Microbiology, Vol. 69, No. 5, 2000, pp 575–581. Translated from Mikrobiologiya, Vol. 69, No. 5, 2000, pp. 686–693.


EXPERIMENTAL ARTICLES

# Establishment of the Phylogenetic Relationships between the Microbial Producers of Cyclodextrin Glucanotransferases Using Complete Amino Acid Sequences

**S. M. Bikbulatova*, A. V. Chemeris*, N. G. Usanov**, and V. A. Vakhitov***

**Department of Biochemistry and Cytochemistry, Ufa Research Center, Russian Academy of Sciences, Ufa, 450054 Russia*

***Institute of Biology, Ufa Research Center, Russian Academy of Sciences, pr. Oktyabrya 69, Ufa, 450054 Russia*

Received November 5, 1999; in final form, February 21, 2000

**Abstract**—A phylogenetic tree was constructed on the basis of the amino acid sequences of the known cyclodextrin glucanotransferases (CDGTs), including those deduced from the nucleotide sequences of *Bacillus* sp. strain 6.6.3 and *Paenibacillus macerans* IB-7 genes encoding α- and β-CDGTs. The tree clearly demonstrates the existence of distinct phylogenetic groups of CDGT-producing microorganisms and the divergence of the α-, β-, and γ-CDGT produced by microorganisms from the genera *Bacillus, Paenibacillus, Brevibacillus*, and *Thermoanaerobacter* from a common ancestor, whereas the CDGT of *Klebsiella pneumoniae* is independent and results from the convergence of different ancestors. The degree of homology of the leader peptide sequences of CDGTs may serve as a criterion of intraspecies relatedness between CDGT-producing microorganisms.

*Key words*: cyclodextrin glucanotransferase, *Bacillus, Paenibacillus*, phylogenetic analysis

About 70 soil microorganisms belonging to the genera *Bacillus*, *Paenibacillus*, *Brevibacillus*, *Micrococcus*, *Thermoanaerobacter* and others are presently known to produce cyclodextrin glucanotransferases (CDGT, EC 2.4.1.19) [1], which degrade starch to cyclodextrins with the formation of 1,4- α-D-glucosidic bonds. Cyclodextrins are widely used in biotechnology, as well as in food, chemical, pharmaceutical, and other industries, due to their unique ability to form complexes with many organic molecules and thus improve their properties [2]. α-, β-, and γ-cyclodextrines, which contain 6, 7, and 8 glucose residues, respectively, are the most abundant cyclodextrins. CDGTs from different organisms differ in the proportion of α-, β-, and γ-cyclodextrin they produce from starch.

The amino acid sequences of nineteen different CDGT were deduced from the nucleotide sequences of their genes cloned in foreign organisms. The X-ray diffraction analysis of some CDGTs showed that they have a specific domain structure [3–5].

The aim of the present work was to elucidate the phylogenetic relatedness between the microbial producers of cyclodextrin glucanotransferases, based on the comparative analysis of the nucleotide sequences of CDGT genes and the deduced amino acid sequences of their products. Relevant data for the α-CDGT of *P. macerans* IB-7 and the β-CDGT of *Bacillus* sp. strain 6.6.3 were obtained in our laboratory, and those for other CDGTs were obtained from the literature and databases available on the Internet.

## MATERIALS AND METHODS

*Bacillus* sp. strain 6.6.3 (= VKM B-1793) and *P. macerans* IB-7 (= VKM B-1794) used in this study were isolated from soil samples collected in Bashkortostan [6, 7]. The nucleotide sequences of DNA fragments containing the α- and β-CDGT genes [8, 9] were determined by the method of Sanger *et al.* [10]. The determination of open reading frames, the multiple alignment of nucleotide and amino acid sequences, and the construction of the phylogenetic tree were performed using the Lasergene software package from DMASTAR Inc. (United States). The nucleotide and amino acid sequences of CDGTs from microorganisms other than *Bacillus* sp. strain 6.6.3 and *P. macerans* IB-7 and the nucleotide sequences of 16S rRNA genes were taken from the literature, DNA Data, and GenBank databases.

