Microbiology (2008), **154**, 2492–2500 **DOI** 10.1099/mic.0.2008/016808-0

Improved annotation of conjugated bile acid hydrolase superfamily members in Gram-positive bacteria

Jolanda M. Lambert,[1,2] Roland J. Siezen,[1,2,3] Willem M. de Vos[1,4] and Michiel Kleerebezem[1,2,4]

Correspondence
Michiel Kleerebezem
Michiel.Kleerebezem@nizo.nl

[1]TI Food and Nutrition, PO Box 557, 6700 AN Wageningen, The Netherlands

[2]NIZO Food Research, PO Box 20, 6710 BA Ede, The Netherlands

[3]Radboud University Nijmegen Medical Centre/NCMLS, Centre for Molecular and Biomolecular Informatics, PO Box 9101, 6500 HB Nijmegen, The Netherlands

[4]Wageningen University, Laboratory of Microbiology, PO Box 8033, 6700 EJ Wageningen, The Netherlands

Most Gram-positive bacteria inhabiting the gastrointestinal tract are capable of hydrolysing bile salts. Bile salt hydrolysis is thought to play an important role in various biological processes in the host. Therefore, correct annotation of bacterial bile salt hydrolases (Bsh) in public databases (EC 3.5.1.24) is of importance, especially for lactobacilli, which are considered to play a major role in bile salt hydrolysis *in vivo.* In the present study, all enzymes listed in public databases that belong to the Bsh family and the closely related penicillin V acylase (Pva; EC 3.5.1.11) family were compared with the sequences annotated as Bsh in *Lactobacillus plantarum* WCFS1, as an example. In Gram-positive bacteria, a clear distinction was made between the two families using sequence alignment, phylogenetic clustering, and protein homology modelling. Biochemical and structural data on experimentally verified Bsh and Pva enzymes were used for validation of function prediction. Hidden Markov models were constructed from the sequence alignments to enable a more accurate prediction of Bsh-encoding genes, and their distinction from those encoding members of the Pva family. Many Pva-related sequences appeared to be annotated incorrectly as Bsh in public databases. This refinement in the annotation of Bsh family members influences the prediction of the function of *bsh*-like genes in species of the genus *Lactobacillus*, and it is discussed in detail.

Received 15 January 2008
Revised 14 April 2008
Accepted 24 April 2008

INTRODUCTION

Bile salts play an important role in lipid digestion in mammals. In the liver, bile salts are synthesized from cholesterol, and conjugated with the amino acids glycine or taurine. Following excretion into the intestinal lumen, the amino acid part of the bile salts can be hydrolysed (i.e. deconjugated) by bile salt hydrolases (Bsh; EC 3.5.1.24), also designated choloylglycine hydrolase or conjugated bile acid hydrolase (CBAH), produced by the intestinal microbiota.

Bacterial Bsh activity has received much attention based on its postulated role in both positive and negative biological processes in the host. For example, bile salt hydrolysis may be involved in serum cholesterol lowering (Pereira & Gibson, 2002). Bile salt deconjugation is the obligatory first reaction in further oxidation and dehydroxylation steps of bile salts by intestinal bacteria, and it includes the production of secondary bile salts that have been linked to various intestinal diseases, such as the formation of gallstones and colon cancer (Ridlon *et al.*, 2006). Moreover, results of *in vitro* studies have suggested that bile salt deconjugation plays a role in mucin production and excretion in the intestinal lumen (Klinkspoor *et al.*, 1999), and this could affect the nutritional environment encountered by the intestinal microbiota, and intestinal transit time (Shimotoyodome *et al.*, 2000).

Based on the biological implications of Bsh activity, it is important to provide a correct annotation of *bsh* genes in bacterial DNA sequences. Notably, Bsh amino acid sequences resemble those of penicillin V acylase (Pva; EC 3.5.1.11), and they belong to the same enzyme superfamily of linear amide CN hydrolases [Pfam CBAH, PF02275 (http://pfam.sanger.ac.uk)]. Although both Bsh and Pva are

Abbreviation: HMM, hidden Markov model.

Supplementary data showing the CLUSTAL_X alignment of sequences in the Bsh and Pva clusters, annotation of Gram-positive CBAH superfamily members in the phylogenetic tree, and HMM data are available with the online version of this paper.

capable of hydrolysing the same type of chemical bond, the overall chemical nature of their substrates is quite different (bile salts and penicillins, respectively; Fig. 1). The steroid moiety of bile salts is significantly more voluminous when compared with the corresponding moiety of penicillin V. Pva-encoding genes may at times be incorrectly annotated as Bsh, as has been found, for example, in *Listeria monocytogenes* (Begley *et al.*, 2005), and this can lead to the unreliable prediction of the presence of Bsh activity in a particular bacterial strain.

To confidently draw conclusions from *in silico* analysis of CBAH superfamily members, information on experimentally established enzyme activity and/or structure of Bsh and Pva enzymes is indispensable (Table 1). For example, experimentally verified Bsh enzymes can be found for *Enterococcus faecium* (Wijaya *et al.*, 2004), *Bifidobacterium* species (Kim *et al.*, 2004; Tanaka *et al.*, 2000), *Clostridium* species (Kirby *et al.*, 1995), *Lactobacillus* species (Christiaens *et al.*, 1992; Elkins & Savage, 1998; Lambert *et al.*, 2007; Leer *et al.*, 1993) and *Lis. monocytogenes* (Dussurget *et al.*, 2002). Experimentally verified Pva-family proteins have been found for *Bacillus* species (Olsson & Uhlen, 1986; Rathinaswamy *et al.*, 2005) and *Lis. monocytogenes* (Begley *et al.*, 2005).

