J. Biochem. 125, 1086-1093 (1999)

Sequence of the *V. parahaemolyticus* Gene for Cytoplasmic *N,N'*-Diacetylchitobiase and Homology with Related Enzymes[1]

Michael H. Wu* and Roger A. Laine*,†,[2]

*Departments of *Biological Sciences, and †Chemistry, Louisiana State University and A&M College and The LSU Agricultural Center, Baton Rouge, LA 70803, USA*

Received February 13, 1999; accepted March 12, 1999

The nucleotide sequence of the gene encoding the cytoplasmic *N,N'*-diacetylchitobiase [EC 3.2.1.14] from *Vibrio parahaemolyticus* (ATCC #27969) has been determined. The deduced peptide sequence of this unusual β-hexosaminidase surprisingly shows minimum evolutionary relationship to two other reported *N,N'*-diacetylchitobiases from vibrios, except in highly conserved regions which are also homologous to lysosomal β-hexosaminidases from eukaryotes including humans. In contrast, the two other β-hexosaminidases from vibrios with reported sequences are much more closely related to each other. This novel 85 kDa cytoplasmic glycosyl hydrolase with restricted specificity participates in the high level utilization of chitin-derived 2-deoxy-2-acetamido-D-glucose (GlcNAc) by vibrios as one of two parallel pathways for the metabolism of *N,N'*-diacetylchitobiose [Bassler, B.L., Yu, C., Lee, Y.C., and Roseman, S. (1991) *J. Biol. Chem.* 266, 24276-24286]. These pathways use chitin-binding proteins for the adherence of the bacterial chitinase to the substrate, and an extracellular chitinase and a periplasmic chitodextrinase to produce *N,N'*-diacetylchitobiose. The *V. parahaemolyticus* cytoplasmic *N,N'*-diacetyl-chitobiase reported herein appears to be a unique protein, lacking a signal sequence, and genetically distant from other known chitinoclastic β-*N,N'*-diacetyl-hexosaminidases. This is consistent with its limited substrate specificity to small GlcNAc terminated oligosaccharides and its cytoplasmic rather than periplasmic localization.

Key words: Chitin degradation, cytoplasmic chitobiase, DNA sequence, *Vibrio parahaemolyticus*, homology.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

Two parallel pathways have been postulated in marine vibrios for the catabolism of chitin, possibly comprising as many as 6-10 enzymes and a number of chemotactic proteins (*1*). In the common part of the pathway, chitin-binding proteins adhere to the substrate (*2, 3*), and extracellular chitinase (*4*) and periplasmic chitodextrinase work together to produce *N,N'*-diacetylchitobiose [$(GlcNAc)_2$] (*5*). In one branch of the pathway, the glycosidase/PTS system cleaves *N,N'*-diacetyl chitobiose to 2-deoxy-2-acetamido-D-glucose (GlcNAc) in the periplasmic space *via* a membrane bound chitobiase (*5-7*) after which GlcNAc is transported and phosphorylated by the PTS (*5*). The second branch, a parallel permease/glycosidase system, resembles the *Escherichia coli lac* permease/β-galactosidase system (*8, 9*), and utilizes an, as yet, unidentified *N,N'*-diacetylchitobiose permease to transport the substrate to the cytoplasm. The transported $(GlcNAc)_2$ is cleaved by the cytoplasmic chitobiase reported here (*10*) and phosphorylated by an ATP-dependent *N*-acetyl-D-glucosamine kinase (*5, 11, 12*). The cytoplasmic system works independently of the PTS (*5*). This report provides the nucleotide sequence and the deduced polypeptide sequence of the gene encoding the cytoplasmic chitobiase [EC 3.2.1.14] from *Vibrio parahaemolyticus* (ATCC #27969), and shows an ancient evolutionary divergence for this unique β-hexosaminidase compared with periplasmic chitobiases.

MATERIALS AND METHODS

Host Bacterial Strains, Vectors, and Phages—*E. coli* strains DH5α and DH5αF' were purchased from Gibco Bethesda Research Laboratories (Gaithersberg, MD). *E. coli* strain JM101, phagemid vectors pBluescript II KS^+, pBluescript II SK^+, and the interference resistant helper phages VCSM13 (kan^r) and R408 were from Stratagene (La Jolla, CA). The plasmid harboring the *V. parahaemolyticus* (ATCC#27969) chitobiase gene, PC120, was constructed in this laboratory and has been described previously (*10*).

Enzymes, Chemicals, and Antibiotics—Restriction endonucleases, T4 DNA ligase, and T4 DNA polymerase were purchased from Gibco Bethesda Research Laboratories (Gaitherberg, MD), New England Biolabs (Beverly, MA), or United States Biochemicals (Cleveland, OH). The T7

[1] The DNA sequence data published here have been deposited with the GenBank/EMBL sequence data bank and are available under accession number U24658 (VPCHB).

[2] To whom correspondence should be addressed at: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. Phone: +1-225-388-4691, Fax: +1-225-388-4695, E-mail: rlaine@usa.net

© 1999 by The Japanese Biochemical Society.

DNA polymerase DNA sequencing kit (Sequenase 2.0) was from United States Biochemicals Corp. (Cleveland, OH). ^{35}S-Deoxyadenosine 5'-[α-thio]-triphosphate was from Du Pont-NEN (Boston, MA). Ampicillin was from Sigma Chemical (St. Louis, MO). Oligonucleotides used as sequencing primers were synthesized on an automated DNA/RNA synthesizer (Applied Biosystems, Model *abi* 394) in GeneLab, School of Veterinary Medicine, Louisiana State University (Baton Rouge, LA).

Construction of Subclones for Sequencing—The chitobiase gene harboring plasmid PC120 (*10*) was digested with the restriction enzymes *Pst*I, *Sac*I, and *Hin*dIII. The 1.6, 2.1, and 3.5 kbp fragments were gel purified and cloned into appropriate sites of pBluescript II SK$^+$ vector for single strand DNA production. Insert orientations of the clones were identified by digestion with appropriate restriction enzymes that cut the inserts asymmetrically (*Hin*dIII for the 2.1 kbp fragment and *Eco*RV for the 3.5 kbp fragment), and the clones thus obtained were named SKS162, SKP21-1, SKP21-2, SKH35-1, and SKH35-2 (Fig. 1). The protocol for rescuing recombinant phagemid using VCSM13 was from Stratagene (La Jolla, CA), except that *E. coli* DH5αF' was used instead of XL1-Blue, and Luria Bertanic medium (LB medium) was used instead of Super-Broth.