## RESULTS AND DISCUSSION

**Analysis of the domain structure of CDGTs.** CDGTs refer to a large group of starch-hydrolyzing enzymes, such as α-amylase, glucoamylase, and pullulanase [3]. The X-ray diffraction analysis of the tertiary structure of some of these hydrolases showed that they have a domain structure. For instance, CDGTs have five

0026-2617/00/6905-0575$25.00 

```
DPDIAVTNKQSFSTDVIYQVFTDRFLDGNPSNNPTGAAYDATCSNLKLYCGGDWQGLINKINDNYFSDLGVTALWISQPVENIFATIN-
...T.....-
.A.T.....N.....I.....F.G..T..R.....-
.A.T.....N.....F.G.....V.....-
A..TS.S...N.....S....A.....F.GS.T..R.....I.....G..TGM.I..I.....YSV..-
A..TS.S...N.....I.....S....A.....F.GS.T..R.....I.....G.LTGM.I..I.....YSV..-
A..TS.S...N.....I.....S....A.....P.F.G..T..R.....I.D....G.LTGM....I.....YSV..-
A..TS.S...N.....I.....S....A.....F.G..T..R.....I.....G.LTGM....I.....YSI...
----D....VNY.K.....IV....S....G...S..IFSQN.ID.HK.....I.D....G.LT...I.....VY.LHP-
----D....VNYTR.....IV....S..D.....I.SQD..D.HK.....I.D....G.LT...I..I.....VY.LHP-
---AGNL..VN.TS..V..IVV...V...T....S..LFSSG.T..RK.....I.....G.LT.M....I.....V.SVM.D
----S....VNY.K.....IV....S....A...S..IFSQN..D.HK.....I...M..G.LT...I.....VY.LHP-
---NENLDNVNYAQEI...IV....Y..D.T...E.TLFSPG.LD.TK.....V.E..E.G.LP.M.I..I...P.I..VMELHP-
S..TS.D..VD.....IV....A..DRT...A.D.FSGDR.....F.....I.D....G.GTGM.....TSV.K-
S..TS.D..VD.....IV....A..DRA...A.D.FSGDR.....F.....I.D....G.GTGM.....TSV.K-
S..TS.D..VN.....IV....A..DRT...A.D.FSGDR.....F.....I.D....G.LTGM.....TSV.K-
S..TS.N..LN....TV..IV....V...SA.....FSSDH.....F.....IT.....G.LTGM.I.....T.V..-
--AEPEETYLD.RKET..FL.L...S..D....AGFNSATYDPN...K.T...LR.....LP--.LKS....SI..TP.ID.VNN.--D
A..TS.S.VVNY.....IV.....DA..P.HTS..K.F.....I.....G.LTGM.I..I.....Y.VLPD

YSGVTNTAYHGYWARDFKKTNPYFGTMADFQNLITTAHA--KGIKIIIDFAP 138 Bacillus sp. 6.6.3.  β AN X66106
.....--.....V..... 138 B. circulans 8  β AN X68326
.G..I.....S....--.....V..... 138 Bacillus sp. ck104  β AN L25256
.....T.....V.....--..... 138 B. licheniformis  β AN X15752
....H..P.....AY...E..K...D....--HD..V..... 138 Bacillus sp. 1011  β AN M17366
....H..P.....A....Q..K...D....--HN..V..... 138 Bacillus sp. 38-2  β AN D00129
.....-..I.....AA...--.N..V..... 137 Bacillus sp. 17-1  β AN M28053
....N.....AY..I.....AA...--.N..V..... 138 B. circulans 251  β AN X78145
-..Y.-.S.....Y.....Y.NFD..DR.MS...S--N...V.M..T. 132 Bacillus sp. 1-1  β AN D13068
-..Y.--S.....Y.R...FY.DFS..DR.MD...S--N...V.M..T. 131 B. ohbensis  β AN D90243
A..SA--S.....P..F...LS...R.VDA...--....V..... 134 B. stearothermophilus  β AN X59042
--.Y.-S.....Y.....NFS..DR.VS...N--.....M..T. 131 Brevibacillus brevis  β AN AF011388
-G.FA--S.....G....R...A..SL...SR..E...N--HD..V....V. 132 Bacillus sp.  γ AN 18991
.P..N..S.....G..DA..DF.....D....--HN..VV..... 138 P. macerans IB-7  α AN AF047363
....N..S.....G..DA..DF.....D.LTL--ITSRSDRLRPQ 138 B. macerans IAM1243  α AN M12777
....N..S.....Q..DA..DF.....D....--HN..VV..... 138 B. macerans IFO3490  α AN X59045
....N.....P.....AA..SFT..S...AA..S--HN..VVM.... 138 B. macerans  α AN E12945
AA.....G.....G..YFRIDEH..NLD..KE.TSLM.SPDYNM.LVL.Y.. 138 Klebsiella pneumoniae  α AN P08704
STFGGS.S.....P..F..SFT.....A....--HN..V..... 139 Thermoanaerobacter sp.  α AN Z35484
```

