadVIGan aIWgniLsvp aIWseVLGyr	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .ercGidDDL .ercGidDDL .dhcGidDDL .gtvGidDDF aqtiGsdSDF periGiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF aqdiGseSNF qnriGerDDF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YaLGGDSIkA FsLGGHSLkA FeLGGHSLkA FrLGGDSItA FrLGGDSItA FrLGGDSItS FrLGGDSLkS FsLGGDSLkS FsLGGDSLkS FsLGGDSLKS FsLGGDSIA FrLGGHSITC FrLGGHSITC FrLGGHSIC	trltgqL. ieVvkrc. iqVvarL. itIisrm. mtmisqV. maVisqV. tllvakI. lhlvaqI. lhlvaqI. lhlvaqI. lhlasqV. tkIsfKI. tkIsfKI. tkIsfMI. iqLiarI. iqLiarI. iqLiarV.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd nn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvn tt gfgqgLgvat rq qlgvIIsied rq rlsvsIsved rq rlssLgved	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFeaPTI IFehTT IFehTT. IFehTT. IFehTT. LFchRTV LFshPTI VLqhTL VFasRT. VFatRTL VFatRTL	(1467) (1028) (590) (601) (3117) (4155) (3078) (3055) (3055) (3055) (3055) (3055) (3055) (3055) (3055) (1993) (1993) (1945) (1945) (1925) (910) (884) (856) (849)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dA GS2.dC GS2.dC GS2.dC ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.asp.dB ACVS.sep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3021) E (3021) E (2961) E (2961) E (1908) E (1908) E (1885) E (851) E (851) E (851) E (851) E (851) E (851) E (851) E (851) E	kqVa kdrVa ksiLv ketLv ketLu kakLa kgkLe kakLa kakLa katLc k	aLWqqLLs kIWndVIG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. rLWasalGt. rLWasalGt. rLWasaLGv. hLWsaqLpg. gIsagLLdis dIWaeLLEip gIWaqVLEia kIWadVLGar aIWgnLSvp aafssLLGvp aafssLLGvp	tgnvtretDF ptelyksSDF .ekiGirDNF .sqiGiqDNF .nkiGvlDNF .qrvGihDDF .ercGidDDL .ercGidDDL .gtvGidDDF aqtiGsdSDF pedGidSDF pdriGvhDF hlsiSrkDNF qrsvSrnDNF aqdiGseSNF qnriGerDDF vqdadaDF	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YsLGGDSIAA FsLGGHSLKA FeLGGHSLKA FrLGGDSIXA FrLGGDSIXS FrCGGDSISS FrCGGDSISS FsLGGDSLKS FsLGGDSLKS FaLGGDSIXA FrLGGHSITC FrLGGHSITC FrLGGHSITC FrLGGHSIC FrLGGSISC FaLGGSLSC FaLGGSLSC	trltgqL. ieVvkrc. iqVvarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhIvaqI. lhIvaqI. lhIvaqI. lhIsqV. tkIsfkI. tkIsfmI. tkIsfmI. tkIsfaa. maIaqaI. iqIiarI. iqIiarI. iqIiarV. ilIiarV.	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hn qigrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvrn tt gfggLgvat rq qlgviIsied rq rlsvsIsved rq qlgqgItlee rq rlslsLgved .r qvarqVtpgq	LFnhPrL LYrySTI LLryPTI LFetPTI LFetPTI VFkhSTI IFdhPT. IFehRT. LFdhPTV LFnRT. LFchRTV LFshPTI VLghtTL VFatRTL VFatRTL VFatRTL VFatRTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3019) (2973) (1993) (1993) (1945) (1945) (1945) (1945) (910) (884) (884) (884) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVT.pen.dC ACVS.cep.dC ACVS.cep.dB ACVS.noc.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (3021) E (3021) E (2997) E (2997) E (2997) E (1908) E (1908) E (1865) E (1865) E (824) E (796) E (789) t (977) E	kqVa kdrVa ksiLv ketLv ketLv kakLa kgkLe kakLc kakLc katLc ksiLc ksiLc kiLkc kiLccc kiLccc kiLccc kiLccc kiLcccc kiLcccc kiLccccc kiLcccccccccc	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. lIWqeVLG1. rLWasaLGt. rLWasaLGt. gIsagLLdis dIWaeLLEmh gIWagVLEia kIWadVLGah eIWadVLGah aIWgnLLSvp afssLLGcd ‡ ‡	