Transactions of the Illinois State Academy of Science
(2017) Volume 110, pp. 9-15

received 11/7/16
accepted 8/28/17

Comparative Analysis of Alternansucrase Genes from *Leuconostoc* Strains

Scott Michael Holt
Western Illinois University, Department of Biological Sciences
Macomb, IL 61455, sm-holt@wiu.edu

ABSTRACT

Alternansucrase from the bacterial genus *Leuconostoc* catalyzes the synthesis of alternan, which has many commercial applications. A comparative analysis was performed to build a consensus of molecular features for gene sequences previously annotated as "alternansucrase." Alternansucrase genes from *L. citreum* isolates formed a distinct clade among glucansucrases. The clade had identical predicted gene and protein size. Amino acid sequence features for the signal peptide, key regions within the catalytic domain, and C-terminal repeat structures were also nearly identical for the clade. Alternansucrase gene from *L. fallax*, however, is more distantly related and possesses features that are clearly distinct from the clade group. Consequently, the gene from *L. fallax* appears atypical and should be designated so when annotated as an "alternansucrase." With tremendous influx of new sequence information due to next generation sequencing technology accurate annotation of gene function becomes extremely important particularly for genome mining applications. This study provided a consensus of molecular features for "alternansucrase" that should be considered during annotation of sequence information.

Key words: Alternansucrase, Dextransucrase, Leuconostoc, Alternan, α-glucan

INTRODUCTION

Glucansucrases made by certain *Leuconostoc* bacterial species catalyze the synthesis of α-glucans from inexpensive, widely available, and renewable agricultural residues such as sucrose. *Leuconostoc* and their corresponding α-glucans have a variety of realized (Alsop 1983, Cargill Inc. 2009, Dolan et al. 2012, Gronwell and Ingelman 1948, Monsan and Paul 1995, Paul et al. 1992) and potential commercial applications (Côté and Holt 2007, Day and Chung 2007, Irague et al. 2012, Lappan and Fogler 1995, Vandamme et al. 2003) which include many food, nonfood, and industrial uses. Understanding the diversity and structure/function relationships of glucansucrases will enhance their biotechnology potential by aiding the development of unique or tailor-made α-glucans for specific applications.

Leuconostoc glucansucrases belong to Glycoside Hydrolase Family 70 (CAZY 2015, Coutinho and Henrissat 1999) and include alternansucrase (EC 2.4.1.140) and dextransucrase (EC 2.4.1.5) among others. Alternansucrase catalyzes the synthesis of the α-glucan called alternan [alternating α-(1→6) and α-(1→3) links] which has novel chemical properties (Côté and Robyt 1982). Alternan polymers from different *Leuconostoc* isolates are structurally very similar with the exception that the alternating α-(1→6) and α-(1→3) links are in different proportions (Bounaix et al. 2009, Jeans et al. 1954, Jeans 1979, Seymore et al. 1979). Alternansucrase may serve as a good model for structure/function studies since *asr* genes are probably highly conserved like all GH70 glucansucrases. In addition, the biochemical characteristics of the few alternan polymers reported in the literature differ only in their linkage proportion (Bounaix et al. 2009, Jeans et al. 1954, Jeans 1979, Seymore et al. 1979). This difference of a single characteristic (link proportion) among *asr* gene products may help simplify future structure/function relationship studies of alternansucrases. In order to relate enzyme structural components (1° and 2°) with functional characteristics (specificities), more *asr* genes must be properly identified and characterized. Until recently, only one alternansucrase gene sequence (*asr*1355) has been reported from *Leuconostoc* strain B-1355 (Arguello-Morales et al. 2000). Since *asr*1355 has been reported (Arguello-Morales et al. 2000), other sequences described as "alternansucrase" from different *Leuconostoc* isolates have become available through Sanger or genomic sequencing (Amari et al. 2014, Bounaix et al. 2009, Kim et al. 2008, Nam et al. 2011). Thus far, alternan formation, alternansucrase enzyme activity, or an alternansucrase gene sequence has only be reported from *Leuconostoc* (Alsop 1983, Côté and Robyt 1982, Jeans et al. 1954, Kim et al. 2008, Nam et al. 2011).

The goal of this study was to perform a comparative analysis of *asr*1355 sequence features with more recently available *asr* sequences (putative) reported in online databases. Phylogenetic relationships and regions of consensus and divergence were identified among these commercially important genes from *Leuconostoc* isolates obtained from across the globe. Information generated from this study will provide more accurate annotation or prediction of what constitutes an *asr* gene. In addition, this study identified new potential targets for site-directed mutagenesis needed to elucidate structure/function relationships of alternansucrase and possibly other glucansucrases.

METHODS

Bacterial Strains and Culture Conditions

Leuconostoc strains B-1498 and B-1501 were obtained from the United States Department of Agriculture, NRRL culture collection, Peoria, IL. *Leuconostoc* strains B-1498 and B-1501 stock cultures were stored at -80°C. A portion of each stock culture was sub-cultured by isolation streak on MRS agar (Oxoid LTD, Hampshire, England) plates (De Man et al. 1960) and incubated at 30°C for 48 h. An isolated colony from each strain was used to inoculate 25-ml MRS broth which was incubated at 30°C overnight with shaking at 200 rpm.

Source of Alternansucrase (*asr*) Genes

Alternansucrase (*asr*) genes are rare. At the time of this study, only six *asr* genes (in-

Table 1. Putative alternansucrase genes used in this study, source information, and characteristics.