**Fig. 1.** Amino acid sequences of the A1 domains of CDGTs in single-letter notation. Identical amino acids and amino acid deletions are indicated by points and dashes, respectively. Invariant amino acids, inherent to all CDGTs under study, are italicized.

domains, A, B, C, D, and E, among which domain D separates the extended domain A into subdomains A1 and A2 of unequal sizes [4]. Because of the different functional activities of the domains, their evolutionary conservativness is also different. This explains the interest researchers show in the analysis of the amino acid sequences of CDGT domains.

Subdomains A1 of different CDGTs contain from 131 to 139 amino acid residues (AARs). The subdomain A1 of the CDGTs of strain 6.6.3 and *P. macerans* IB-7, as well as that of the majority of other CDGTs, has 138 AARs. The most conserved AARs of CDGTs are located in the middle of subdomain A1. Some subdomains A1 lack two to four amino acid residues at their upstream end, whereas subdomains A1 of the *B. stearothermophilus, Thermoanaerobacter* sp., and *Klebsiella pneumoniae* CDGTs have an additional asparagine residue in their central part. The *K. pneumoniae* subdomain A1 also has two additional amino acid residues at its downstream end. As can be seen from Fig. 1, allied CDGTs (i.e., those which produce different types of cyclodextrins in similar proportions) are distinguished by specific substitutions of certain amino acids. At the same time, subdomains A1 have 32 so-called invariant (i.e., inherent to all subdomains A1) amino acids, which made up about 23% of the amino acid residues of A1. These invariant amino acid residues are dominated by glycine, aspartate, tyrosine, and asparagine residues. Most invariant amino acid residues are located in the high-homology regions of subdomain A1, and only some are located in the variable regions of this subdomain.

Domain B is the shortest domain (Fig. 2), comprising typically 64 AARs, except the domain B of *Bacillus* sp. DSM 5850 γ-CDGT, which lacks six amino acid residues in its variable region and the domain B of

```
NHTSPAMETDTSFAENGKLYDNGTLVGGY-TND--TNGYFHHNGGSDFSSL-ENGIYKNLYDLADFNH  64
.................R..........-...--.............-.............  64
.....................N.....-...--.............-.....L..  64
.....................N.....-...--.............T.-.....L..  64
......SSD........R.....N.L...-...--.QNL...Y..T...TI-.....H...L..  64
......SSD.P......R.....N.L...-...--.QNL...Y..T...TI-.....H...L..  64
......SLDGP.....N..RDE...-...--.HNL......T...TT-.....L..  64
......SSDQP......R.......L...-...--.QNL......T...TT-.....L..  64
..S...L..NPNYV...AI....A.L.N.-S..--QQNL......T....Y-.DS..R.......YDL  64
..S...L...P.Y....AV.ND.V.I.N.-S..--P.NL......T....Y-.DS..R.......YDL  64
......S..NP.YM...R.......L...-...--A.M.......TT....-.D...R..F....L..  64
..S.L.L..NPNYV...A..N..A.L.N.-S..--R.KL......T....Y-.DS..R.......YDL  64
.....VDI------...A.....R...H.-S..--SED..YT.....Y-.DS..R......SL.Q  58
......DRDNPG.....A.....S.L.A.-S..--.A.L......T...TI-.D.....I..  64
P.V.GRAG.NPG.....A.....S.L.A.-S..--.A.L......T...TI-.D.....I..  64
......DRDNPG.....GM....S.L.A.-S..--.A.L......T...TI-.D.....I..  64
...N..SS..P......A..N....L.K.-S..--.A.L......T...TT-.S.....I.Q  64
..SNAND.-----N.F.A..RD.VFITD.P..VAANT.WY.....VTNWNDFFQVKNH..FN.S.L.Q  63
......S...PTYG...R.....V.L...-...--.......Y..TN...Y-.D...R..F....LDQ  64
```