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .grvGihDNF .ercGidDL .ercGidDL .gtvGidDDL .gtvGidDDF aqtiGsdSDF peeiGiySDF vdriSiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF qrsvSrnDNF qriGerDDF vqdadaDF ##	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YaLGGDSIKA FsLGGHSLKA FsLGGHSLKA FrLGGHSLKA FrLGGDSIKS FrCGGDSISS FrCGGDSISS FsLGGDSLKS FsLGGDSLKS FaLGGDSIC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSLA ‡ ±‡±±‡	trltgqL. ieVwkrc. iqVaarL. itIisrm. mtmisqV. maVisqV. tllvakI. lhIaaqI. lhIvaqI. lrIvgdI. lhIasqV. tkIsfmI. tkIsfmI. tkIsfmI. tkIsfaa. maIaqaI. iqIiarI. iqIiarV. illiarV. mkLaaqIs *	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv yh qigrkVtvrd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa tr algvaVsvrn tt gfgqgLgvat rq qlgviIsied rq rlsvsIsved rq rlsglLgved .r qvarqVtpgq	LFnhPrL LYrySTI LLnyPTI LFetPTI LFetPTI IFdhPT. IFdhPT. IFdhRTA IYlhRsV LFdhPTV LFohRTV LFshPTI VLghtTL VFstRTL VFqtKTL VFqtKTL VFqtRTL VFqlRTL VFqlRTL VFqlRTL VFqlRTL VFqlRTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3019) (2973) (1993) (1993) (1993) (1993) (1945) (1925) (910) (884) (884) (856) (8849) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.cep.dC ACVS.noc.dC ACVS.noc.dC ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Consensus	(1410) E (971) E (533) E (544) E (2013) E (2013) E (3059) E (4097) E (3021) E (2997) E (2997) E (1908) E (1908) E (1885) E (1865) E (824) E (786) E (789) t (977) E E	kqVa kdrVa kdrVa ketLv ketLv ketLv ketLc kakLa kak	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. kIWeeVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLGi. rLWasaLGt. rLWasaLGt. gIsagLLGis dIWaeLLEmh gIWseLLEip gIWaqVLEia kIWadVLGan eIWadVLGan aIWgniLsvp aafssLLGcd ‡ ‡ -IWVLG L. F	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .nkiGv1DNF .eliGitDNF .ercGidDL .ercGidDL .ercGidDL .gtvGidDDL .gtvGidDDL .gtvGidDDF paqtiGsdSDF pdriGvDDF hlsiSrkDNF qrsvSrnDNF qrdiGseSNF qnriGerDDF vqdadaDF ***	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YsLGGDSIkA FsLGGHSLkA FtLGGHSLkA FrLGGDSILS FrLGGDSILS FrLGGDSILS FrLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSICA FrLGGHSICC FrLGGHSICC FrLGGHSLA ‡ ‡‡‡‡‡ ‡	trltgqL. ieVvkrc. iqVarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhIvaqI. lrIvgdI. lhIasqV. tkIsfmI. tkIsfmI. tkIsfmI. tkIsfaa. maIaqaI. iqIiarI. iqIiarV. jlliarV. mkIaaqIs *	.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv ye ymqiemplnv hn qvgckItvrd yr algrkVtvrd hn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa tr algvaVsvrn tt gfgqgLgvat rq qlgviIsied rq rlsvsIsved rq rlsvsIsved rq rlsvsLgvq .r qvarqVtpgq	LFnhPrL LYrySTI LLryPTI LFeaPTI LFetPTI VFkhSTI IFdhPT. IFdhPT. IFdhRTA IYlhRsV LFdhPTV LFrhRT. LFchRTV LFshPTI VLqhtTL VFatRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL VFqtRTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3055) (3019) (2973) (1993) (1993) (1993) (1945) (1925) (910) (884) (856) (849) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.cep.dC ACVS.roc.dC ACVS.sap.dB ACVT.pen.dB ACVS.cep.dB ACVS.cep.dB ACVS.sap.