Source Organism	Strain Origin	Accession Number Genbank	Gene Size (bp)	AA	Mol. Wt (kDa)	glucan linkage % α(1→3) ; α(1→6)
L. citreum NRRL B-1355	FL, USA[1]	AJ250173[2]	6174	2057	228,971	43 ; 57 [1,3,4]
L. citreum NRRL B-1498	LA, USA[5]	KF360257 Holt, this study	6174	2057	228,798	38 ; 62 [1,3,4]
L. citreum NRRL B-1501	LA, USA[5]	KF360258 Holt, this study	6174	2057	229,113	33 ; 67 [1,3,4]
L. citreum KM20	Korea[6]	NC_010471[7]	6174	2057	228,864	Not determined
L. citreum LBAE C11	France[8]	NZ_CAGF01000008[9]	6174	2057	229,094	40 ; 60 [9]
L. fallax[T] KCTC 3537	Germany[10]	NZ_AEIZ01000002[11]	6525	1874	208,192	Not determined

[1]Jeans et al. 1954, [2]Arguello-Morales et al. 2000, [3]Jeans 1979, [4]Seymore et al. 1979, [5]McCleskey et al. 1947, [6]Choi et al. 2003, [7]Kim et al. 2008, [8]Gabriel et al. 1999, [9]Bounaix et al. 2009, [10]Martinez-Murcia and Collins 1991, [11]Nam et al. 2011.

cluding *asr*1355) were reported in the literature or online databases and are listed in Table 1. All *asr* genes were considered putative except *asr*1355 which has been cloned, sequenced, and its protein product biochemically characterized (Arguello-Morales et al. 2000). *asr* gene sequences from strains *Leuconostoc* B-1498 and 1501 were sequenced for this study. asr gene sequences from Leuconostoc strains KM20, C11, and KCTC 3537 were determined through genome sequencing from previous studies (Amari et al. 2014, Kim et al. 2008, Nam et al. 2011). The linkage distribution from alternan products for strains B-1355, B-1498, and B-1501 (Jeans et al. 1954, Jeans 1979, Seymour et al. 1979), and C11 (Bounaix et al. 2009) were reported from previous studies. The alternan products reported for the previous studies were prepared from partially purified enzymatic culture extracts (Bounaix et al. 2009, Jeans et al. 1954, Jeans 1979, Seymour et al. 1979). Alternan production from *Leuconostoc* strains KM20 and 3537 has not been reported (Kim et al. 2008, Nam et al. 2011).

DNA Extraction, PCR, and Cloning

MRS broth cultures were used to isolate genomic DNA from *Leuconostoc* B-1498 and B-1501 according to manufacturer's instructions (MOBIO, LaJolla, CA). Putative *asr* genes from *Leuconostoc* B-1498 (*asr*1498) and B-1501 (*asr*1501) chromosomal DNA were amplified by PCR. PCR primers used to amplify *asr*1498 and *asr*1501 were prepared from non-coding regions of previously characterized *asr*1355 cloned from *Leuconostoc* strain B-1355 (Arguello-Morales et al. 2000) (Forward 5'-TAT ATG GGT GAT AGA TGC AC-3' and Reverse 5'-ATA ATG TTA CCC TCC TTT GT-3'). The predicted PCR product size based on the *asr*1355 gene is 6270 bp, which included non-coding regions. PCR reaction included ddH_2O, AccuTaq LA 10x buffer, DNA (100 ng), $MgSO_4$ (1.0 mM), primers (12.0 pm), dNTP mix (200 μM), and JumpStart RedAccutaq LA DNA Polymerase (Sigma Aldrich, St Louis, MO, 0.5 unit). PCR conditions were performed as previously reported (Kossman et al. 2000). Initial denaturation at 95°C for 2 min preceded cycle conditions. Cycle conditions (35 cycles) included denaturation at 95°C for 45 s, primer annealing at 50°C for 45 s, primer extension at 68°C using ramped time starting at 7 m ending at 8 min 42 s. Final extension was at 68°C for 15 min. The amplified *asr*1498 and *asr*1501 PCR products were purified and cloned into a TOPO vector (Invitrogen Inc, Carlsbad, CA) according to manufacturer's instructions.

Sequencing, Bioinformatics, and Phylogenic Analysis

Putative *asr*1498 and *asr*1501 cloned genes were sequenced from both ends by primer walking using Sanger chemistry. Purified PCR products were also used as templates to sequence portions of the *asr* genes not satisfactorily resolved using the cloned genes as template. Sequence reads (12-13 reads for each strand) for *asr*1498 and *asr*1501 were assembled using CodonCode and DNAbaser. Comparative analyses of *asr*1355 with putative *asr*1498 and *asr*1501 genes (this study) and three other sequences described as "alternansucrase" (*asr*KM20, *L. citreum* KM20; *asr*C11, *L. citreum* LBAE C11; *asr*3537, *L. fallax* KCTC 3537) available from online databases were analyzed by BLAST, ORF Finder, PrediSi, ClustalW2, CD-Domain, and C-DART. At the time of this study, only six alternansucrase gene sequences have been reported in the literature or in online databases. For phylogenetic analyses shown in figures 2, 3, S1 and S2, full-length amino acid sequences of putative alternansucrase genes and other select glucoside hydrolase family 70 enzymes (available from online databases) were first aligned by MUSCLE and files were converted to MEGA (version6) file format. MEGA6 software was then used to prepare unrooted phylogenetic trees from the aligned full-length amino acid sequences by Neighbor-Joining and Maximum Likelihood methods. The Neighbor-Joining algorithm parameters "complete deletion," "amino acid p-distance," and "bootstrap" (1000 replications) were used in the analyses. The Maximum Likelihood algorithm parameters of "complete deletion," "amino acid substitution model" (Jones-Taylor-Thornton (JTT) model), "Nearest-Neighbor-Interchange (NNI) tree inference option", and "bootstrap" (1000 replications) were used in the analyses. No bootstrap "cutoff" value was used during the preparation of phylogenetic trees so that all data can be shown.