**Fig. 2.** Amino acid sequences of the B domains of CDGTs. Designations as in Fig. 1.

```
SAA----SDADNTDFANKSGMSLLDFRFNSAVRNVFRDNTSN-MYALDSMINSTATDYNQVNDQVTFIDNHDMDRF----------
...----.....................................-.....................-----------
...----......E...E.............D........-........TG..A.....................----------
...----P.........E.....................-.......LTA..A.....................----------
VNE----ISPEYHQ...E.....AQKA.Q......D.-..G.KA.LEGSEV..A.....................----------
VNE----ISPEYHQ...E.....-..P.AQKA.Q......D.-..G.KA.LEGSEV..A.....E..----------
VNE----VS.E.HK...V.....AQK..Q..K...D.-..G.K..LEGS....A.ME.....E..----------
VNE----VSPE.HK...E.....AQK..Q......D.-..G.KA.LEGS.A..A..D.....E..----------
.GE----V.PQ.HH...E.....Q.GQTI...LK.R...-W.DFNE..T..EKE..E.I.....S..----------
.GE----V.PQ.HH...E.....Q.GQTI.D.LM.GS..-W.DFNE..A..EE..DE.I.....S..----------
ENE----V..N.HY...E.....GQKL.Q.L.N.SD.-W.GFNQ..QD..SA.DE.L.....----------
TGE----V.PQ.HH...E.....Q.GQTI.S.LK.R...-W.DFNE..K..EK..DE.I.....S..----------
AGG----..-EYHY.I.N....A....YAQV.QD.L.N.DGT-..D.ETVLRE.ESV.EKPQ.....I...----------
ADQ----T.G..IK...E...N....EYAQE..E....K.ET-.KD.YEVLA..ESQ.DYI.NM.....----------
ADQ----T.G..IK...E...N....EYAQE..E....K.ET-.KD.YEVLA..ESQ.DYI.NM.....----------
ADQ----T.G..IK...E...N....EYAQE..E....K.ET-.KD.YEVLA..ESQ.DYI.NM.....----------
PDE----MTQ..IN...Q...H....A.AQEI.E....KSET-.TD.N.V.S..GSS..YI.NM.....----------
..NTTTGV.GNAI.Y..T..SA....G.RDTLERVLVGRSGNT.KT.N.YLIKRQ.VFTSDDW..V.M.....A.IGTALRSNATTFGPGNN
TNE----V.PN..Y...E.....AQK..Q......DT-..G.....Q...A...FI..M.....----------
```

```
-----KTSAVNNRRLEQALAFTLTSRGVPAIYYGTEQYLT---------GNGDPDNRAKMPSFSKSTTAFNVISKLAPLRKSNPAIAY 204
-----.....................................---------.....................................  204
-----.....................................R---------.....................T.....................  204
-----.....................................---------.....................G.....................  204
-----H..NGDR.K.....................S...MS----------.GN......RI..S.TT...YQ..QN.....  204
-----H..NGDR.K.....................S...MS----------.GN......RI..S.TT...YQ..QN.....  203
-----HNNSA.R.K.....................MS----------.GN......RI....TT...YQ.SK.....  204
-----HA.NA.R.K.....................MS----------.GT......RI....T....YQ..Q.......C.....  204
-----SVGSSS..QTDM...VL.......T.....V.----------.GN..E..KPLKT.DR..NSYQI.....S..QT.S.LG.  204
-----SFEQSS..HTDI...VL.......T.....----------.GN..E..KP.SD.DRT.NSYQI..T..S..QN...LG.  204
-----MIDGGDP.KVDM...VL.......N.....M.----------.....N..KM.S..N.N.R.YQ..Q..SS..RN...L..  204
-----SMVVF.F-QTDI...VL.......T.....----------.GN.....KP.KT.DR..NSYKIT....S..QR.S.LG.  203
-----SRNGHST.TTDLG...L.......T......I.M.----------.D......KM.NT.D...V.YQI.QQ.SS..QE.R....  204
-----QVAGSGT.AT.....L.....M.----------.D...N...M.T..NTG...YK..QA.....  205
-----QVAGSGT.AT.....L.....M.----------.D...N...M.T..NTG...YK..QA.....  205
-----QVAGSGT.AT.....L.....M.----------.D...N...M.T..NTG...YK..QA.....  205
-----QQAGAST.PT.....V.....M.----------.....N..GM.TG.DTNK..YK..KA.....L..  205
ETGGSQSE.FAQK.IDLG.VA.M.V..I.....H.AANFTSNSFGQVGS..Y..E...G.DTESE..SI.KT.GD....S...QN 244
-----Y.GGS-T.PV.....M.----------.....Y...M.T..DTT...Y...K.....  203
```