Among bacterial species that are capable of bile salt hydrolysis, lactobacilli, in particular, have been reported to play a major role in bile salt deconjugation *in vivo* (Bongaerts *et al.*, 2000; De Smet *et al.*, 1995; Pereira *et al.*, 2003; Tannock *et al.*, 1994). In *Lactobacillus*-free mice, the Bsh activity in the ileal content has been found to be reduced by almost 90 % when compared with *Lactobacillus*-reconstituted mice (Tannock *et al.*, 1989). Moreover, one of the commonly used criteria in selection of candidate probiotic strains is their ability to hydrolyse bile salts (Begley *et al.*, 2006). There is a growing interest in the use of bile-salt-hydrolysing *Lactobacillus* strains in fermented milk products because they could potentially lead to reduction of serum cholesterol concentrations in humans (Pereira & Gibson, 2002). Consequently, the accuracy of annotation of *bsh* genes in lactobacilli deserves special attention.

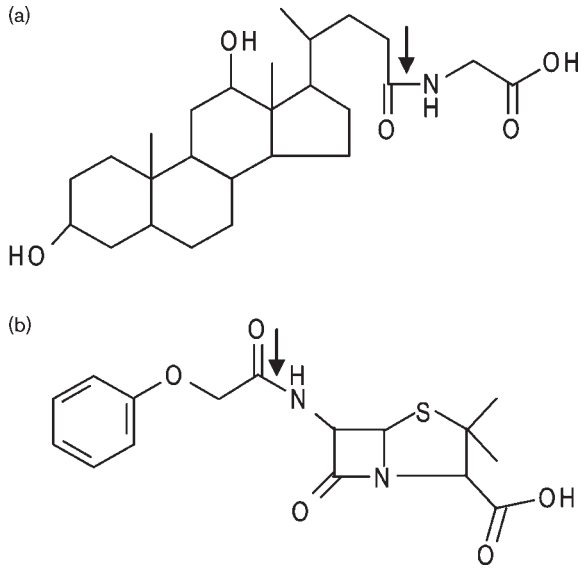

Fig. 1. Chemical structure of bile salts and penicillins. As examples, the bile salt glycodeoxycholic acid (a) and penicillin V (b) are shown. The bond that is hydrolysed by Bsh and Pva is indicated by an arrow.

In this study, a combination of *in silico* methods was employed to distinguish accurately between Bsh- and Pva-family members. All protein sequences of CBAH superfamily members in public databases were compared, and their functionality was predicted using phylogenetic profiling, sequence alignment and 3D-modelling techniques, incorporating experimentally verified Bsh and Pva proteins. In addition, hidden Markov models (HMMs) were constructed that enabled the distinction between Bsh and Pva proteins. Annotation of Bsh-family members in several *Lactobacillus* species is based mainly on sequence homology, and it has been found to be mostly incorrect, as exemplified for the four CBAH superfamily paralogues found in the model organism *Lactobacillus plantarum* WCFS1 (Kleerebezem *et al.*, 2003; Lambert *et al.*, 2007). In this work, we propose a refined annotation of Bsh- and Pva-family members in lactobacilli.

METHODS

Sequence alignment, phylogenetic profiling, and construction of HMMs. For phylogenetic profiling of Bsh- and Pva-family members, the sequences of putative CBAH superfamily members were retrieved from the ERGO database (Overbeek *et al.*, 2003) using BLASTP (Altschul *et al.*, 1990), and, as seed, the sequence of the experimentally verified Bsh1 (lp_3536) of *Lb. plantarum* WCFS1 (Kleerebezem *et al.*, 2003; Lambert *et al.*, 2007). Multiple sequence alignments were created with MUSCLE (Edgar, 2004), using default settings (www.drive5.com/muscle). CLUSTAL_X (Thompson *et al.*, 1997) was used to display multiple sequence alignments, and to create phylogenetic trees, and these trees were then displayed with LOFT (van der Heijden *et al.*, 2007). HMMs were constructed from the multiple sequence alignments by using HMMER 2.3.2 (Durbin *et al.*, 1998).

3D structure homology modelling. The 3D structures used as templates for homology modelling were Bsh from *Bifidobacterium longum* [2hez.pdb (Protein Data Bank identification no.; www.rcsb.org/pdb/home/home.do)], Bsh from *Clostridium perfringens* in complex with reaction products taurine and deoxycholate (2bjf.pdb), and Pva in complex with diethane diol from *Bacillus sphaericus* (2pva.pdb). Homology modelling was performed with WHATIF/YASARA Twinset (www.yasara.com).

RESULTS

Phylogenetic profiling

The sequences of the CBAH superfamily members [retrieved using the sequence of the experimentally verified Bsh1 of *Lb. plantarum* WCFS1 (Lambert *et al.*, 2007) as

Table 1. Experimentally verified Bsh- and Pva-family proteins

Sequence identification*	Bacterial strain	Reference	Remarks
Bsh family			
Q6R974	*Bif. bifidum* ATCC 11863	Kim *et al.* (2004)	
Q9KK62	*Bif. longum* SBT2928	Tanaka *et al.* (2000)	Crystallized, Kumar *et al.* (2006)
P54965	*C. perfringens* 13	Coleman & Hudson (1995)	Crystallized, Rossocha *et al.* (2005)
Q83YZ2	*E. faecium* FAIRE-E 345	Wijaya *et al.* (2004)	
Q9F660	*Lb. johnsonii* 100-100	Elkins & Savage (1998)	α Subunit
P97038	*Lb. johnsonii* 100-100	Elkins & Savage (1998)	β Subunit
Q8Y5J3	*Lis. monocytogenes* EGDe	Dussurget *et al.* (2002)	
S51638 (GenBank)	*Lb. plantarum* LP80	Christiaens *et al.* (1992)	
Q06115	*Lb. plantarum* WCFS1	Lambert *et al.* (2007)	Bsh1
Pva family			
P12256	*Bac. sphaericus* ATCC 14577	Olsson & Uhlen (1986)	Crystallized, Suresh *et al.* (1999)
Q2HPP6	*Bac. subtilis* NCIMB	Rathinaswamy *et al.* (2005)	Crystallized
Q8Y9S7	*Lis. monocytogenes* EGDe	Begley *et al.* (2005)	