Sequence Analysis—Sequencing gel data were assembled and analyzed using Staden's algorithm (*13*), an integrated part of the GCG software, Unix® version 8.0 (Genetic Computing Group, Madison, WI). The identified chitobiase open reading frame (ORF) and the predicted polypeptide sequence were then used as primary query sequences to search available nucleic acid and protein depository databases. The secondary structures of chitobiases from *V. parahaemolyticus*, other vibrios, as well as higher organisms were predicted using GCG programs based on the report of Chou and Fasman (*14*).

RESULTS

Nucleotide Sequence Determination—The sequencing clones and the orientations of the inserts are shown in Fig. 1A. The two phagemids, SK21-1 and SK21-2, contain the 2.1 kbp *Pst*I fragments but in opposite orientations. Both SKS162 and SKH351 overlap with the above two clones to facilitate the assembly of the sequencing gel data. The restriction map of PC120 as determined previously (*10*) and from the sequence data is depicted in Fig. 1B. Using the Staden (*13*) algorithm as part of the GCG package on the Unix platform, a single ORF was identified from sequence data from the clones shown in Fig. 1. The area sequenced and the coding region of *V. parahaemolyticus* chitobiase are also shown. The 47 amino acid residues at the amino terminus of the polypeptide are located between the *Pst*I and *Sac*I restriction sites, indicating that the expression of the chitobiase gene on PC120 is driven by the *V. parahaemolyticus* chitobiase gene promoter instead of the *lac* promoter on pUC18, the parental cloning vector. The nucleotide sequence and the deduced amino acid sequence of the gene are shown in Fig. 2. The amino terminal sequence of the polypeptide from Edman degradation (*10*) perfectly matched that deduced from the DNA sequence, suggesting the lack of a signal peptide at the amino terminus. This agrees with the cytoplasmic localization of the enzyme.

By attaching a signal peptide upstream of the first initiator AUG methionine codon, the enzyme has been shown to permeate the two membranes of Gram negative *E. coli* (to be published elsewhere), further indicating that the wild type chitobiase of *V. parahaemolyticus* is cytoplasmic. The molecular mass of the deduced 741 aa polypeptide is 85 kDa, agreeing with that determined by SDS-PAGE (*10*).

Sequence Homology with Other Chitobiases—The deduced polypeptide sequence of *V. parahaemolyticus* was used as a query sequence to search the GenBank/EMBL and Swiss-Prot genetic data bases using FASTA, a GCG program based on the algorithm of Pearson and Lipman (*15*). Limited homologies, aligning in a 60 amino acid area of a composite map from residues 341-400, were found between this enzyme and those of *V. harveyi*, an outer membrane protein (*7, 16*), and *V. vulnificus* (*17*). Except for lysosomal β-*N*-acetylhexosaminidases from human and rabbit, which cleave GlcNAc from a number of glycosides and are related to each other (*18*), a remarkable homology was found among 31 out of 930 amino acids to all sequenced β-hexosaminidases in the composite map locations 296-325, 341-359, 380-400, and 458-465 (Fig. 3). Especially interesting in this region is the arginine residue at position 359 conserved in all enzymes listed except lysosomal chitobiases from human and rabbit. It has been shown (*19*) that Arg^{178} and Arg^{211} (aligned at position 359 in Fig. 3) in the α- and β-subunits of human β-hexosaminidase, respectively, are "active" residues, part of the catalytic sites, but do not participate in substrate binding. This highly conserved region may be associated with the enzyme active sites. Figure 3 also shows that periplasmic