**Fig. 3.** Amino acid sequences of the central parts of subdomains A2. Designations as in Fig. 1.

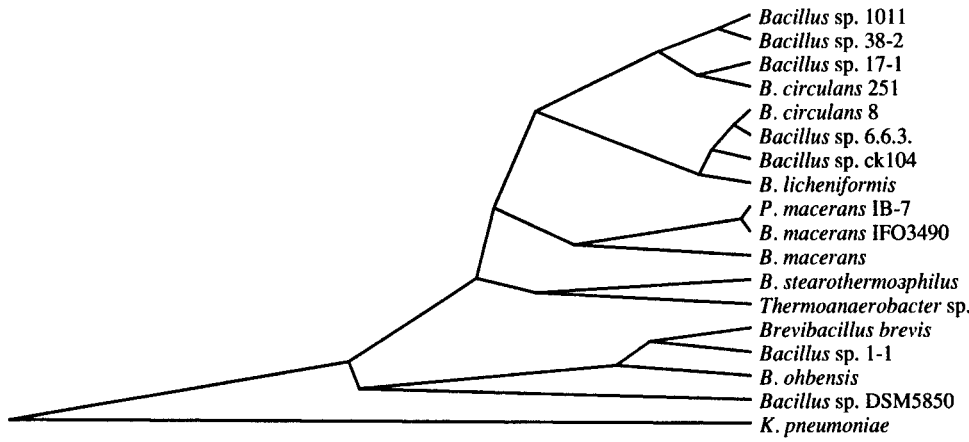


**Fig. 4.** Phylogenetic tree of CDGT-producing microorganisms from the genera *Bacillus, Paenibacillus, Brevibacillus, Thermoanaerobacter,* and *Klebsiella* derived from the analysis of the amino acid sequence homology of their CDGTs.

*K. pneumoniae* α-CDGT, which lacks five AARs in the analogous region (actually, other regions of these two domains B strongly differ). Domains B have 12 invariant amino acid residues, including four glycine residues. It is noteworthy that two adjacent histidine residues in the conservative region of γ-CDGT domains B are substituted by tyrosine and threonine residues.

Subdomain A2 is the longest CDGT domain (Fig. 3 shows only the central part of this subdomain), containing usually 203–205 AARs. The only exception is the subdomain A2 of *K. pneumoniae* CDGT, which is made up of 244 AARs and contains four insertions from 4 to 21 AARs in size. Subdomain A2 includes 43 invariant amino acid residues located in different regions of this subdomain (presumably, these regions are important for the functional activity of CDGTs). Unlike domains A1 and B, whose invariant amino acid residues are dominated by a glycine residue, the invariant AARs of subdomain A2 are dominated by the catalytically active aspartate, arginine, and asparagine residues.

Domain C differs from domains A and B in having a more conservative *N*-terminus. Amino acid homology in domain C decreases in the direction toward its *C*-terminus, especially in the case of *K. pneumoniae* CDGT, although the last 15 *C*-terminal amino acid residues exhibit relatively high homology. Domains C of most CDGTs have 88–89 AARs except those of *K. pneumoniae* CDGT and γ-CDGT, which have 86 and 90 AARs, respectively. The number of invariant amino acid residues in this domain is as low as eight.