*Sequence accession no. in the Swiss-Prot database, except where indicated.

seed] were analysed along with sequences of experimentally studied Bsh and Pva enzymes that are not present in the ERGO database (Overbeek *et al.*, 2003) (Table 1). The superfamily tree shows that sequences derived from Gram-negative and Gram-positive organisms were clearly divided in two separate groups.

Focusing on the group of sequences derived from Gram-positive organisms, two clusters were identified (Fig. 2). The first cluster, hereafter called the Bsh cluster, consisted of sequences that appeared phylogenetically closely related, and contained all Bsh proteins that had been biochemically verified, including Bsh1 of *Lb. plantarum* WCFS1. The Bsh cluster contained 25 sequences, of which 23 are annotated in the public NCBI database (http://ncbi.nlm.nih.gov) as Bsh, and one is annotated as penicillin acylase (see supplementary Table S1, available with the online version of this paper). For one sequence, no function is annotated; however, this protein has been shown experimentally to be a Bsh (Dussurget *et al.*, 2002). Furthermore, this cluster contained 10 sequences derived from lactobacilli, of which nine sequences are annotated as Bsh, and one sequence is annotated as Pva. The second cluster, hereafter called the Pva cluster, consisted of sequences that generally appeared to be less tightly related phylogenetically, as compared with those in the Bsh cluster. All experimentally verified Pva proteins were encompassed within the Pva cluster. This cluster contained 49 sequences, of which 11 are annotated as Pva, 30 are annotated as Bsh, one is annotated as an ABC transporter, and seven sequences lack a functional annotation (Table S1). Furthermore, eight sequences of lactobacilli were grouped in the Pva cluster, of which seven are annotated as Bsh, and one is annotated as Pva. The three paralogous proteins previously annotated as Bsh (Bsh2, Bsh3 and Bsh4; Kleerebezem *et al.*, 2003), which were predicted to be present in addition to the experimentally verified Bsh1 (Lambert *et al.*, 2007) in *Lb. plantarum* WCFS1, were grouped phylogenetically in the Pva cluster, with Bsh2 and Bsh3 being more closely related to each other than to Bsh4.

Importantly, all experimentally verified Bsh and Pva enzymes were grouped correctly during *in silico* analysis, strongly suggesting that Bsh and Pva proteins can be distinguished correctly using the phylogenetic profiling technique.

3D protein homology modelling, and sequence alignment

The 3D structures of the crystallized Bsh of *Bif. longum* (Kumar *et al.*, 2006) (2hez.pdb), Bsh of *C. perfringens* (Rossocha *et al.*, 2005) (2bjf.pdb), and Pva of *Bac. sphaericus* (Suresh *et al.*, 1999) (2pva.pdb) were superimposed, and the catalytic residues and binding pockets were compared (Fig. 3a, b). Based on multiple sequence alignment (see supplementary Fig. S1, available with the online version of this paper), the catalytic and binding-pocket residues of all Bsh and Pva family members were defined. The residues present in the substrate-binding pocket appeared to be more conserved in the Bsh cluster compared with the Pva cluster. Catalytic and putative substrate-binding residues of the Bsh and Pva clusters, and the four *Lb. plantarum* enzymes, are compared in Tables 2 and 3.

In Fig. 3(a), the substrate-binding pocket of *Bif. longum* Bsh is shown, with a deoxycholate molecule (as an example of the steroid moiety of bile salts) modelled in the same orientation as that in the *C. perfringens* 3D structure (2bjf.pdb). The side chains lining this pocket are shown to come mainly from three surface loops, and they are predominantly hydrophobic. The substrate-binding pocket is also hydrophobic in other sequences in the Bsh cluster (Table 2); this is in agreement with the hydrophobic nature of the steroid moiety of bile salts. As expected, the