RESULTS AND DISCUSSION

Five gene sequences described as "alternansucrase" from different *Leuconostoc* isolates were compared with *asr* from *L. citreum* B-1355 (*asr*1355) for phylogenetic relationships and sequence features. This study is the first comparative analysis of multiple asr genes from *Leuconostoc* strains isolated from three continents. All *asr* genes mentioned in this study were considered putative except *asr*1355 (Arguello-Morales et al. 2000).

Sequence Analysis and Phylogenetic Relationships of Alternansucrase Genes

The predicted gene lengths of *asr*1498, *asr*1501, *asr*KM20, and *asr*C11 were the same as for *asr*1355 (6174 nt) all encod-

ing a deduced protein product (ASR) of 2057 amino acids (Table 1). The putative alternansucrase gene from *Leuconostoc fallax* NCTC 3537 (*asr*3537), however, has a reported gene length of 5625 nt and a predicted protein product of 1874 amino acids (Table 1). The estimated molecular mass for each ASR is approximately 229 kDa except from *asr*3537 which is 209 kDa (Table 1). Nucleotide and protein sequence alignments of the *asr* genes examined in this study with *asr*1355 indicated a very high degree of residue conservation with percent identities ranging above 97% except for *asr*3537 which had 66% nt identity and 52% amino acid identity. The putative promoter and RBS sites located upstream of the predicted translational start codon regions were also highly conserved among all *asr* genes examined except for *asr*3537 (Fig 1., Arguello-Morales et al. 2000).

Putative ASR proteins sub-clustered into three groups (*asr*1498 and *asr*1501, *asr*KM20 and *asr*C11, and *asr*1355) when compared among themselves but *asr*3537 formed its own divergent branch (Fig 2). Topology of the neighbor-joining tree generated in Figure 2 was congruent with a maximum likelihood tree (Supplemental Fig. S1), however, bootstrap values were slightly different as can be expected because each tree methodology uses a different algorithm. ASR clustering profile may be related to the proportion of α(1→3) links within each alternan polymer reported by some of the *Leuconostoc* strains. ASR from *Leuconostoc* strains B-1498 and B-1501 make less than 40%, C11 makes 40%, and B-1355 makes over 40% α(1→3) links (Table 1). More information is needed, however, to confirm this observation.

ASRs examined in this study formed a distinct clade when compared among other select GH70 glucansucrases except for *asr*3537 which formed its own branch (Fig 3). Topology of the neighbor-joining tree generated in Figure 3 was congruent with a maximum likelihood tree (Supplemental Fig. S2), however, bootstrap values were slightly different as can be expected because each tree methodology uses a different algorithm. The GH70 glucansucrases selected for the tree analysis are commonly used in other studies (Arigmon et al. 2013, Hoshino et al. 2012). All ASR from the *asr*1355 group (*asr*KM20, *asr*C11, *asr*1498, *asr*1501, *asr*1355) originated from *Leuconostoc citreum* species which suggests a monophyletic relationship similar to streptococcal glucansucrases (Fig 3., Arigmon et al. 2013).

```
Strain        -35                  -10                  RBS        Start
1355  TATATGGGGTGATAGATGCACCAAATACTGTATCATGTCTGGTCACATGAAAGGGAGAATAATTAATGAAA
KM20  TTATTGTGGTGATAGATGCACCAAATACTGTATCATGTTTGGTCACATGAAAGGGAGAGTAATTAATGAAA
C11   TTATTGTGGTGATAGATGCACCAAATACTGTATCATGTTTGGTCACATGAAAGGGAGAGTAATTAATGAAA
1498  TATATGGGGTGATAGATGCACCAAATACTGTATCATGTTTGGTCACATGAAAGGGAGAGTAATTAATGAAA
1501  TATATGGGGTGATAGATGCACCAAATACTGTATCATGTTTGGTCACATGAAAGGGAGAGTAATTAATGAAA
3537  TACGGGTTTTACATGGCCCAAAAGTTTGAAAATTTATTTTTTCATATCCATAATGGGAGAGATTAATGAGA
```

Figure 1. Putative promoter (-10, -35), ribosomal binding site (RBS), and start regions for putative *asr* genes in *Leuconostoc* strains are shown in bold (Arguello-Morales et al. 2000).

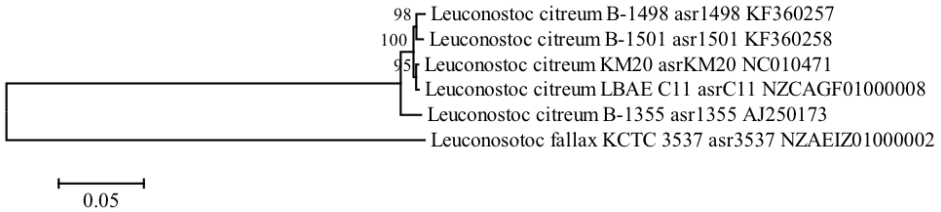

Figure 2. Phylogenetic analysis of full-length amino acid putative alternansucrase sequences. The tree was generated by the Neighbor-Joining Method and is drawn to scale. Bootstrap values (%) are shown above branch points (1000 replications). The scale bar represents substitutions per site. *Leuconostoc* species, strain, *asr* gene, and accession number are shown at nodes.