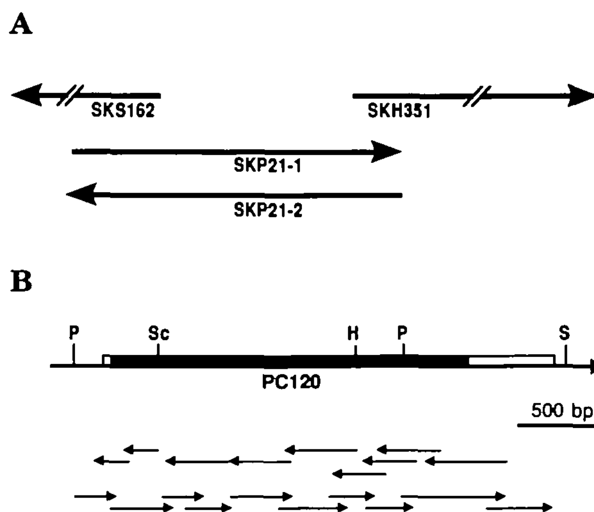

Fig. 1. **Restriction map and nucleotide sequencing strategy of the sequenced region containing the *N,N'*-diacetylchitobiase gene from *V. parahaemolyticus*.** (A) Inserts in subclones used to generate single stranded DNA for sequencing. Arrows represent insert orientation relative to the chromosomal DNA fragment in PC120 (*10*), arbitrarily determined as that of the *lacZ'* on pUC18. (B) Restriction sites used to generate the inserts shown above and the sequencing strategy. Arrows below the restriction map are individual sequencing gel data used in sequence assembly. The sequence presented in this paper is boxed. The solid portion indicates the chitobiase ORF. Only restriction sites used for subcloning are shown. P, *Pst*I; S, *Sal*I; Sc, *Sac*I; H, *Hin*dIII.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

```
                                                TGGAGCTCACGCGTGCGCCGCTCTAGACTAGTGG  -177
ATCCCCGGCTGCAGCTGATATAGAAATGTTTCTCGTGGCATTAAACAAGCGCGGCGCAACGCGTATTTGTTTAATGGGC  -98

                                                                          -35
AGCATTGCTGAACGTATTTTACGTTGGCTTTCGCACCAGTTCAGCAATGGGTTGTTGAGCACAGTTTGACGCGATTGGG  -19

        -10             +1                                               SD
CGC AAT TAT GTT GCA GGT AAC TGA CAT ACT TTA CTC GCT CAG TGC GGC GTA GGA GTA GAT   42

ATG GAA TAT CGT GTT GAT CTC GTC GTC CTA TCA GAA CAA AAG CAA AAC TGC CGT TTC GGA  102
Met Glu Tyr Arg Val Asp Leu Val Val Leu Ser Glu Gln Lys Gln Asn Cys Arg Phe Gly   20

CTG ACT TTC CAT AAT TTG AGC GAT CAA GAT CTC CAC AAT TGG AGC CTG ATT TTT GCT TTT  162
Leu Thr Phe His Asn Leu Ser Asp Gln Asp Leu His Asn Trp Ser Leu Ile Phe Ala Phe   40

GAT CGC TAC ATT CTG CCG GAT AGT ATT TCG AAT GGT CAG CTC AAG CAA ATT GGC AGC TAC  222
Asp Arg Tyr Ile Leu Pro Asp Ser Ile Ser Asn Gly Gln Leu Lys Gln Ile Gly Ser Tyr   60

TGC ACC CTC AAA CCA GAA GGG TTG GTG CTG GCA GCT AAC CAT CAT TTT TAC TGC GAG TTC  282
Cys Thr Leu Lys Pro Glu Gly Leu Val Leu Ala Ala Asn His His Phe Tyr Cys Glu Phe   80

AGT ATT GGT TCG AAC CCA TTC CGT TAC TAT TCT GAT GGA TTC AAT GAA GCC TTG GTC AAC  342
Ser Ile Gly Ser Asn Pro Phe Arg Tyr Tyr Ser Asp Gly Phe Asn Glu Ala Leu Val Asn  100

TTT GAA GTC AAC GGC AAC CTT CAG CGA GCT CAA GTC GAT GTC ACG CCG ATC GTA TTG GCT  402
Phe Glu Val Asn Gly Asn Leu Gln Arg Ala Gln Val Asp Val Thr Pro Ile Val Leu Ala  120

TCA CCG TAC CGT GAA CGT AGT GAG ATC CCT TCC AGC TTG ACG CAT GCG CAC GCT TTG TTG  462
Ser Pro Tyr Arg Glu Arg Ser Glu Ile Pro Ser Ser Leu Thr His Ala His Ala Leu Leu  140

CCA AAA CCA AAC CAT ATA GAA GTC AGC GAT CAC TGC TTT AGC TTT AAT CAT CAC GCC GGC  522
Pro Lys Pro Asn His Ile Glu Val Ser Asp His Cys Phe Ser Phe Asn His His Ala Gly  160

GTT GCG GTT TAT TCA AAC CTA GCC AAT TCA GCT AAA GAG TGG TTA CTT GAA GAG CTT AAG  582
Val Ala Val Tyr Ser Asn Leu Ala Asn Ser Ala Lys Glu Trp Leu Leu Glu Glu Leu Lys  180

CGC ATT CAT CAA TTT GAG TTC GCA TCA GAC AAT GGC AGT CAG ATC ATC TTC AAA GGC AAC  642
Arg Ile His Gln Phe Glu Phe Ala Ser Asp Asn Gly Ser Gln Ile Ile Phe Lys Gly Asn  200

CCA ACC TTG GAT GAA GGC GCT TAC AAG CTG AAG GTA GCA GAA GAG TCG ATC AAA ATT GAA  702
Pro Thr Leu Asp Glu Gly Ala Tyr Lys Leu Lys Val Ala Glu Glu Ser Ile Lys Ile Glu  220

GCA GGC TCT TCG TCT GGT TTT ACC CAT GCT TGT GCA ACG TTG TTG CAA CTG ATC AAA GTT  762
Ala Gly Ser Ser Ser Gly Phe Thr His Ala Cys Ala Thr Leu Leu Gln Leu Ile Lys Val  240

GGC GAT CAA CCA GCC TCA ATG GAA GTG GTT TGC TGT TCA ATC AAA GAC CGA CCA CGT TTT  822
Gly Asp Gln Pro Ala Ser Met Glu Val Val Cys Cys Ser Ile Lys Asp Arg Pro Arg Phe  260

CGT TAC CGC GGT ATG ATG CTA GAT TGT GCT CGC CAT TTT CAC TCC GTT GAG CAA GTC AAA  882
Arg Tyr Arg Gly Met Met Leu Asp Cys Ala Arg His Phe His Ser Val Glu Gln Val Lys  280

CGT TTG ATC AAC CAG TTG GCT CAC TAC AAG TTC AAT ACA TTC CAT TGG CAC CTT ACC GAT  942
Arg Leu Ile Asn Gln Leu Ala His Tyr Lys Phe Asn Thr Phe His Trp His Leu Thr Asp  300

GAT GAA GGT TGG CGA ATT GAG ATC AAA TCA TTG CCT CAA CTA ACC GAT ATT GGC GCA TGG 1002
Asp Glu Gly Trp Arg Ile Glu Ile Lys Ser Leu Pro Gln Leu Thr Asp Ile Gly Ala Trp  320

CGT GGG TTG GAT GAA ACC AAT GAG CCA CAG TAC TCG CAC CTT GCT GAG CGG TTA CGG CGG 1062
Arg Gly Leu Asp Glu Thr Asn Glu Pro Gln Tyr Ser His Leu Ala Glu Arg Leu Arg Arg  340

TTT TTA CAC TCA AGA AGA CAT CAA AGA CGT GGT TGC CTT TGC TTC GAA ACG AGG CAT CAC 1122
Phe Leu His Ser Arg Arg His Gln Arg Arg Gly Cys Leu Cys Phe Glu Thr Arg His His  360

TGT TAT CCT GAA ATC GAT GTA CCA GGG CAC TGC CGA GCT GCC ATC AAG TCG TTA CCA CAC 1182
Cys Tyr Pro Glu Ile Asp Val Pro Gly His Cys Arg Ala Ala Ile Lys Ser Leu Pro His  380

CTA TTG GTA GAA GCA GAA GAT ACC ACC GAA TAC CGC AGC ATT CAG CAT TAC AAC GAC AAC 1242
Leu Leu Val Glu Ala Glu Asp Thr Thr Glu Tyr Arg Ser Ile Gln His Tyr Asn Asp Asn  400

AAA GTT TTG GAA GAA GTC TCG CGT TGT GTC ATT AAC CCA GCT CTG CCG GGG AGC TAT GAG 1302
Val Ile Asn Pro Ala Leu Pro Gly Ser Tyr Glu Phe Ile Asp Lys Val Leu Glu Glu Val  420

TTT ATC GAT TCC CTG CCC CTT ATG TTC ATA TCG GTG CGG ATA AGT ACT AAC GGC GTA TGG 1362
Ser Arg Cys Ser Leu Pro Leu Met Phe Ile Ser Val Arg Ile Ser Thr Asn Gly Val Trp  440

TCA AAA AGC CCT GCA TGC CAA GCA CTA ATG GAA CAA CTG GGT TAC AGC GAC TAC AAA GAG 1422
Ser Lys Ser Pro Ala Cys Gln Ala Leu Met Glu Gln Leu Gly Tyr Ser Asp Tyr Lys Glu  460

TTA CAA GGG CAC TTC TTG CGT CAT GCC GAA GAC AAA CTG CGC AAA CTT GGC AAG CGC ATG 1482
Leu Gln Gly His Phe Leu Arg His Ala Glu Asp Lys Leu Arg Lys Leu Gly Lys Arg Met  480
```