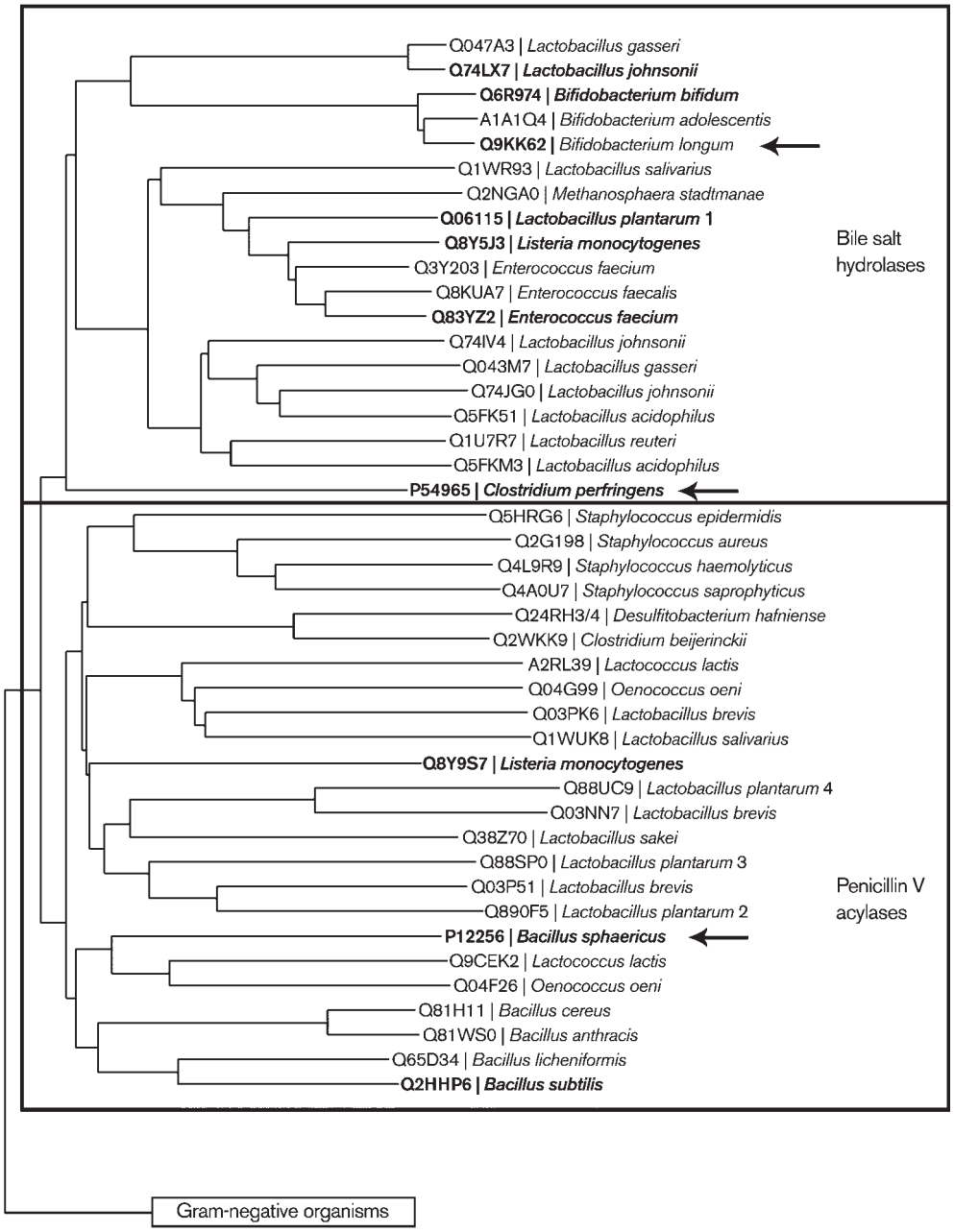

Fig. 2. Family tree of Bsh and Pva family members in Gram-positive bacteria, with Swiss-Prot accession numbers. For clarity, one representative strain is shown when sequences of different strains of the same organism are identical or nearly identical. The activity of sequences of strains shown in bold was experimentally verified, and strains that have sequences with known 3D structures are marked with arrows. The *bsh1*, *bsh2*, *bsh3* and *bsh4* genes of *Lb. plantarum* WCFS1 are indicated by the numbers 1, 2, 3 and 4, respectively.

predicted catalytic and substrate-binding residues of the experimentally verified Bsh1 enzyme of *Lb. plantarum* WCFS1 are most similar to those of the *Bif. longum* and *C. perfringens* Bsh enzymes, as shown in Table 2 and Fig. 3(c).

Clear differences were found in the 3D structure of the substrate-binding pockets of proteins of the Bsh and Pva clusters. First, loop 3 differs in length and orientation, as illustrated in Fig. 3(b). In the Pva of *Bac. sphaericus*, loop 3 is folded inward, severely reducing the size of the binding pocket. Second, the side chains of loops 2 and 3 lining the binding pocket are far more variable in size and hydrophilicity/hydrophobicity in the Pva group (Table 3). The catalytic residues, however, appear to be highly