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Consensus	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (297) E (2961) E (2961) E (2961) E (1934) E (1934) E (1938) E (1938) E (1865) E (851) E (851) E (851) E (851) E (851) E (824) E (789) t (977) E	kqVa kdrVa ssiLv ketLv ketLa katLa katLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc ssiLc v	aLWqqLLs kIWndVLG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLGi. lIWqeVLGi. rLWasaLGt. rLWasaLGv. hLWsaqLpg. gIsagLLdis dIWaeLLEip gIWaqVLEia kIWadVLGah eIWadVLGA eIWadVLGA eIWadVLGA eIWadVLGA eIWA	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .nkiGv1DNF .eliGitDNF .ercGidDDL .dhcGidDDL .dtrGidDDL .gtvGidDDF peeiGiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF qrsvSrnDNF qriGerDDF vqdadaDF ## GDDF S SNL	FqqGGDSLlA FlsGGDaynA YsLGGDSIqA YsLGGDSIkA FsLGGHSLkA FeLGGHSLkA FrLGGDSItS FrLGGDSItS FrLGGDSItS FsLGGDSLkS FsLGGDSLkS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FsLGGDSICA FrLGGHSICC FrLGGHSICC FrLGGHSICA FrLGGHSLA Y HLS	trltgqL. ieVvkrc. iqVarL. itlisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhIvaqI. lhIasqV. tkIsfKI. tkIsfKI. tkIsfaa. maIaqaI. iqIiarI. iqIiarV. illiarV. illiarV. illiarV. jliarV.	<pre>.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd nn qvgckItvrd yr algrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvrn tt gfgggLgvat rq qlgviIsied rq rlsvsIsved rq rlsvsLyved .r qvarqVtpgq </pre>	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI IFehPT. IFehPT. IFehPT. IFchRTA IYlhRsV LFdhPTV LFshPTI VLGhTTL VFatRTL VFATTL VFATTL	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3078) (3055) (3019) (2973) (1993) (1993) (1993) (1945) (1925) (910) (884) (884) (849) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.noc.dC ACVS.noc.dC ACVS.sap.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Consensus	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (2901) E (2901) E (2961) E (1934) E (1908) E (1934) E (1908) E (1885) E (824) E (789) t (977) E E	SkqVa SdrVa SsiLv SetLv SetLv SetLa SatLc SatLc SiLc SiLc SiLc SiLc SiLc SiLc SiLc Si	aLWqqLLs kIWndVIG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLGi. IIWqeVLGi. rLWasaLGt. rLWasaLGt. rLWasaLGy. hLWsaqLpg. gIsagLLdis dIWaeLLEip gIWaqVLEia kIWadVLGan eIWadVLGan aIWgniLsvp aafssLLGcd ‡ ‡ -IWVLG L L E	tgnvtretDF ptelyksSDF .ekiGirDNF .ekiGikDNF .sqiGiqDNF .nkiGvlDNF .eliGitDNF .ercGidDDL .ercGidDDL .dhcGidDDL .dtrGidDDF peeiGiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF qnriGerDDF vqdadaDF \$ SNL	FqqGGDSL1A F1sGGDsupA YsLGGDSIQA YsLGGDSIQA FsLGGHSLKA FeLGGHSLKA FeLGGHSLKA FrLGGDSITS FrLGGDSITS FrCGGDSISS FrCGGDSISS FsLGGDSLKS FsLGGDSLKS FsLGGDSLKS FrLGGHSITC FrLGGHSITC FrLGGHSITC FrLGGHSITC FrLGGHSITC FrLGGHSIC FrLGGHSIC FrLGGHSIC FrLGGHSIC FrLGGHSIC FrLGGHSIC FrLGGSISC FalGGDSLA ¥ ##### # F-LGGDSI-A Y H L S C	trltgqL. ieVvkrc. iqVarL. itlisrm. mtmisqV. maVisqV. tllvakI. hlaaqI. hlvaqI. lhIvaqI. lhIvaqI. lhIvsfKI. tkIsfKI. tkIsfKI. tkIsfKI. iqLiarI. iqLiarI. iqLiarV. mkLaaqIS. * -LI- V V L	<pre>.h qagyeaqlsd .h kagylIklsm hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvrd yr algrkVtvrd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvrn tt gfgqgLgvat rq qlgviIsied rq rlsvsIsved rq rlssLgved .r qvarqVtpgq </pre>	LFnhPrL LYrySTI LLkyPTI LFeaPTI LFetPTI IFetPTI IFdhPT. IFehRTa IYlhRsV LFdhPTV LFshPTI VLGhTTL VFqtkTL VFqtkTL VFqtkTL VFqtkTL VFqtkTL VTqt R LFPTI VL R L IY S V	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3078) (3055) (3019) (2973) (1993) (1993) (1945) (1925) (910) (884) (856) (849) (1035)