Fig. 2 (continued on next page)

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

```
CTG GGT TGG GAA GAA GCA CAG CAT GGC GAC AAA GTC AGC AAA GAC ACA GTG ATC TAT TCG  1542
Leu Gly Trp Glu Glu Ala Gln His Gly Asp Lys Val Ser Lys Asp Thr Val Ile Tyr Ser   500

TGG TTA AGC GAA GAA GCG GCG TTG AAC TGC GCG CGC CAA GGT TTC GAT GTG GTG CTA CAA  1602
Trp Leu Ser Glu Glu Ala Ala Leu Asn Cys Ala Arg Gln Gly Phe Asp Val Val Leu Gln   520

CCT GCG CAA ACC ACC TAC TTA GAT ATG ACC CAA GAT TAC GCA CCA GAA GAA CCG GGC GTG  1662
Pro Ala Gln Thr Thr Tyr Leu Asp Met Thr Gln Asp Tyr Ala Pro Glu Glu Pro Gly Val   540

GAT TGG GCT AAC CCA TTG CCG CTA GAA AAA GCT TAC AAC TAT GAA CCA CTC GCT GAA GTG  1722
Asp Trp Ala Asn Pro Leu Pro Leu Glu Lys Ala Tyr Asn Tyr Glu Pro Leu Ala Glu Val   560

CCA GCC GAC GAT CCA ATA CGT AAA CGC ATT TGG GGC ATT CAA ACA GCA TTG TGG TGC GAA  1782
Pro Ala Asp Asp Pro Ile Arg Lys Arg Ile Trp Gly Ile Gln Thr Ala Leu Trp Cys Glu   580

ATC ATC AAC AAC CAG TCT CGT ATG GAC TAC ATG GTC TTC CCG CGC TTA ACC GCA ATG GCA  1842
Ile Ile Asn Asn Gln Ser Arg Met Asp Tyr Met Val Phe Pro Arg Leu Thr Ala Met Ala   600

GAA GCA TGT TGG ACA GAC AAG CAA CAC CGA GAC TGG ACC GAC TAT TTA TCA CGT TTG AAA  1902
Glu Ala Cys Trp Thr Asp Lys Gln His Arg Asp Trp Thr Asp Tyr Leu Ser Arg Leu Lys   620

GGA CAC CTA CCG CTG CTT GAT TTG CAG GGA GTG AAT TAC CGT AAC CGT GGA AGT AAT ACA  1962
Gly His Leu Pro Leu Leu Asp Leu Gln Gly Val Asn Tyr Arg Asn Arg Gly Ser Asn Thr   640

GAG CAT TGT AGT AGA AGC ATC ACG CTT GAA GAG TTT TTA AAT TTT GGC TGC AGA CGC AGC  2022
Glu His Cys Ser Arg Ser Ile Thr Leu Glu Glu Phe Leu Asn Phe Gly Cys Arg Arg Ser   660

TTT GTA AAA AGG AAT ACA CAA ATG AAA TAC GGC TAT TTC GAT AAC GAG AAT CGT GAA TAC  2082
Phe Val Lys Arg Asn Thr Gln Met Lys Tyr Gly Tyr Phe Asp Asn Glu Asn Arg Glu Tyr   680

GTC ATT ACT CGC CCT GAT GTA CCT GCT CCT TGG ACC AAC TAC CTA GGT ACA GAA AAA TTC  2142
Val Ile Thr Arg Pro Asp Val Pro Ala Pro Trp Thr Asn Tyr Leu Gly Thr Glu Lys Phe   700

TGT ACC GTT ATC TCG CAT AAC GCA GGT GGC TAT TCG TTC TAC AAC TCT CCA GAA TAC AAC  2202
Cys Thr Val Ile Ser His Asn Ala Gly Gly Tyr Ser Phe Tyr Asn Ser Pro Glu Tyr Asn   720

CGT GTT ACT AAG TTC CGT CCA AAT GCG ACA TTT CGA TCG CCC AGG ACA CTA CGT TTA CCT  2262
Arg Val Thr Lys Phe Arg Pro Asn Ala Thr Phe Arg Ser Pro Arg Thr Leu Arg Leu Pro   740

ACG TGA TGA TGA GACGGGAGATTACGGTCAATCTCTTGGCAACCAGTTGCAAAAGCCTAGACGAAGCGAACTACG  2337
Thr Ter Ter Ter                                                                   741

AAGTTCGTCATGGTTTGTCGTACTCTAAATTCAAGTGTGAATACAGCGGCATTAGCGCAACCAAAACGCTCTTTGTACC  2416
AAAAGGCGAAGATGCAGAAATTTGGGATGTGGTCATCAAGAACACCTCTGACAAACCGCGTACGATCAGTGCATTCTCA  2495
TTTGTTGAGTTCTCGTTCAGCCACATTCAATCAGATAACCAAAACCACCAGATGTCTCTCTACTCTGCTGGTACGTCAT  2574
ACAACACAGGCGTGTTGAATACGACCTGTACTACAACACTAACGATTCGAAGGCTTCTACTACCTAGCGTCAACGTTTG  2653
ATCCAGATTCATACGACGGTCCAACGTGATAGCTTCTAGGTCTATACCGCGACGAAGCAAACCCACTAGCAGAGTAGAA  2732
CAGGTAAGTGTTTACACGCACAACGTGTTACACCACTGTGGCTTTGCACGCATT                           2786
```