Domain D (83–87 AARs) is absent in some starch-hydrolyzing enzymes, for instance, in α-amylase [4] and *K. pneumoniae* CDGT. Amino acid homology in domain D decreases in the direction toward its C-terminus. Like the invariant amino acid residues of domains A1 and B, the 12 invariant AARs of domain D are dominated by glycine residues.

Domain E is more homologous than domains C and D. This can be explained by the possible involvement of this domain in catalytic functions. Homology within domain E decreases in the direction toward its *C*-terminus, especially in the case of the domain E of *K. pneumoniae* CDGT, which has 98 AARs in comparison with 101 to 105 AARs present in other CDGTs. The eleven invariant amino acid residues of this domain are located in its upstream half.

Analysis of the homology of different CDGTs and some of their domains[1] led to some generalizations. In particular, mature CDGTs has 118 invariant amino acid residues (the *K. pneumoniae* CDGT being disregarded, their number increases to 173), of which 20% are glycine residues. The absence of domain D in the α-CDGT of *K. pneumoniae* may be indicative of its insignificant functional role. Noteworthy is the absence of six amino acid residues in the variable region of the domain B of γ-CDGT. However, the γ-specificity of this enzyme can hardly be related only to the deletion mentioned, since there are a number of amino acid substitutions in this CDGT.

The comparative analysis of the amino acid sequences of mature CDGTs allowed for all the CDGT-producing microorganisms under study to be divided into the following six groups: (I) *Bacillus* sp. (strain 6.6.3.), *B. circulans* (strain 8), *B. licheniformis, Bacillus* sp. (strain ck104); (II) *Bacillus* sp. (strains 1011, 38-2, 17-1), *B. circulans* (strain 251); (III) *B. stearothermophilus, B. ohbensis, Bacillus* sp. (strain 1-1); *Brevibacillus brevis, Thermoanaerobacter* sp.; (IV) *Bacillus* sp. (strain DSM 5850); (V) *P.* (B.) *macerans* (several strains); and (VI) *K. pneumoniae.*

All but one of the enzymes produced by microorganisms of the first three groups are β-CDGTs. Group IV includes γ-CDGT-producing bacteria, and the last two groups are made up of α-CDGT-producing species. The appropriateness of this conventional division is

[1] Complete information on the comparative analysis of all CDGT domains is given on the laboratory's web site: http://www.anrb.ru/molgen/pgs&e/cgtase.htm

Amino acid homology of the domains of some CDGTs with those of the β-CDGT of *Bacillus* sp. 6.6.3. (%)

| Strain | Domains | | | | | | CDGT |
|---|---|---|---|---|---|---|---|
| *Bacillus* sp. 6.6.3. (β) | A1 | B | A2 | C | D | E | |
| *B.circulans* 8 (β) | 98.5 | 98.4 | 99.0 | 93.2 | 97.7 | 95.2 | 97.7 |
| *B.licheniformis* (β) | 94.9 | 95.3 | 94.1 | 85.2 | 93.0 | 84.6 | 91.5 |
| *Bacillus* sp. 38-2 (β) | 76.0 | 75.0 | 66.2 | 67.0 | 53.5 | 75.0 | 68.8 |
| *Thermoanaerobacter* sp. (α) | 66.6 | 71.8 | 72.0 | 64.8 | 50.0 | 73.8 | 67.4 |
| *P. macerans* (α) | 70.3 | 70.3 | 65.2 | 61.4 | 65.1 | 49.0 | 66.4 |
| *Bacillus* sp. 1-1 (β) | 62.3 | 60.9 | 61.3 | 51.1 | 41.8 | 55.7 | 59.8 |
| *Brevibacillus brevis* (β) | 64.5 | 64.0 | 60.3 | 53.4 | 51.1 | 58.6 | 59.2 |
| *Bacillus* sp. (γ) | 50.0 | 60.9 | 55.9 | 44.3 | 38.4 | 55.7 | 50.4 |
| *K. pneumoniae* (α) | 39.1 | 28.1 | 23.0 | 19.3 | – | 25.0 | – |