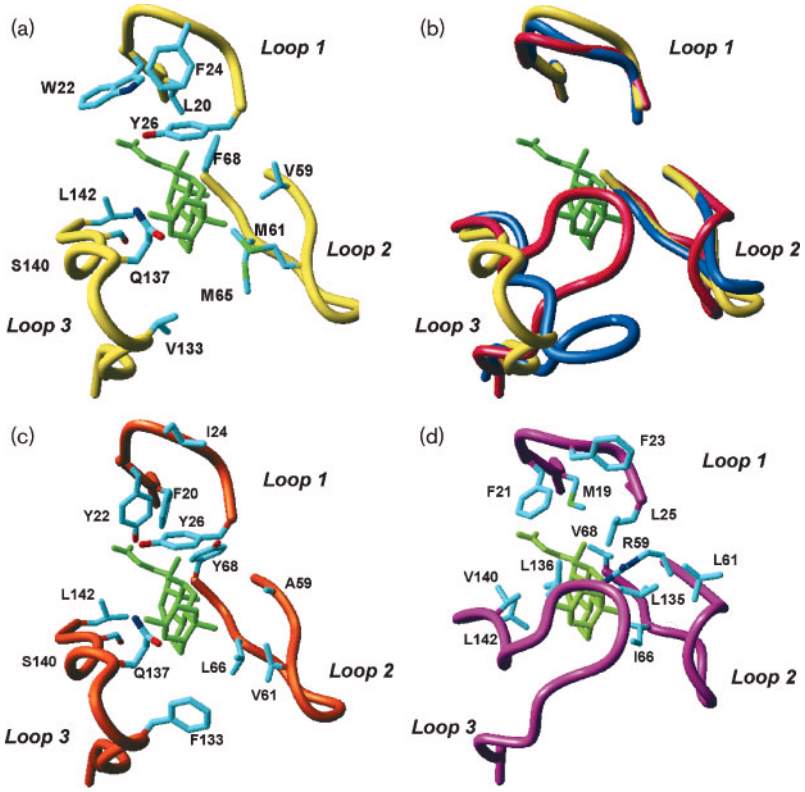

Fig. 3. 3D structure comparison of Bsh and Pva proteins. (a) Substrate-binding pocket of Bsh of *Bif. longum* (2hez.pdb), showing only side chains of residues in three loops (yellow) presumed to be relevant in binding substrates such as deoxycholate (green), which is modelled in the same orientation as that found in the 3D structure of the binding pocket of *C. perfringens* (2bjf.pdb). (b) Main chain backbone superposition of the three loops of the substrate-binding pockets of Bsh of *Bif. longum* (yellow) and *C. perfringens* (blue), and Pva of *Bac. sphaericus* (red). (c) Substrate-binding pocket of Bsh1 of *Lb. plantarum* WCFS1, as predicted by homology modelling, using the 3D structure of the binding pocket of Bsh of *Bif. longum* as a template. (d) Substrate-binding pocket of Bsh2 of *Lb. plantarum* WCFS1, as predicted by homology modelling, using the 3D structure of the binding pocket of Pva of *Bac. sphaericus* as a template.

conserved in the two respective groups. Five out of the six catalytic residues appear to be identical in both the Bsh and the Pva cluster, whereas the catalytic residue at position 82 in the 3D model was usually asparagine in the Bsh cluster, and tyrosine, or occasionally threonine (four out of 27), in the Pva cluster (Fig. S1). Predicted binding-pocket residues of Bsh2 and Bsh3 of *Lb. plantarum* WCFS1 appeared to resemble those of the 3D structure of Pva, as illustrated in Table 3 and Fig. 3(d), whereas the residues of Bsh4 appeared to fit less well to that of either Bsh of *Bif. longum* or Pva of *C. perfringens*.

Thus, using 3D modelling, the separation of Bsh family members in the Bsh and the Pva clusters, as found by phylogenetic profiling, was confirmed based on 3D models of the substrate-binding pocket and, consequently, the ability to accommodate bile salt molecules.

HMMs for distinction between Bsh and Pva family members

From the CLUSTAL_X alignments, separate and distinctive HMMs were constructed for the proteins of the Bsh and Pva clusters (see supplementary material SL1, available with the online version of this paper). Using the Bsh HMM to annotate the CBAH superfamily members, only proteins of the Bsh cluster were found as best hits. Likewise, only proteins of the Pva cluster were found as best hits using the Pva HMM. Subsequently, the two HMMs were used for a search of all publicly available sequenced genomes. Using the Bsh HMM, all best hits found in the sequenced genomes were already present in the Bsh cluster of the phylogenetic tree (Fig. 2), and no additional hits were found (E value $<1.80\times10^{-179}$). Furthermore, with the Bsh HMM, a clear separation was found in the E value of the members of the Bsh cluster (E value $<1.80\times10^{-179}$) and subsequent hits (E value $>2.00\times10^{-108}$), which were members of the Pva cluster (see supplementary material SL2, available with the online version of this paper). Likewise, using the Pva HMM, all best hits (E value $<6.60\times10^{-132}$) were found in the sequenced genomes consisting of members of the Pva cluster. Again, a clear separation was found in the E values of the Pva sequences (E value $<6.60\times10^{-132}$) and subsequent hits (E value $>8.20\times10^{-97}$), which were either members of the Bsh cluster or Gram-negative sequences not present in our phylogenetic tree. These findings strongly suggest that the HMMs presented here enable an accurate prediction of the functionality of any CBAH superfamily member, classifying it as either a Bsh enzyme (EC 3.5.1.24) or a Pva-related enzyme (EC 3.5.1.11).

DISCUSSION

In this study, we strived to improve the annotation of CBAH superfamily members by employing both phylogenetic clustering and 3D modelling techniques, in addition to the more conventional method of sequence alignment. The *in silico* analyses employed here consistently

Table 2. Catalytic and binding-pocket residues in Bsh family members: Bsh-cluster members

For active site residues, the first number designates the residue number in the template 3D structure used in homology modelling (*Bif. longum* Bsh) (Fig. 3), whereas the number in parentheses is the residue number used in CLUSTAL_X multiple sequence alignment (Fig. S1, supplementary material).

Residue	*Bif. longum*	*C. perfringens*	Bsh1, *Lb. plantarum*	Most common residue in family	Remarks
Catalytic					
2 (2)	C	C	C	C	Conserved
18 (18)	R	R	R	R	Conserved
21 (21)	D	D	D	D	Conserved
82 (82)	N	N	N	N	*Lb. gasseri*: S
177 (176)	N	N	N	N	Conserved
229 (229)	R	R	R	R	Conserved
Binding pocket: loop 1					
20 (20)	L	M	F	L, F	All very hydrophobic
22 (22)	W	I	Y	Y, L, W	All hydrophobic
24 (24)	F	Y	I	I, F	Mostly hydrophobic, but variable
26 (26)	Y	F	Y	Y, F	All hydrophobic
Binding pocket: loop 2					
58 (58)	G	G	T	G, A, S	Neutral to hydrophobic
59 (59)	V	T	A	A, V, I	All hydrophobic
61 (61)	M	F	V	M, V	Mostly hydrophobic, but variable
66 (66)	M	T	L	L	Mostly hydrophobic
68 (68)	F	A	Y	F, Y	All hydrophobic
103 (103)	F	Y	F	F, Y	All hydrophobic
Binding pocket: loop 3					
133 (133)	Deletion	I	F	F, I	All hydrophobic
137 (137)	Q	I	L	L	All hydrophobic
140 (140)	S	T	S	S, A	Neutral to hydrophobic
142 (142)	L	L	L	L	Conserved; very hydrophobic

distinguished between the Bsh and the homologous Pva enzymes. The Bsh enzymes are considered to be especially relevant for microbes that reside in the mammalian intestinal system, where lactobacilli are considered to be among the most important participants in bile salt deconjugation *in vivo* (Tannock *et al.*, 1989).