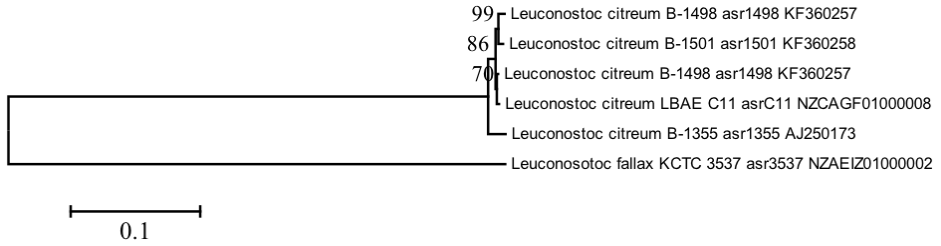

Figure S1. Phylogenetic analysis of full-length amino acid putative alternansucrase sequences. The tree was generated by the Maximum Likelihood Method and is drawn to scale. Bootstrap values (%) are shown above branch points (1000 replications). All bootstrap values are shown. The scale bar represents substitutions per site. *Leuconostoc species*, strain, *asr* gene, and accession number are shown at nodes.

Sequence Features of Predicted Alternansucrase Gene Products

All *asr* included in this study possess amino acid domain features typical of GH 70 glucansucrases which includes a signal peptide, an N-terminal variable region containing repeats elements, a highly conserved catalytic core (AmyAc Family), and a C-terminal region harboring repeat elements (Van Hijum et al. 2006). It is beyond the scope of this article to discuss in detail the basic domain structures typical of GH 70 glucansucrases as excellent reviews have been prepared (Korakli and Vogel 2006, Monchois et al. 1999, Remaud-Simeon et al. 2000, Van Hijum et al. 2006). This study will focus on sequence features specific to *asr*.

Signal Peptide of Predicted Alternansucrase Gene Products

The presence of a signal peptide region usually indicates that the protein may be translocated across the bacterial cell membrane. The N-terminal signal peptide region includes aa 1 through 39 for the *asr*1355 group and consists of a canonical conserved

segment typical of Gram-Positive bacteria within the Firmicutes phylum (von Heijne and Abrahmsen 1989, von Heijne 1990). Signal peptide cleavage is predicted to occur after residue 39 for the *asr*1355 group (PredicSeq). The signal peptide region is highly conserved with only one non-identical conserved residue for *asr*1355 occurring at residue 38. The *asr*3537 signal peptide is longer by one amino acid and cleavage is predicted to occur after residue 40 (PredicSeq).

Catalytic Domain of Predicted Alternansucrase Gene Products

GH70 family enzymes possess an (β/α)8 barrel protein configuration with alternating β-sheets and α-helices (Van Hijum et al. 2006). The enzyme family also has conserved catalytic triad residues (D, E, D) thought to be directly involved in catalysis and a retaining reaction mechanism (retention of stereochemistry, Van Hijum et al. 2006). Amino acid sequence flanking the catalytic triad residues (D615, E673, D767; ASR1355 numbering) within the catalytic core region are typically highly conserved among glucansucrases of GH70 family (Fabre et al. 2006). Consequently, divergent residues within the areas may be related to functional characteristics and help identify molecular features specific to the protein product. The four highly conserved β-sheet regions (common to all GH 70 enzymes, β3, β4, β5, and β7) within the catalytic domain and the catalytic triad residues (β4-D, β5-E, β7-D) thought to play a significant role in catalysis were identified among the *asr* genes used in this study by alignment with *asr*1355 (Fig 4., Arguello-Morales et al. 2000, Van Hijum et al. 2006).

All members of the *asr*1355 group possess nearly identical residues within the four regions of the catalytic domain (β3, β4, β5, and β7) thought to play an important role in catalysis (Fig 4). ASR3537, however, possesses divergent residues in regions β4, β5, and β7 all downstream of the catalytic triad amino acids (D615, E673, D767) when compared to the *asr*1355 group (Fig 4). The molecular signature of *asr*3537 in key regions of the catalytic domain is quite different from members of the asr1355 group and shows an increasing level of divergence from β4 to β7. For example, *asr*3537 is divergent from *asr*1355 at one position