Note: The homology of the whole α-CDGT of *K. pneumoniae* cannot be determined because this enzyme lacks domain D.

confirmed by constructing the phylogenetic tree of CDGT-producing microorganisms on the basis of the comparative analysis of the amino acid sequences of mature CDGTs (Fig. 4). It can be seen that α-CDGTs largely diverge from other bacillar CDGTs. The CDGT of *Thermoanaerobacter* sp. is close to that of *B. stearothermophilus*, and the CDGT of *Brevibacillus brevis* is close to the enzymes produced by *Bacillus* sp. 1-1 and *B. ohbensis*. At the same time, the three α-CDGTs of *P. (B.) macerans* strains, two of which are virtually identical and a third enzyme that has a great number of substitutions, form a distinct group subdivided into two subgroups.

The evolutionary conservativeness of CDGT domains is apparent from the table, which shows the homology levels of the amino acid sequences of some CDGTs with reference to the β-CDGT of *Bacillus* sp. 6.6.3. Domains A and B are characterized by the highest degree of homology, while domains C, D, and E are much less homologous. The amino acid homology of the whole β-CDGT of *Bacillus* sp. 1-1 is as low as 59.8%, and that of the α-CDGT of *P. macerans* IB-7 is 66.4%. The homology of the γ-CDGT of *Bacillus* sp. DSM 5850 is also low. In spite of the low homology of *K. pneumoniae* CDGT, it however contains regions showing a high homology with other CDGTs, suggesting a convergent character of its evolution. It should be noted that all domains are characterized by a great number of substitutions of similar amino acids, such as leucine, isoleucine, and valine.

**Analysis of the leader peptide sequences of CDGTs.** The homology of the leader peptide sequences of CDGTs, which are less conservative in their evolution than other parts of these enzymes, may serve as a criterion of the intraspecies relatedness of CDGT-producing microorganisms from the genera *Bacillus* and *Paenibacillus*. The leader peptides of the CDGTs studied, which have several positively charged amino acid residues at the *N*-terminus, hydrophobic central part, and alanine residue at the *C*-terminus, are typical of all gram-positive bacteria. In Fig. 5, leader peptides are grouped in accordance with their degree of amino acid sequence homology.

In the first group, leader peptide sequences are 34 AARs in size, with a homology varying from 82 to 97%.

In the group of CDGTs from alkaliphilic strains, leader peptide sequences are shorter (27 AARs) and widely vary in homology (from 55 to 100%).

The leader peptide sequences of CDGTs from *Bacillus* sp. 1-1 and *B. ohbensis* have 29 AARs and possess the same starting codon, TTG. It is noteworthy that the structurally similar leader peptides of *B. stearothermophilus* CDGT and of the unique γ-CDGT of *Bacillus* sp. DSM 5850 have different sizes (31 and 28 AARs, respectively).

Of particular interest is the group of the *P. (B.) macerans* α-CDGTs, among which three CDGTs have similar leader peptides, and the leader peptide of the fourth CDGT has a great number of amino acid substitutions (its homology is as low as 51.8%). Comparing this leader peptide with the leader peptides of enzymes from other groups did not reveal any regions with noticeable homology, suggesting that the strain producing this fourth CDGT belongs to a distinct bacterial group within the species *P. (B.) macerans*. We failed to reveal any leader peptide homology between different groups of bacillar CDGTs, nor was there any homology between the CDGTs produced by members of other genera, *K. pneumoniae, Brevibacillus brevis*, and *Thermoanaerobacter* sp. (relevant data are not provided in this paper).