The construction of a phylogenetic tree is an excellent way of visualizing the relatedness of sequences, and it is more informative than the study of sequence homology alone. In our study, phylogenetic clustering clearly separated Gram-positive CBAH superfamily members into two groups: a Bsh cluster that is predicted to contain Bsh enzymes, and a Pva cluster, the members of which are predicted to be Pva-related proteins. Interestingly, it appeared that all strains that were represented in the Bsh cluster, and not in the Pva cluster, were typical gut-related bacteria (i.e. bifidobacteria, *Enterococcus faecalis, E. faecium* and *Lactobacillus acidophilus*), while strains represented in the Pva cluster, and not the Bsh cluster, were not typical gut-related bacteria. This finding is in good agreement with the reported correlation between the presence of Bsh activity in lactic acid bacteria and isolation from the intestine or faeces (Tanaka *et al.*, 1999). The prediction of the functionality of the CBAH superfamily members by phylogenetic profiling was reinforced by the fact that the functionality of all experimentally verified Bsh and Pva proteins matched the *in silico* clustering.

Our results indicate that the sequences annotated as Bsh in the Pva cluster are likely to be incorrectly annotated in the public NCBI (http://ncbi.nlm.nih.gov) and KEGG (www.genome.jp/kegg) databases, and that they are probably Pva or Pva-related enzymes. Particularly for *Lactobacillus* spp., the refinement of the annotation of functionality of members of this family of enzymes could have an important impact on their anticipated influence on gastrointestinal bile salt metabolism and its cognate physiological consequences in the host. According to our current analysis, the present Pva annotation of one sequence, the *Lactobacillus gasseri* ATCC 33323 CBAH superfamily sequence [Swiss-Prot accession no. Q047A3 (www.expasy.ch/sprot/)], should be corrected to Bsh, while the sequences in the Pva cluster currently annotated as Bsh in *Lactobacillus salivarius* UCC118 (Swiss-Prot accession no. Q1WUK8), *Lactobacillus brevis* ATCC 367 (Swiss-Prot accession nos Q03NN7 and Q03P51), *Lactobacillus sakei* 23K (Swiss-Prot accession no. Q38Z70), and *Lb. plantarum*

Table 3. Catalytic and binding pocket residues in Bsh-family members: Pva-cluster members

For active site residues, the first number designates the residue number in the template 3D structure used in homology modelling (*Bac. sphaericus* Pva) (Fig. 3), whereas the number in parentheses designates the residue number used in CLUSTAL_X multiple sequence alignment (Fig. S1, supplementary material).

Residue	*Bac. sphaericus*	Bsh2, *Lb. plantarum*	Bsh3, *Lb. plantarum*	Bsh4, *Lb. plantarum*	Most common residue in family	Remarks
Catalytic						
1 (4)	C	C	C	C	C	Conserved
17 (20)	R	R	R	R	R	Conserved
20 (23)	D	D	D	D	D	Conserved
82 (87)	Y	Y	Y	Y	Y	Occasionally T
175 (184)	N	N	N	N	N	Except *Desulfitobacterium hafniense* (Q)
228 (238)	R	R	R	R	R	Conserved
Binding pocket: loop 1						
19 (21)	M	M	M	M	M	Conserved; very hydrophobic
21 (24)	F	F	F	F	F, W	All hydrophobic
23 (26)	M	V	F	T	F, Y	Mostly hydrophobic, but variable
25 (29)	P	F	L	T	L, F, M	Mostly hydrophobic, but variable
Binding pocket: loop 2						
58 (53)	G	G	G	G	G	Conserved
59 (54)	S	R	R	R	R	Very variable
61 (56)	D	L	L	H	L, M	Very variable
66 (71)	V	I	I	L	I, L, D	Variable
68 (73)	Y	T	V	A	A, F, G	Neutral to hydrophobic
Binding pocket: loop 3						
135 (144)	I	L	L	D	L, D, T, I	Variable
136 (145)	L	I	L	T	L, I, T	Mostly hydrophobic, but variable
140 (149)	P	P	V	Y	P, A, T	Variable
142 (151)	L	L	L	F	L	All very hydrophobic

WCFS1 [Swiss-Prot accession nosQ890F5 (Bsh2); Q88SP0 (Bsh3), and Q88UC9 (Bsh4)], should be corrected to Pva related (Table S1). The reannotation of these sequences was further validated by experimental evidence of the presence of Bsh activity in bacterial strains. For example, for *Lb. brevis* ATCC 367, the original annotation predicted the presence of two Bsh sequences and one Pva sequence and, thus, this strain was expected to show Bsh activity. However, our reannotation predicted the presence of Pva-related sequences only; thus, *Lb. brevis* ATCC 367 should, according to our annotation, show no Bsh activity. Indeed, using a Bsh plate assay that has been described earlier (Dashkevicz & Feighner, 1989), *Lb. brevis* ATCC 367 showed no Bsh activity, suggesting that our reannotation of the Bsh sequences to Pva-related sequences was correct (data not shown).