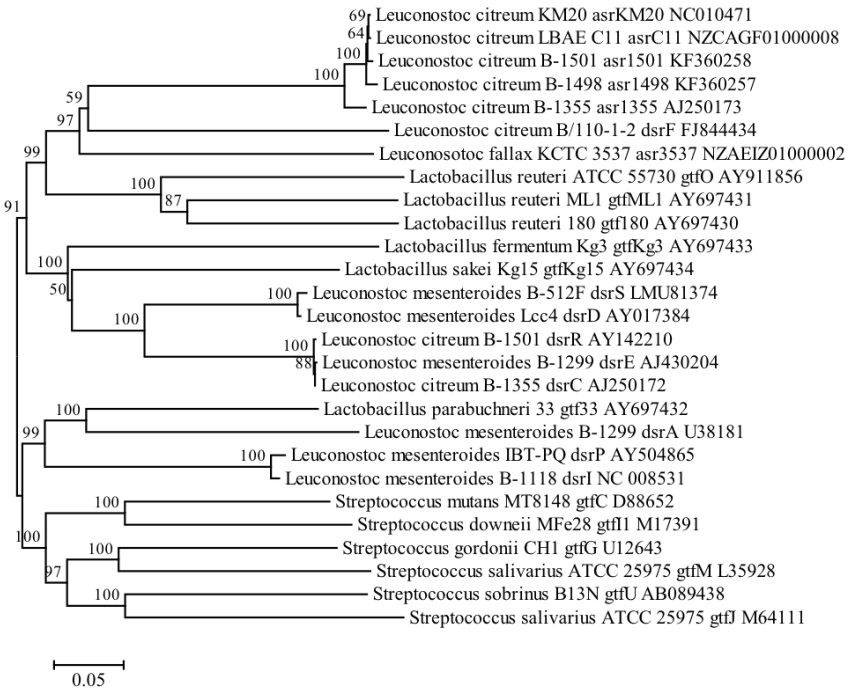

Figure 3. Phylogenetic analysis of full length amino acid sequences of select glucoside hydrolase family 70 enzymes. The tree was generated by the Neighbor-Joining Method and is drawn to scale. Bootstrap values (%) are shown above branch points (1000 replications). The scale bar represents substitutions per site. Bacterial species and strain origin, glucansucrase gene, and accession number are shown at nodes.

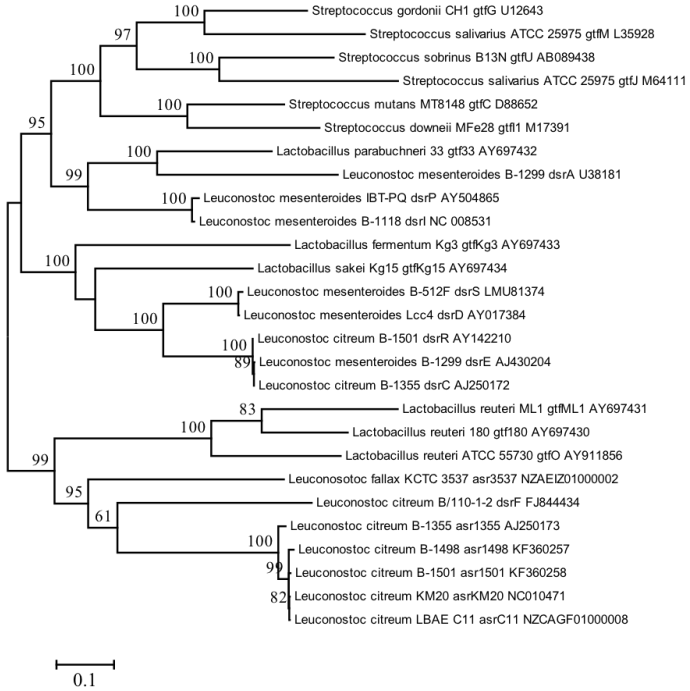

Figure S2. Phylogenetic analysis of full length amino acid sequences of select glucoside hydrolase family 70 enzymes. The tree was generated by the Maximum Likelihood Method and is drawn to scale. Bootstrap values (%) are shown above branch points (1000 replications). All bootstrap values are shown. The scale bar represents substitutions per site. Bacterial species and strain origin, glucansucrase gene, and accession number are shown at nodes.

in β4 (G557), three positions in β5 (F616, V619, K621) and five positions in β7 (S687, N692, A694, S695, I696) all downstream of the catalytic triad residues (Fig 4). No divergent residues were detected in β3 region for *asr*1355 group and *asr*3537. Clearly β3 region is not important for defining "alternansucrase" or for future structure/function studies since this region is highly conserved and does not contain a catalytic triad residue.

Variation among the *asr*1355 group genes within the catalytic domain was detected but was outside of the four main conserved regions (β3, β4, β5, and β7) thought to play a direct role in catalysis (Arguello-Morales et al. 2000, Van Hijum et al. 2006). The less conserved regions included β6 through α6 (α-helix 6, 710-764, seven site variations), α8 into beginning of hinge region (834-861, seven site variations), and end of hinge into β1 region (1082-1094, six site variations). For example the sequence variation noted within region β6 through α6 (Fig 5., 710-764) includes at least one residue variation for each *asr*1355 group member used in this study. Most of the residue variations are considered semi-conserved based on amino acid biochemical properties. Alternan products from different *Leuconostoc* isolates are structurally very similar with the exception that the alternating α-(1→6) and α-(1→3) links are in different proportions (Table 1, Bounaix et al. 2009, Jeans et al. 1954, Jeans 1979, Seymore et al. 1979). Differences in alternan link proportion among *Leuconostoc* isolates may be due to one or more residue variations within the catalytic core regions of *asr* genes. Thus, the β6 through α6 sequence segment, which includes at least one residue variation for each *asr*1355 group member used in this study, may serve as an alternative to the four main conserved regions (β3, β4, β5 and β7) for future site-directed mutagenesis studies. The unique specificity of alternansucrase, which includes both the overall alternating link structure and the proportion of each link, may be due to a few amino acid differences found in the catalytic core. Consequently, comparative sequence analysis among alternansucrase genes that make alternan products with different link distributions is important to help identify divergent residues among the primary amino acid structures.

```
                    β4                    β5                    β7                    β3
ASR-1355 607-ANFDGIRVDAVDNVDADLLKI 668-HLSILEDWNGKDPQY 760-YSFVRAHDYDAQDPIRKA 1168-ADWVPDQIY
ASR-KM20 607-ANFDGIRVDAVDNVDADLLKI 668-HLSILEDWNGKDPQY 760-YSFIRAHDYDAQDPIRRA 1168-ADWVPDQIY
ASR-C11  607-ANFDGIRVDAVDNVDADLLKI 668-HLSILEDWNGKDPQY 760-YSFIRAHDYDAQDPIRRA 1168-ADWVPDQIY
ASR-1501 607-ANFDGIRVDAVDNVDADLLKI 668-HLSILEDWNGKDPQY 760-YSFIRAHDYDAQDPIRRA 1168-ADWVPDQIY
ASR-1498 607-ANFDGIRVDAVDNVDADLLKI 668-HLSILEDWNGKDPQY 760-YSFIRAHDYDAQDPIRRA 1168-ADWVPDQIY
ASR-3537 546-ANFDSIRIDAVGNVDADLLKI 607-HISILENWDFDSVDK 679-YSFIRAHDSKVQENLASI 1065-ADWVPDQLY
           ****.**:*** ********     *:****:*: .. :     ***:**** ..*: :       *******:*
```