Fig. 2. **Nucleotide sequence encoding the *N,N′*-diacetylchitobiase gene of *V. parahaemolyticus*.** The figure shows the coding strand (non-template strand) of the DNA. Putative regulatory elements and the amino terminal sequence of the predicted polypeptide as determined by Edman degradation (*10*) are underlined. Nucleotide numbering is based on the putative transcription start site (+1). −35 and −10: −35 and −10 regions of the promoter, respectively; SD: Shine-Dalgarno site or ribosome binding site (RBS). Note the three consecutive translation Opal stop codons (TGA) at the end of the ORF.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

chitobiases from *V. harveyi* and *V. vulnificus* show extensive homology with each other, while their homology with the cytoplasmic chitobiase from *V. parahaemolyticus* is much lower. This implies that cytoplasmic chitobiases from vibrios took a very early and different line of evolution than periplasmic chitobiases, and that these signal-sequence-containing enzymes are more closely related to β-hexosaminidases from higher organisms.

The structural gene and the deduced amino acid sequence of the *V. parahaemolyticus* chitobiase were progressively piled up (*20, 21*) to those of chitobiases and β-hexosaminidases from other organisms including other vibrios and higher organisms. The results are shown as a sequence alignment (Fig. 3A) as well as a dendrogram (Fig. 3B). Relationships among enzymes from these organisms were obtained using either the DNA or amino acid sequences for comparison (only the amino acid data are shown). The clustering relationships as visualized in Fig. 3B show the uniqueness of the cytoplasmic chitobiase from *V. parahaemolyticus* among the chitobiases and β-hexosaminidase from all three vibrios.

Comparison of Secondary Structures among Chitobiases from Different Organisms—The secondary structures of chitobiases from various species, including human and other vibrios, as predicted using the method of Chou and Fasman are shown in Fig. 4. Although very little amino acid identity was found in the linear polypeptide sequences of chitobiases from *V. harveyi*, *V. vulnificus*, and *Homo sapiens*, chitobiases from these three organisms ranging from prokaryotes to highly evolved eukaryotes, seem similar in their general secondary structure. The Chou-Fasman plots remarkably show nearly a mirror image in overall patterns of secondary and turn structures between human and the closely related pair of *V. harveii* and *V. vulnificus*, although the amino/carboxyl terminus orientation is opposite. *V. parahemolyticus*, on the other hand, shows a plot closer to that of the human enzyme in its secondary patterning and amino/carboxyl orientation. This indicates a common evolutionary ancestor for these enzymes among those organisms, as suggested for other systems by Somerville and Colwell (*17*). In contrast with these three chitobiases, the cytoplasmic chitobiase from *V. parahaemolyticus* does not seem to follow the secondary structural pattern, underscoring its separate line of evolution.

DISCUSSION

Roseman *et al.* showed that *V. furnissii* possesses a separate *N,N′*-diacetylchitobiase with a cytoplasmic localiza-