Furthermore, the members of the Pva and Bsh clusters were not only phylogenetically different but also structurally different. 3D modelling and sequence alignment showed clear differences in the 3D structure of the substrate-binding pockets of Bsh and Pva enzymes, as illustrated in Fig. 3. In particular, one loop of the substrate-binding pocket (loop 3) in Pva is folded inwards as compared with the Bsh enzymes. As a consequence, it is unlikely that bile salts can be accommodated in the substrate-binding pocket of Pva (Fig. 1). Compared with members of the Bsh cluster, the members of the Pva cluster were found to be less related to each other during phylogenetic analysis. This finding was reflected in a higher variability in the residues of the binding pocket of members of the Pva cluster, as compared with members of the Bsh cluster, and this suggests that there is considerable variation in putative penicillin derivatives or similar molecules that can be bound. As expected, the predicted substrate-binding pocket of the experimentally verified Bsh1 of *Lb. plantarum* (Lambert *et al.*, 2007) structurally resembled the Bsh enzymes of *Bif. longum* (Kumar *et al.*, 2006) and *C. perfringens* (Rossocha *et al.*, 2005). The limited degree of variation in the residues of the binding pocket, and thus a more strict substrate preference, for members of the Bsh cluster is reflected by the fact that Bsh1 of *Lb. plantarum* WCFS1 is able to hydrolyse bile salts, but not penicillin V or penicillin G (Lambert *et al.*, 2008).

In addition, the paralogues Bsh2, Bsh3 and Bsh4 were clustered in the Pva group during phylogenetic analysis, with Bsh2 and Bsh3 being more related to each other than to Bsh4. This is in agreement with the finding that the

predicted substrate-binding pocket of Bsh2 and Bsh3 appeared to show a greater structural resemblance to that of the Pva enzyme of *Bac. sphaericus* (Suresh *et al.*, 1999) than that of Bsh4. It is possible that Bsh4 encodes an enzyme related to Pva that is capable of acylating a substrate other than penicillin V.

Notably, one of the catalytic residues appeared to differ between the Bsh and Pva cluster. This finding facilitates accurate prediction of the functionality of the Bsh family members as Bsh or Pva (-related) from their amino acid sequences. Novel and discriminative HMMs were constructed to distinguish between Bsh- and Pva-family members. Thereby, our analyses defined the discriminative characteristics of the Pva and Bsh enzyme families, despite the fact that this did not appear to be feasible using the available conserved motif databases. For example, the conserved domain search tool (Marchler-Bauer *et al.*, 2005) provided by NCBI (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) aims at detection and prediction of conserved domains in proteins, and includes domains imported from SMART, Pfam and COG databases. However, this tool attributes 'penicillin V acylase, also known as conjugated bile salt acid hydrolase' to the best-conserved domain found in all experimentally verified Bsh and Pva proteins used here (Table 1), and thereby fails to distinguish between Bsh-like and penicillin-acylase-like domains. Analogously, SMART, Pfam and COG databases do not provide a correct discriminative domain description for all experimentally verified Bsh and Pva proteins (data not shown).

In conclusion, we have provided evidence that various Pva-related sequences are wrongly annotated as Bsh in various Gram-positive bacteria, including several lactobacilli. For intestinal lactobacilli in particular, the capacity to deconjugate bile acids has been suggested to be of importance for their survival capacity in the gastrointestinal tract of mammals, and it is proposed to strongly influence gut physiology. Therefore, appropriate annotation of the Bsh enzymes encoded in the genomes of *Lactobacillus* and other lactic acid bacteria is important to predict their *in situ* functionality in the gastrointestinal tract. The methodology presented here provides the *in silico* tools to effectively distinguish the genes encoding the important Bsh enzymes from those encoding closely related enzymes with Pva-like activities.

ACKNOWLEDGEMENTS

The authors would like to thank Bernadet Renckens and Michiel Wels for their technical support, and valuable contribution to this manuscript.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Begley, M., Sleator, R. D., Gahan, C. G. & Hill, C. (2005). Contribution of three bile-associated loci, *bsh*, *pva*, and *btlB*, to gastrointestinal persistence and bile tolerance of *Listeria monocytogenes*. *Infect Immun* **73**, 894–904.

Begley, M., Hill, C. & Gahan, C. G. (2006). Bile salt hydrolase activity in probiotics. *Appl Environ Microbiol* **72**, 1729–1738.

Bongaerts, G. P., Severijnen, R. S., Tangerman, A., Verrips, A. & Tolboom, J. J. (2000). Bile acid deconjugation by lactobacilli and its effects in patients with a short small bowel. *J Gastroenterol* **35**, 801–804.

Christiaens, H., Leer, R. J., Pouwels, P. H. & Verstraete, W. (1992). Cloning and expression of a conjugated bile acid hydrolase gene from *Lactobacillus plantarum* by using a direct plate assay. *Appl Environ Microbiol* **58**, 3792–3798.

Coleman, J. P. & Hudson, L. L. (1995). Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens*. *Appl Environ Microbiol* **61**, 2514–2520.

Dashkevicz, M. P. & Feighner, S. D. (1989). Development of a differential medium for bile salt hydrolase-active *Lactobacillus* spp. *Appl Environ Microbiol* **55**, 11–16.

De Smet, I., Van Hoorde, L., Vande Woestyne, M., Christiaens, H. & Verstraete, W. (1995). Significance of bile salt hydrolytic activities of lactobacilli. *J Appl Bacteriol* **79**, 292–301.

Durbin, R., Eddy, S. R., Krogh, A. & Mitchison, G. (1998). *Biological Sequence Analysis: Probablistic Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press.

Dussurget, O., Cabanes, D., Dehoux, P., Lecuit, M., Buchrieser, C., Glaser, P. & Cossart, P. (2002). *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol Microbiol* **45**, 1095–1106.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.

Elkins, C. A. & Savage, D. C. (1998). Identification of genes encoding conjugated bile salt hydrolase and transport in *Lactobacillus johnsonii* 100–100. *J Bacteriol* **180**, 4344–4349.