Figure 4. A comparison of the four highly conserved regions (β4,β5,β7,β3) within the catalytic core of GH70 enzymes and putative alternansucrase enzymes. The GH70 catalytic triad residues (D,E,D) thought to be directly involved with catalysis are shown in bold. Nonconserved divergent residues of ASR-3537 versus ASR-1355 are shown in bold and underlined. Nonconserved divergent residues of ASR-1355 versus consensus sequence are underlined.

```
1355  710-NNRSNMWYFLDTGYYLNGDLNKKIVDKNRPNSGTLVNRIANSGDTKVIPNYSFVR
KM20  710-NNRSNMWYFLDTGYYLNGDINKKIVDKNRQNSGTLVNRIANAGDTQVIPNYSFIR
C11   710-NNRSNMWYFLDTGYYLNGDINKKIVDKNRQNSGTLVNRIANAGDTQVIPNYSFIR
1501  710-NNRSNMWYFLDTGYYLNGDLNRKIVDKNRQNSGTLVNRIANAGDTQVIPNYSFIR
1498  710-NDRSNMWYFLDTGYYLNGDLNKKIVDKNRQNSGTLVNRIANAGDTQVIPNYSFIR
        *:******************:*:******* ***********:***:*******:*
```

Figure 5. Region of amino acid sequence variation among the ASR-1355 group that is located outside of catalytic regions (β3, β4 β5, and β7) in β6 through beginning of β7. Variable residues among the ASR-1355 group consensus are shown in bold.

Table 2. C-terminal APY repeat elements associated with *Leuconostoc citreum* glycoside hydrolases.

ASR-1355 group	ASR-3537	DSR-F	ISL-A (GH68)
1508-DGLFANAPY	1407-DGIYKDAPL	1377-DGLFANAPY	953-DGLFLNAP
1600-DGLFLNAPY	1499-DGFYLSAPY	1471-DGLFLNAPY	1032-DGLFLNAPY
1679-DGLFLNAPY	1581-DGMYLNAPY		1111-DGLFLNAPY
1757-DGLFLNAPY	1660-DGMYLNAPY		1190-DGLFLNAPY
1836-DGLWLNAPY	1737-DGMYLNAPY		1269-DGLWLNAPY
1916-DGMFKTAPY	1816-DGIYLNAPY		1348-DGMFKTAPY
1995-DGVFSGAPY			1428-DGVFSGAPY

ASR-1355 group (Janecek et al. 2000), ASR-3537 (Nam et al. 2011), and DSR-F (Vidal et al. 2011) are family GH70 enzymes and ISL-A (Olivares-Illana et al. 2003) belongs to family

C-Terminal Region of Predicted Alternansucrase Gene Products

Comparative analysis of C-terminal protein region indicates that the *asr*1355 group harbors two different amino acid repeat elements that include the CW-like repeat which precedes the APY repeat elements (Fabre et al. 2006, Janecek 2000, Joucla et al. 2006). The CW-like repeat (cell-wall binding) is a conserved region of 20 amino acids in length containing a high proportion of aromatic and glycine residues. The APY repeat is a conserved region of 79 amino acids containing three contiguous residues of alanine (A), proline (P), and tyrosine (Y). The *asr*1355 group harbors seven APY repeat elements, however, *asr*3537 possesses six (Table 2). The APY repeat element was first described in ASR from *L. citreum* B-1355 but has also been found in another GH70 protein DSR-F from *L. citreum* B/110-1-2 (Table 2, 2 APY repeats, Vidal et al. 2011). The APY repeat has also been found in a GH68 protein inulosucrase (ISLA) from *L. citreum* CW28 (Table 2, seven APY repeats, (Olivares-Illana et al. 2003). Function of APY repeat is unknown but may be associated with enhancing thermostability of inulosucrase (Joucla et al. 2006) and/or cell wall anchoring in ASR (Fabre et al. 2006, Vidal et al. 2011). The C-terminal, CW-like repeat elements seem to influence ASR enzyme activity much more than the APY repeats (Fabre et al. 2006). Deletion of C-terminal CW and APY repeats resulted in a loss of 99.5 % activity but deletion of

only the APY results in all activity restored (Fabre et al. 2006). The APY repeats do not seem to play a role in the unique ASR specificity or in deciding the distribution of the alternating α-(1→6) and α-(1→3) links among different ASR sources (Fabre et al. 2006).