HMWP2 AngR TY1 GS1 GS2.dA GS2.dB GS2.dC GS2.dD ACVS.asp.dC ACVS.pen.dC ACVS.pen.dC ACVS.noc.dC ACVS.asp.dB ACVS.noc.dB ACVS.asp.dA ACVS.asp.dA ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dB ACVS.cep.dA ACVS.cep.dA ACVS.cep.dA ACVS.noc.dA EntF Consensus	(1410) E (971) E (533) E (544) E (977) E (2013) E (3059) E (3059) E (2907) E (2907) E (2907) E (2915) E (1934) E (1908) E (1865) E (1865) E (824) E (789) t (977) E E	kqVa kdrVa satLv aetLv aetLv aetLc aetLc actLc siLc siLaLg sdLa sqLa sqLa sqLa sqLa sqLa sqLa sqLa sq	aLWqqLLs kIWndVIG sIWqnVLGi. tIWqdVLGi. eIWhnVLGv. eIWkdVLG1. lIWqeVLG1. lIWqeVLG1. rLWasaLGv. hLWsaqLpg. gIsagLLdis dIWaeLLEmh gIWseLLEip gIWaqVLEia kIWadVLGan aIWgniLsvp aafssLLGcd ‡ ‡ -IWVLG L L E	tgnvtretDF ptelyksSDF .ekiGirDNF .griGiqDNF .nkiGvlDNF .grvGihDDF .ercGidDDL .ercGidDDL .dhcGidDL .gtvGidDDF aqtiGsdSDF periGiySDF pdriGvhDDF hlsiSrkDNF qrsvSrnDNF aqdiGseSNF qnriGerDDF vqdadDF \$ SNL	FqqGGDSL1A FlsGGDaynA YsLGGDSIQA YsLGGDSIAA FsLGGHSLKA FeLGGHSLKA FeLGGHSLKA FrLGGDSIAS FrLGGDSIAS FrLGGDSLKS FsLGGDSLKS FsLGGDSLKS FaLGGDSLKS FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FrLGGHSICC FaLGGHSLAA Y H L S C	trltgqL. ieVvkrc. iqVvarL. itIisrm. mtmisqV. maVisqV. tllvakI. lhlaaqI. lhlvaqI. lhlasqV. tkIsfkI. tkIsfkI. tkIsfaa. iqIiarI. iqIiarI. iqIiarV. mkLaaqIs * -LI- V V L	<pre>.h qagyeaqlsd .h kagylIklsm bs .yqlkLetkd hs .yqlkLetkd nk ecnvdIplrl hk efdveLplkv hk ecqteVplrv ye ymqiemplnv hh qigrkVtvd nn qvgckItvrd nn qvgckItvrd qr algrkVtvkd qr eierkVsvky he vfgrtIsvsa he sfnraVsvsa tr algvaVsvrn t gfqqgLgvat rq qlgviIsied rq rlsvsIsved rq qlqqgItlee rq rlslsLgved .r qvarqVtpgq V L</pre>	LFnhPrL LYrySTI LLnyPTI LFeaPTI LFetPTI IFehPT. IFehPT. IFehPT. IFhRT. LFchRTV LFshPTI VLghtTL VFssRT. VFatRTL VFqtkTL VFqtkTL VFqtkTL VFqtkTL VFqtkTL IF-PTI VL R L IY S V	(1467) (1028) (590) (601) (1035) (2071) (3117) (4155) (3055) (3055) (3055) (3055) (3055) (3055) (1993) (1993) (1993) (1993) (1945) (1945) (1945) (1945) (1945) (884) (884) (856) (849) (1035)

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:

- (Y,F)(T,S)SG(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)GXX(V,I,L)(A,S,T)(R,K,E)GY
- IX (I,V,L)E(Y,F)LGRXDXQ(V,I,L)K(I,V,L)RGXR(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 (≥ 28 of 30) or of family A (≥ 21 of 22). CS I 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 ACVSs, 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, KX(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 (<u>OLTFTEDW</u>) 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 ironregulated protein HMWP2 is an ORF of 6,126 nt encoding a protein of approximately 228,000 Da. A typical ribosomebinding 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 ACVSs 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-adenylateenzyme 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 β -lactam antibiotics, formation of the tripeptide δ -(L- α -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'-phosphopantheteine 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'-phosphopantheteine 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 versiniabactin (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. Tekaïa) 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.