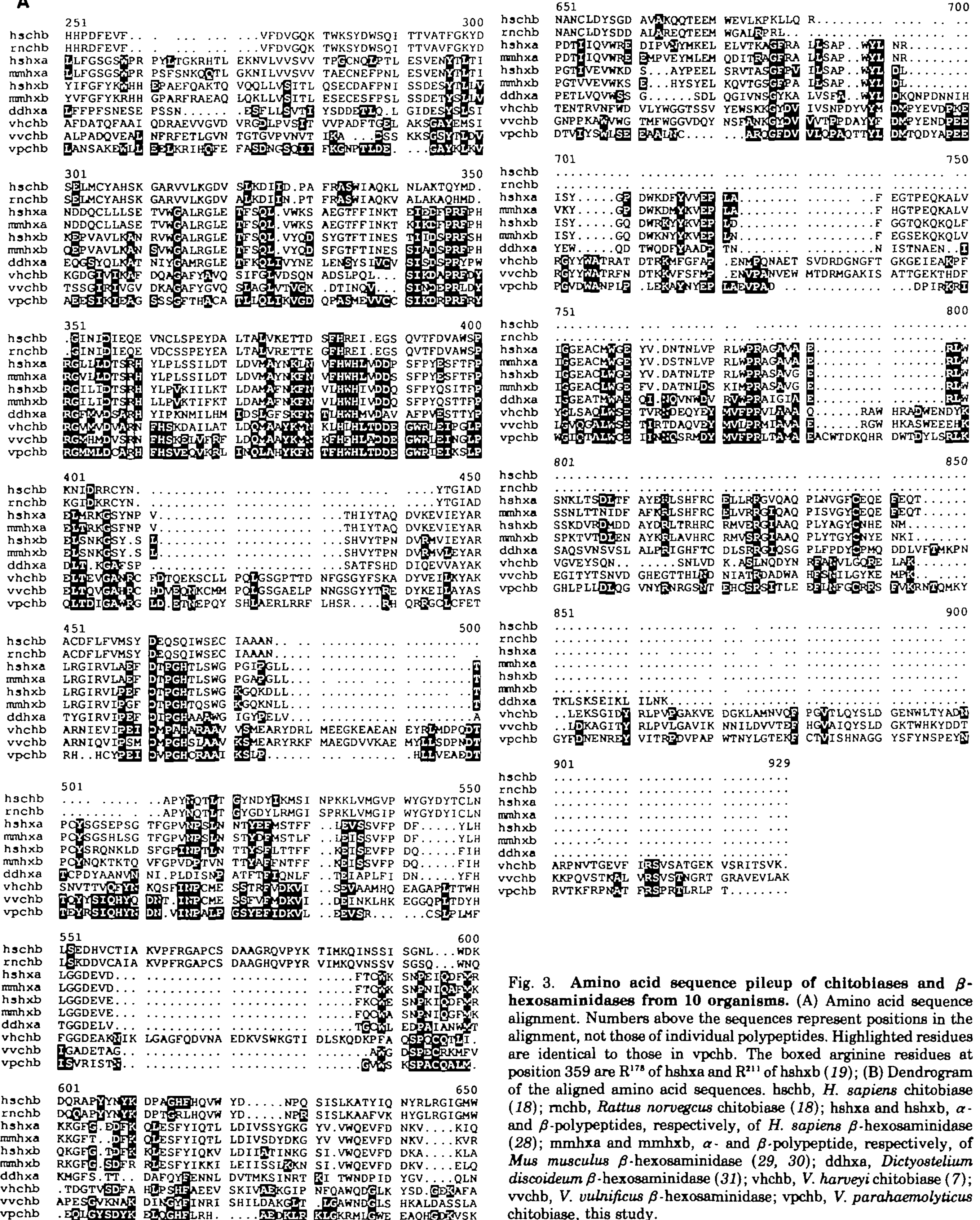

Fig. 3. **Amino acid sequence pileup of chitobiases and β-hexosaminidases from 10 organisms.** (A) Amino acid sequence alignment. Numbers above the sequences represent positions in the alignment, not those of individual polypeptides. Highlighted residues are identical to those in vpchb. The boxed arginine residues at position 359 are R^{178} of hshxa and R^{211} of hshxb (*19*); (B) Dendrogram of the aligned amino acid sequences. hschb, *H. sapiens* chitobiase (*18*); rnchb, *Rattus norvegcus* chitobiase (*18*); hshxa and hshxb, α- and β-polypeptides, respectively, of *H. sapiens* β-hexosaminidase (*28*); mmhxa and mmhxb, α- and β-polypeptide, respectively, of *Mus musculus* β-hexosaminidase (*29, 30*); ddhxa, *Dictyostelium discoideum* β-hexosaminidase (*31*); vhchb, *V. harveyi* chitobiase (*7*); vvchb, *V. vulnificus* β-hexosaminidase; vpchb, *V. parahaemolyticus* chitobiase, this study.

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

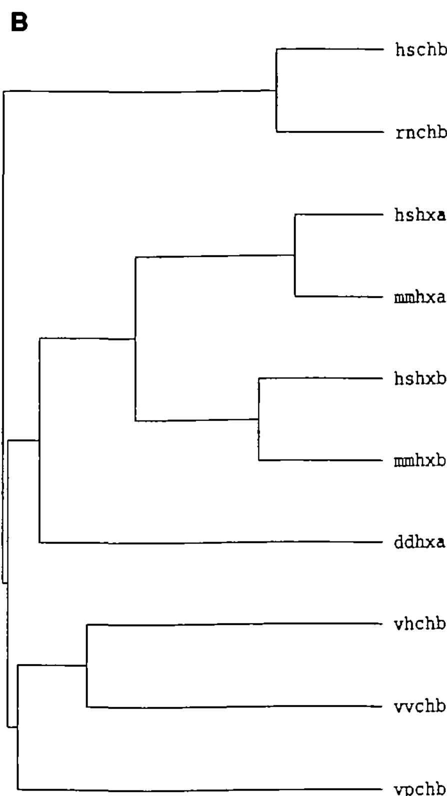

Fig. 3B

tion that depends on a $(GlcNAc)_2$ permease (*1, 2, 5*). We isolated a similar *N,N′*-diacetylchitobiase from the cytoplasm of *V. parahaemolyticus* (*10*). The lack of signal peptide at the amino terminus in the gene confirms this assignment. It appears that a hydrophobic patch is absent throughout the entire length of the polypeptide sequence. In addition, a naturally secreted endochitinase has been characterized and the gene cloned from *V. parahaemolyticus* (*4*). The limited digest product of chitin cleavage by this chitinase is *N,N′*-diacetylchitobiose, the substrate of periplasmic and cytoplasmic *N,N′*-diacetylchitobiases. A chitin binding protein that may facilitate *V. parahaemolyticus* adherance to chitin has also been reported (*3*). The involvement of a permease is supported by *in vitro* data from *V. furnisii* (*5*). Complementation analysis (*22*) showed that a *lac* permease defective *E. coli* strain (LE392, $lacY^-$ mutant) carrying plasmids with chromosomal DNA fragment from *V. harveyi* are able to hydrolyze *o*-nitrophenyl-D-galactoside (ONPG), implying that the putative chitopermease is equivalent in function to the *lacY* gene product and is able to translocate lactose across the cytoplasmic membrane of *E. coli*. Identification and characterization of the permease protein in *V. parahaemolyticus* will establish the complete chitinoclastic pathway in this organism.

Although it has been shown that the PTS exists in *V. parahaemolyticus* (*23*) and all vibrios tested (*23, 24*), the periplasmic form of chitobiase has not yet been isolated from *V. parahaemolyticus* (*23*).

It is worth noting that sequences in the −10 and −35 region of the *V. parahaemolyticus* chitobiase gene promoter are nearly identical to the prokaryotic consensus promoter sequences (5′-TATAAT-3′ for −10 and 5′-TTGACA-3′ for −35) (*25, 26*), especially in the −35 region where the *V. parahaemolyticus* chitobiase gene has a G instead of an A as in the consensus sequence, suggesting the high expression of the gene driven by its own promoter. This agrees with the high yield of the enzyme using PC120, the clone with the chromosomal insert from *V. parahaemolyticus*. On the other hand, the chitobiase level in *V. parahaemolyticus* itself is at least 30-fold lower (Zhu, B.C., personal communication) than in PC120. One possibility is that the expression of the chitobiase gene is regulated in the original organism but not when cloned into *E. coli*, although other alternatives such as a higher gene dosage in *E. coli* carrying PC120 may also play a role. In contrast to the chitin degradation systems of *V. harveyi* and *V. vulnificus*, in which catalytic and non-catalytic proteins are organized as simple *chi*-operons (*6, 22*), the secreted chitinase and cytoplasmic chitobiase genes in *V. parahaemolyticus* do not seem to be in the same cistron, since both have their own promoters and ribosome binding sites; nor are they in close vicinity, since sequences 1,000 nt upstream and 500 nt downstream of the chitobiase ORF do not overlap with the chitinase gene (data not shown) or other sequences known in the *Vibrio chi* operons. The regulatory effect involved in the chitobiase gene of *V. parahaemolyticus* may be *in trans*.