Kim, G. B., Miyamoto, C. M., Meighen, E. A. & Lee, B. H. (2004). Cloning and characterization of the bile salt hydrolase genes (*bsh*) from *Bifidobacterium bifidum* strains. *Appl Environ Microbiol* **70**, 5603–5612.

Kirby, L. C., Klein, R. A. & Coleman, J. P. (1995). Continuous spectrophotometric assay of conjugated bile acid hydrolase. *Lipids* **30**, 863–867.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M. & other authors (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc Natl Acad Sci U S A* **100**, 1990–1995.

Klinkspoor, J. H., Mok, K. S., Van Klinken, B. J., Tytgat, G. N., Lee, S. P. & Groen, A. K. (1999). Mucin secretion by the human colon cell line LS174T is regulated by bile salts. *Glycobiology* **9**, 13–19.

Kumar, R. S., Brannigan, J. A., Prabhune, A. A., Pundle, A. V., Dodson, G. G., Dodson, E. J. & Suresh, C. G. (2006). Structural and functional analysis of a conjugated bile salt hydrolase from *Bifidobacterium longum* reveals an evolutionary relationship with penicillin V acylase. *J Biol Chem* **281**, 32516–32525.

Lambert, J. M., Bongers, R. S. & Kleerebezem, M. (2007). Cre-*lox*-based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. *Appl Environ Microbiol* **73**, 1126–1135.

Lambert, J. M., Bongers, R. S., De Vos, W. M. & Kleerebezem, M. (2008). Functional analysis of four bile salt hydrolase and penicillin

acylase family members in *Lactobacillus plantarum* WCFS1. *Appl Environ Microbiol* (in press).

Leer, R. J., Christiaens, H., Verstraete, W., Peters, L., Posno, M. & Pouwels, P. H. (1993). Gene disruption in *Lactobacillus plantarum* strain 80 by site-specific recombination: isolation of a mutant strain deficient in conjugated bile salt hydrolase activity. *Mol Gen Genet* **239**, 269–272.

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D. & other authors (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* **33**, D192–D196.

Olsson, A. & Uhlen, M. (1986). Sequencing and heterologous expression of the gene encoding penicillin V amidase from *Bacillus sphaericus. Gene* **45**, 175–181.

Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E., Jr, Liolios, K., Joukov, V., Kaznadzev, D. & other authors (2003). The ERGO genome analysis and discovery system. *Nucleic Acids Res* **31**, 164–171.

Pereira, D. I. & Gibson, G. R. (2002). Effects of consumption of probiotics and prebiotics on serum lipid levels in humans. *Crit Rev Biochem Mol Biol* **37**, 259–281.

Pereira, D. I., McCartney, A. L. & Gibson, G. R. (2003). An *in vitro* study of the probiotic potential of a bile-salt-hydrolyzing *Lactobacillus fermentum* strain, and determination of its cholesterol-lowering properties. *Appl Environ Microbiol* **69**, 4743–4752.

Rathinaswamy, P., Pundle, A. V., Prabhune, A. A., Sivaraman, H., Brannigan, J. A., Dodson, G. G. & Suresh, C. G. (2005). Cloning, purification, crystallization and preliminary structural studies of penicillin V acylase from *Bacillus subtilis. Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 680–683.

Ridlon, J. M., Kang, D. J. & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* **47**, 241–259.

Rossocha, M., Schultz-Heienbrok, R., von Moeller, H., Coleman, J. P. & Saenger, W. (2005). Conjugated bile acid hydrolase is a tetrameric N-terminal thiol hydrolase with specific recognition of its cholyl but not of its tauryl product. *Biochemistry* **44**, 5739–5748.

Shimotoyodome, A., Meguro, S., Hase, T., Tokimitsu, I. & Sakata, T. (2000). Decreased colonic mucus in rats with loperamide-induced constipation. *Comp Biochem Physiol A Mol Integr Physiol* **126**, 203–212.

Suresh, C. G., Pundle, A. V., SivaRaman, H., Rao, K. N., Brannigan, J. A., McVey, C. E., Verma, C. S., Dauter, Z., Dodson, E. J. & Dodson, G. G. (1999). Penicillin V acylase crystal structure reveals new Ntn-hydrolase family members. *Nat Struct Biol* **6**, 414–416.

Tanaka, H., Doesburg, K., Iwasaki, T. & Mierau, I. (1999). Screening of lactic acid bacteria for bile salt hydrolase activity. *J Dairy Sci* **82**, 2530–2535.

Tanaka, H., Hashiba, H., Kok, J. & Mierau, I. (2000). Bile salt hydrolase of *Bifidobacterium longum* – biochemical and genetic characterization. *Appl Environ Microbiol* **66**, 2502–2512.

Tannock, G. W., Dashkevicz, M. P. & Feighner, S. D. (1989). Lactobacilli and bile salt hydrolase in the murine intestinal tract. *Appl Environ Microbiol* **55**, 1848–1851.

Tannock, G. W., Tangerman, A., Van Schaik, A. & McConnell, M. A. (1994). Deconjugation of bile acids by lactobacilli in the mouse small bowel. *Appl Environ Microbiol* **60**, 3419–3420.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

van der Heijden, R. T., Snel, B., van Noort, V. & Huynen, M. A. (2007). Orthology prediction at scalable resolution by phylogenetic tree analysis. *BMC Bioinformatics* **8**, 83.

Wijaya, A., Hermann, A., Abriouel, H., Specht, I., Yousif, N. M., Holzapfel, W. H. & Franz, C. M. (2004). Cloning of the bile salt hydrolase (*bsh*) gene from *Enterococcus faecium* FAIR-E 345 and chromosomal location of bsh genes in food enterococci. *J Food Prot* **67**, 2772–2778.

Edited by: J. M. van Dijl