CONCLUSIONS

Phylogenetic and sequence analyses indicated that *asr*1498, *asr*1501, *asr*C11, and *asr*KM20 possess the same molecular features as *asr*1355 and should therefore be designated "alternansucrase". Although isolated from three different countries, members of the *asr*1355 group were highly conserved in the four key regions (β3, β4, β5 and β7) within the catalytic core and possess features that are distinctive among the GH70 family enzymes. This study also identified less-conserved regions within the catalytic domain which may present new targets for structure/function studies and provide insight into why ASRs from different *Leuconostoc* isolates make different linkage proportions in alternan products.

On the other hand, *asr*3537 is more distantly related and possesses molecular features that are distinct from the *asr*1355 group including a longer signal peptide, divergent residues in key catalytic domain regions, and fewer APY repeat elements in the C-terminal domain. Consequently, the gene from *L. fallax* appears atypical and, at the very least, should be designated so when annotated as an "alternansucrase." The protein product of *asr*3537 may not produce a "typical" alternan product (alternating α-(1→6) and α-(1→3) links (Côté and Robyt 1982) based on the conserved molecular features of the *asr*1355 group. The only way to verify this information is for an additional study to clone and express the protein product of *asr*3537 and perform enzymatic studies since the gene product has not been determined. This additional study will help determine if *asr*3537, with distinct molecular features, can produce a protein product with similar function as the apparent *asr*1355 grouping.

With tremendous influx of new sequence information due to next generation sequencing technology accurate annotation of structural genes becomes extremely important particularly for genome mining applications. This study provided a consensus of molecular features for "alternansucrase" that should be considered during annotation of sequence information. In addition, potential new targets for structure/function studies were identified. Knowledge of structure/function relationships will allow the development of alternan polymers that vary in proportions of α-(1→6) and α-(1→3) glycosidic links and how these modifications alter chemical properties. The outcome may result in the production of tailor-made glucans with chemical properties specific for certain applications.

ACKNOWLEDGMENTS

This study was funded, in part, by the University Research Council of Western Illinois University.

LITERATURE CITED

Alsop RM (1983) Industrial Production of Dextrans. Progress in Industrial Microbiology, In Bushell ME (ed) Elsevier Science Publishers, pp1-42.

Amari M, Gabriel V, Robert H, Morel S, Moulis C, Gabriel B, Remaud-Simeon M, and Fontagne-Faucher C (2014) Overview of the glucansucrase equipment of *Leuconostoc citreum* LBAE-E16 and LBAE-C11, two strains isolated from sourdough. FEMS Microbiol Lett 362:1-8.

Arguello-Morales MA, Remaud-Simeon M, Pizzut S, Sarcabal P, Monsan PF (2000) Sequence analysis of the gene encoding alternansucrase, a sucrose glucosyltransferase from *Leuconostoc mesenteroides* NRRL B-1355. FEMS Microbiol Lett 182:81-85.

Arigmon S, Alekseyenko AV, DeSalle R, Caufield PW (2013) Phylogenetic analysis of glucosyltransferases and implications for the coevolution of mutans streptococci with their mammalian hosts. PLoS ONE 8(2). e56305 doi:10.1371/journal.pone.0056305.

Bounaix M-S, Gabriel V, Morel S, Robert H, Rabier P, Remaud-Siméon M, et al (2009) Biodiversity of exopolysaccharides produced from sucrose by sourdough lactic acid bacteria. J Ag Food Chem 57:10889-10897.

Cargill, Inc. (2009) FDA CFSAN/Office of food additive safety agency response letter to Cargill, Inc. regarding Notification of GRAS determination for sucromalt. Gras Notice No 000258.

CAZY (2015) Carbohydrate Active Enzymes. Architecture et Fonction des Macromolécules Biologiques (AFMB). http://www.cazy.org/. Accessed July 2015.

Choi I-K, Jung S-H, Kim B-J, Park S-Y, Kim J, Han H-U (2003) Novel *Leuconostoc citreum* starter culture system for the fermentation of kimchi, a fermented cabbage product. Antonie Van Leeuwenhoek 84:247-253.

Côté, GL, Robyt JF (1982) Isolation and partial characterization of an extracellular glucansucrase from *Leuconostoc mesenteroides* NRRL B-1355 that synthesizes an alternating (1→6),(1→3)-D-glucan. Carbohydr Res 101:57-74.

Côté GL, Holt SM (2007) Prebiotic oligosaccharides via alternansucrase acceptor reactions. U.S. Patent, 7182954.

Coutinho PM, Henrissat B (1999) Carbohydrate active enzymes: an integrated database approach. In: Gilbert HJ,

Davis GJ, Henrissat B, Svensson B. editors. Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry, Cambridge, United Kingdom, pp 3-12.

Day D, Chung C (2007) Isomaltoologosaccharides from *Leuconostoc* as neutraceuticals. U.S. Patent, 7291607.

De Man J C, Ragosa M, Sharpe M E (1960) A medium for the cultivation of lactobacilli. J Appl Bacteriol 23:130-135.

Dolan LC, Gietl E, La Cognata U, Landschutze V, Matulka RA. Safety evaluation of Fibermalt (2012) Food Chem Toxicol 50:2515–2523.