Figure 3A reveals two regions highly conserved among all chitobiases and hexosaminidases except for human and rabbit lysosomal chitobiases, which are exoglycosidases that split the GlcNAc-β-D-(1-4)GlcNAc chitobiose core of asparagine-linked glycoproteins (*18, 27*). It is highly likely that amino acid residues in one or both of these regions participate in the catalysis of their respective substrates, especially the region from position 341-400 that includes the catalytic arginine residues in the α- and β-subunits of human β-hexosaminidase (*19*).

This report shows the uniqueness of the cytoplasmic chitobiase gene of *V. parahaemolyticus* compared with other chitobiase genes cloned from vibrios or other organisms. A cytoplasmic chitobiase activity was proposed by Roseman's group (*5*) and isolated and cloned by Zhu *et al.* (*10*). The sequence and homology comparisons in this report establish an evolutionary relationship among similar enzymes and elaborate the extent of genetic investment in chitin degradation by vibrios.

This work is published as part of the requirement for the Ph.D. dissertation research of MHW.

REFERENCES

1. Yu, C., Bassler, B., and Roseman, S. (1993) Chemotaxis of the marine bacterium *Vibrio furnissii* to sugars: a potential mechanism for initiating the chitin catabolic cascade. *J. Biol. Chem.* **268**, 9405-9409
2. Yu, C., Lee, A.M., Bassler, B.L., and Roseman, S. (1991) Chitin utilization by marine bacteria: a physiological function for bacterial adhesion to immobilized carbohydrates. *J. Biol. Chem.* **266**, 24260-24267
3. Gildemeister, O., Zhu, B., and Laine, R. (1994) Chitovibrin: a chitin-binding lectin from *Vibrio parahaemolyticus*. *Glycoconjugate J.* **11**, 518-526
4. Laine, R.A., Ou, C.-Y., and Jaynes, J.M. (1994) Molecular clone of a chitinase gene from *Vibrio parahaemolyticus*, US Patent #5,352, 607
5. Bassler, B.L., Yu, C., Lee, Y.C., and Roseman, S. (1991) Chitin

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012

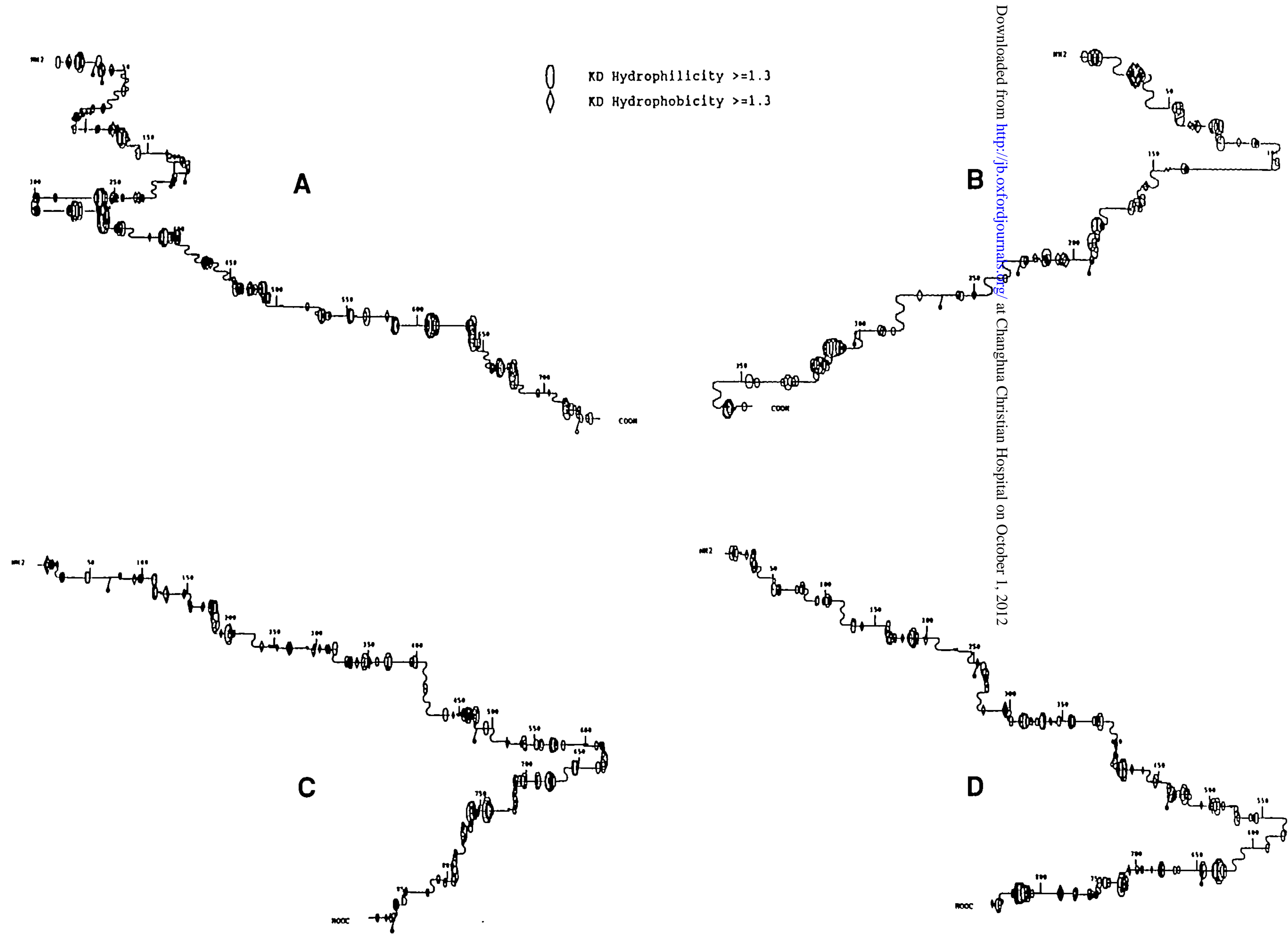

Fig. 4. **Chou-Fasman prediction of the secondary structures of chitobiases.** Relative positions and the amino and carboxyl termini are indicated. (A) *V. parahaemolyticus*, this study; (B) *H. sapiens* (*18*); (C) *V. harveyi* (*7*); (D) *V. vulnificus* (*17*).