- 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.
- 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.
- 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.
- Carniel, E., A. Guiyoule, I. Guilvout, and O. Mercereau-Puijalon. 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.
- Carniel, E., A. Guiyoule, O. Mercereau-Puijalon, 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.
- Carniel, E., O. Mercereau-Puijalon, 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.
- Conchas, R. F., and E. Carniel. 1990. A highly efficient electroporation system for transformation of *Yersinia*. Gene 87:133– 137.
- 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 lactamdurans* are clustered together in an organization different from the same genes in *Acremonium chrysogenum* and *Penicillium chrysogenum*. Mol. Microbiol. 5:1125–1133.
- Crosa, J. H. 1989. Genetics and molecular biology of siderophore-mediated iron transport in bacteria. Microbiol. Rev. 53:517-530.
- de Almeida, A. M. P., A. Guiyoule, I. Guilvout, I. Iteman, G. Baranton, and E. Carniel. 1993. Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. Microb. Pathog. 14:9-21.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387–395.
- 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 a-aminoadipyl-cysteinyl-valine synthetase and linkage to the *pcbC* and *penDE* genes. J. Biol. Chem. 265:16358-16365.
- 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.
- Driscoll, J. R., and H. W. Taber. 1992. Sequence organization and regulation of the *Bacillus subtilis menBE* operon. J. Bacteriol. 174:5063-5071.
- 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 *Methanothrix 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.
- Favaloro, 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 α-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the *pcbC* gene as a cluster of early cephalosporin biosynthesic genes and evidence of multiple functional domains. J. Bacteriol. 173:2354–2365.
- Harley, C. B., and R. P. Reynolds. 1987. Analysis of E. coli promoter sequences. Nucleic Acids Res. 15:2343-2361.
- Heesemann, J. 1987. Chromosomal-encoded siderophores are required for mouse virulence of enteropathogenic *Yersinia* species. FEMS Microbiol Lett. 48:229–233.
- 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.
- Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237-244.
- 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.
- 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.
- 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 TonBdependent 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 hydropathic 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.
- Loyoza, 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.
- 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. γ-(L-α-aminoadipyl)-L-cysteinyl-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.
- 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.
- 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.
- Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85:2444-2448.
- 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.
- Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13:319–353.
- 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.
- 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.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406-425.
- 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.
- Schaffer, M. H., B. E. Noyes, C. A. Slaughter, G. C. Thorne, and S. J. Gaskell. 1990. The fruitfly *Drosophila melanogaster* contains a novel charged adipokinetic-hormone-family peptide. Biochem. J. 269:315-320.
- 52. Schlumbohm, W., T. Stein, C. Ullrich, J. Vater, M. Krause, M. A. Marahiel, V. Kruft, 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.
- Schwyn, B., and J. B. Neilands. 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.
- Silver, S., and M. Walderhaug. 1992. Gene regulation of plasmid- and chromosome-determined inorganic ion transport in bacteria. Microbiol. Rev. 56:195-228.
- 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.
- Staggs, T. M., and R. D. Perry. 1992. Fur regulation in Yersinia species. Mol. Microbiol. 6:2507-2516.
- 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.
- 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.
- 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 en-

coding a novel *trans*-activator of siderophore biosynthesis. J. Bacteriol. **170**:1913–1919.

- 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.
- 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.
 Wee, S., J. B. Neilands, M. L. Bittner, B. C. Hemming, B. L.
- 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.
- 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.