Fabre, E. Joucla G, Moulis C, Emond S, Richard G, Potocki-Veronese G, et al (2006) Glucansucrases from GH family 70: What are the determinants of their specificity? Biocatal Biotransform 24:137-145.

Gabriel V, Lefebvre D, Vayssier Y, Faucher C (1999) Characterization of the microflora from natural sourdoughs. Microbiol Aliments Nutr 17:171-179.

Grönwall, JT & Ingelman BG-A (1948) Manufacture of infusion and injection fluids. U.S. Patent, 2437518.

Hoshino T, Fujiwara T, Kawabata S (2012) Evolution of cariogenic character in *Streptococcus mutans*: orizontal transmission of glycosyl hydrolase family 70 genes. Sci Rep 2:518. doi:10.1038/srep00518

Irague R, Rolland-Sabaté A, Tarquis L, Doublier JL, Moulis C, Monsan P, et al (2012) Structure and property engineering of α-D-glucans synthesized by dextransucrase mutants. Biomolecules 13:187-195.

Janecek S, Svensson B, Russel RRB (2000) Location of repeat elements in glucansucrases of *Leuconostoc* and *Streptococcus* species. FEMS Microbiol Lett 192:53-57.

Jeanes A, Haynes WC, Wilham CA, Rankin JC, Melvin EH, Austin MJ, et al (1954) Characterization and classification of dextrans from ninety-six strains of bacteria. J Am Chem Soc 76:5041-5052

Jeanes A (1979) The α-D-glucopyranosidic linkages of dextran: comparison of percentages from structural analysis by periodate oxidation and by methylation. Carbohydr Res 74:31-40.

Joucla G, Pizzut S, Monsan P, Remaud-Simeon M (2006) Construction of a fully active trun-

cated alternansucrase partially deleted of its carboxy-terminal domain. FEBS Lett 580:763-768.

Kim JF, Jeong H, Lee JS, Choi SH, Ha M, Hur CG, et al (2008) Complete genome sequence of *Leuconostoc citreum* KM20. J Bacteriol 190:3093-3094.

Korakli M, Vogel RF (2006) Structure/function relationship of homopolysaccharide producing glycansucrases and therapeutic potential of their synthesized glycans. Appl Microbiol Biotechnol 71:790-803.

Kossmann J, Welsh T, Quanz M, Knuth K. (2000) Nucleic acid molecules encoding alternansucrase. WO/2000/047727.

Lappan, RE, Fogler HS (1995) Reduction of porous media permeability from in situ *Leuconostoc mesenteroides* growth and dextran production. Biotechnol Bioeng 50:6-15.

Martinez-Murcia AJ, Collins MD (1991) A phylogenetic analysis of an atypical *Leuconostoc*: Description of *Leuconostoc fallax* sp. nov. FEMS Microbiol Lett 82:55-60.

McCleskey CS, Faville LW, Rex RO (1947) Characteristics of *Leuconostoc mesenteroides* from cane juice. J Bacteriol 54:697-708.

Monchois, V, Willemot, R-M, Monsan, P (1999) Glucansucrases: mechanism of action and structure-function relationships. FEMS Microbiol Rev 23:131-151.

Monsan PF, Paul F (1995) Oligosaccharides feed additives. In: Wallace RJ, Chesson A, (ed) VCH, Weinheim, Germany. Biotechnology in animal feeds and animal feeding. pp 233-245.

Nam SH, Choi SH, Kang A, Kim DW, Kim DS, Kim RN et al (2011) Genome sequence of *Leuconostoc fallax* KCTC 3537. J Bacteriol 193:588-589

Olivares-Illana V, Lopez-Munguia A, Olvera C. (2003) Molecular characterization of inulosucrase from *Leuconostoc citreum*: A fructosyltransferase within a glucosyltransferase. J Bacteriol 185:3606-3612.

Paul F, Lopez-Munguia A, Remaud M, Pelenc V, & Monsan P (1992) Method for the production of α(1-2) oligodextrans using *Leuconostoc mesenteroides* B-1299, U.S. Patent, 5141858.

Remaud-Simeon, M, Willemot, R-M, Sarcabel, P, de Montalk, GP, Monsan P (2000) Glucansucrases: molecular engineering and oligosaccharide synthesis. J Mol Catal B: Enzym 10:117-128.

Seymour FR, Knapp RD, Chen ECM, Bishop SH, Jeanes A (1979) Structural analysis of *Leuconostoc* dextranscontaining 3-O-α-D-glucosylated α-D-glucosyl residues in both linear-chain and branch-point positions, or only in branch-point positions, by methylation and by ^{13}C-N.M.R. spectroscopy. Carbohydr Res 74:41-62.

Van Hijum SAFT, Krajk S, Ozimek LK, Dijkhuizen L, van Geel-Schutten IGH (2006) Structure-function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. Microbiol Mol Biol RevM70:157-176.

Vandamme EC, Renard F, Arnaut N, Vekemans, Tossut P (2003) Process for obtaining improved structure build-up of baked products. U.S. Patent, 6627235.

Vidal RF, Moulis C, Escalier P, Remaud-Simeon M, Monsan P (2011) Isolation of a gene from *Leuconostoc citreum* B/110-1-2 encoding a novel dextransucrase enzyme. Curr Microbiol 62:1260-1266.

von Heijne G, Abrahmsen L (1989) Species specific variation in signal peptide design. Implications for protein secretion in foreign hosts. FEBS Lett 244:439-446.

von Heijne G (1990) The signal peptide. J Membr Biol 115:195-201.