utilization by marine bacteria: degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. *J. Biol. Chem.* **266**, 24276-24286
6. Wortman, A.T., Somerville, C.C., and Colwell, R.R. (1986) Chitinase determinants of *Vibrio vulnificus*: gene cloning and applications of a chitinase probe. *Appl. Environ. Microbiol.* **52**, 142-145
7. Soto-Gil, R.W., and Zyskind, J.W. (1989) *N,N′*-Diacetylchitobiase of *Vibrio harveyi*: primary structure, processing, and evolutionary relationships. *J. Biol. Chem.* **264**, 14778-14783
8. Zabin, I. and Fowler, A.V. (1980) Beta-galactosidase, the lactose permease protein, and thiogalactoside transacetylase in *The Operon* (Miller, J.H., and Reznikoff, W.S., eds.) 2nd ed, pp. 89-122, Cold Spring Harbor Laboratory
9. Zabin, I. (1980) Beta-galactosidase and the lactose operon in *The Evolution of Protein Structure and Function* (Sigman, D.S., and Brazier, M.A.B., eds.) Vol. 21, pp. 49-62, Academic Press, New York
10. Zhu, B.C., Lo, J.Y., Li, Y.T., Li, S.C., Jaynes, J.M., Gildemeister, O.S., Laine, R.A., and Ou, C.Y. (1992) Thermostable, salt tolerant, wide pH range novel chitobiase from *Vibrio parahaemolyticus*: isolation, characterization, molecular cloning, and expression. *J. Biochem.* **112**, 163-167
11. Asensio, C. and Ruiz-Amil, M. (1966) *N*-Acetyl-D-glucosamine kinase: II. *Escherichia coli. Methods Enzymol.* **9**, 421-425
12. Barkulis, S.S. (1966) *N*-Acetyl-D-glucosamine kinase: I. *Streptococcus pyogenes. Methods Enzymol.* **9**, 415-420
13. Staden, R. (1980) A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* **8**, 3673-3694
14. Chou, P.Y. and Fasman, G.D. (1988) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enz.* **47**, 45-147
15. Pearson, W. and Lipman, D. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448
16. Jannatipour, M., Soto-Gil, R.W., Childers, L.C., and Zyskind, J.W. (1987) Translocation of *Vibrio harveyi N,N′*-diacetylchitobiase to the outer membrane of *Escherichia coli. J. Bacteriol.* **169**, 3785-3791
17. Somerville, C.C., and Colwell, R.R. (1993) Sequence analysis of the beta-*N*-acetylhexosaminidase gene of *Vibrio vulnificus*: evidence for a common evolutionary origin of hexosaminidases. *Proc. Natl. Acad. Sci. USA*, **90**, 6751-6755
18. Fisher, K. and Aronson, N.J. (1992) Cloning and expression of the cDNA sequence encoding the lysosomal glycosidase di-*N*-acetylchitobiase. *J. Biol. Chem.* **267**, 19607-19616
19. Brown, C.A. and Mahuran, D.J. (1991) Active arginine residues in beta-hexosaminidase: identification through studies of the B1 variant of Tay-Sachs disease. *J. Biol. Chem.* **266**, 15855-15862
20. Feng, D.F. and Doolittle, R.F. (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**, 351-360
21. Needleman, S. and Wunsch, C. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453
22. Soto-Gil, R.W. and Zyskind, J.W. (1984) Cloning of *Vibrio harveyi* chitinase and chitobiase genes in *Escherichia coli in Chitin, Chitosan, and Related Enzymes* (Zikakis, J.P., ed.) pp. 209-223, Academic Press, New York
23. Meadow, N.D., Revuelta, R., Chen, V.N., Colwell, R.R., and Roseman, S. (1987) Phosphoenolpyruvate:glycose phosphotransferase system in species of *Vibrio*, a widely distributed marine bacterial genus. *J. Bacteriol.* **169**, 4893-4900
24. Meadow, N.D., Fox, D.K., and Roseman, S. (1990) The bacterial phosphoenol-pyruvate: glycose phosphotransferase system. *Annu. Rev. Biochem.* **59**, 497-542
25. Hawley, D. and McLure, W. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* **11**, 2237-2255
26. Rosenberg, M. and Court, D. (1975) Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* **13**, 319-353
27. Kuranda, M.J. and Aronson, N.N., Jr. (1986) A di-*N*-acetylchitobiase activity is involved in the lysosomal catabolism of asparagine-linked glycoprotein in rat liver. *J. Biol. Chem.* **261**, 5803-5809
28. Korneluk, R., Mahuran, D., Neote, K., Klavins, M., O'Dowd, B., Tropak, M., Willard, H., Anderson, M., Lowden, J., and Gravel, R. (1986) Isolation of cDNA clones coding for the alpha-subunit of human beta-hexosaminidase. *J. Biol. Chem.* **261**, 8407-8413
29. Beccari, T., Hoade, J., Orlacchio, A., and Stirling, J.L. (1992) Cloning and sequence analysis of a cDNA encoding the alpha-subunit of mouse beta-*N*-acetylhexosaminidase and comparison with the human enzyme. *Biochem. J.* **285**, 593-596
30. Yamanaka, S., Johnson, O.N., Norflus, F., Boles, D.J., and Proia, R.L. (1994) Structure and expression of the mouse beta-hexosaminidase genes, *Hexa* and *Hexb. Genomics* **21**, 588-596
31. Graham, T.R., Zassenhaus, H.P., and Kaplan, A. (1988) Molecular cloning of the cDNA which encodes beta-*N*-acetylhexosaminidase A from *Dictyostelium discoideum*: complete amino acid sequence and homology with the human enzyme. *J. Biol. Chem.* **263**, 16823-16829

Downloaded from http://jb.oxfordjournals.org/ at Changhua Christian Hospital on October 1, 2012