From the foregoing, it can be inferred that the homology of the leader peptides of CDGTs may serve as a criterion for establishing the intraspecies relatedness of CDGT-producing microorganisms. The phylogenetic trees constructed using homology data for mature CDGTs and their leader peptides (Fig. 6) do not substantially differ, which is further evidence of the

```
   (34 AARs)
 1 MFQMAKRAFLSTTLTLGLLAGSALPFLPASAAYA  Bacillus sp. 6.6.3.
 2 ..............................V..  B.circulans 8
 3 .......VL......FS.............I..  B.licheniformis
 4 .......VL.......S....G...Y.....I..  Bacillus sp. ck104

   (27 AARs)
 5 MKRFMKLTAVWTLWLSLTLGLLSPVHA        Bacillus sp. 1011
 6 ...........................        Bacillus sp. 38-2
 7 ..KIS...TALA.S...A.S..G.A..        Bacillus sp. 17-1
 8 ..K.L.S..ALA.G..........AQ.        B.circulans 251

   (29 AARs)
 9 LNDLNDFLKTILLSFIFFLLLSLPTVAEA      Bacillus sp. 1-1
10 .KN.TVL....P.ALLL.I......A.Q.      B.ohbensis

   (31 and 28AARs)
11 MRRWLSLVLSMSFVFSAIFIVSDTQKVTVEA    B.stearothermophilus
12 .I.R..FS.VVL.LI.FLV..-NPEY-.-..    Bacillus sp. DSM5850

   (27 AARs)
13 MKSRYKRLTSLALSLSMALGISLPAWA        P.macerans IB-7
14 ...........................        B.macerans IAM1243
15 ...........................        B.macerans IFO3490
16 ..KQV.W...VSM.VGI...AA..V..        B.macerans
```

**Fig. 5.** Deduced amino acid sequences of the leader peptides of CDGTs. For designations, see the legend to Fig. 1.

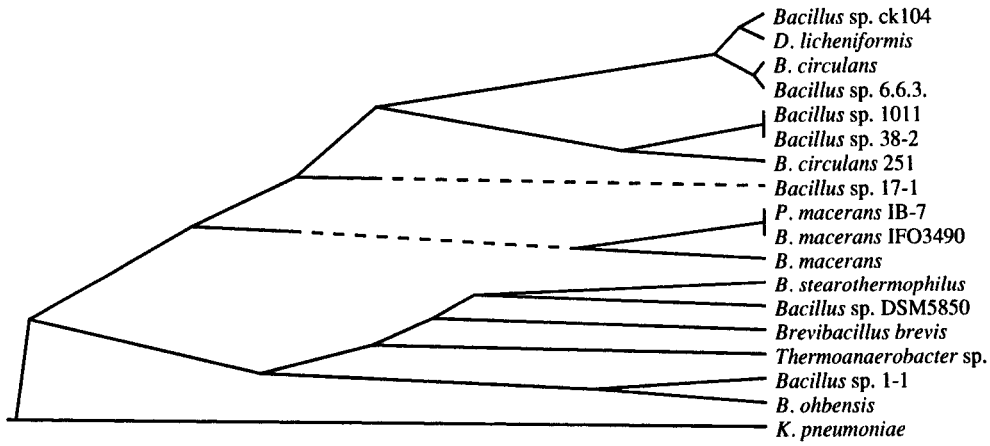


**Fig. 6.** Phylogenetic tree of CDGT-producing microorganisms from the genera *Bacillus, Paenibacillus, Brevibacillus, Thermoanaerobacter,* and *Klebsiella* derived from the analysis of homology of the leader peptide sequences of their CDGTs.

appropriateness of using short regions of leader peptides for elucidating the phylogenetic relationships of CDGT-producing microorganisms.

The phylogenetic trees constructed using the amino acid sequences of CDGTs and the nucleotide sequences of 16S rRNA of some bacilli (Fig. 7) are very similar, except for the position of *P. macerans* and *B. stearothermophilus.* This can be explained by different rates of mutations in the genes that encode macromolecules with different metabolic functions. The phylogenetic tree derived from the 16S rRNA sequence data is in good agreement with those constructed by other authors [11, 12].

Thus, the comparative analysis of amino acid sequences of α-, β-, and γ-CDGT made it possible to reveal some specific features in their structural organization, in particular, the existence of high-homology and diverged regions. The phylogenetic trees constructed in this study clearly demonstrated the existence of quite distinct phylogenetic groups of CDGT-producing microorganisms and the divergence of the α-, β-, and γ-CDGTs produced by microorganisms from the genera *Bacillus, Paenibacillus, Brevibacillus,*

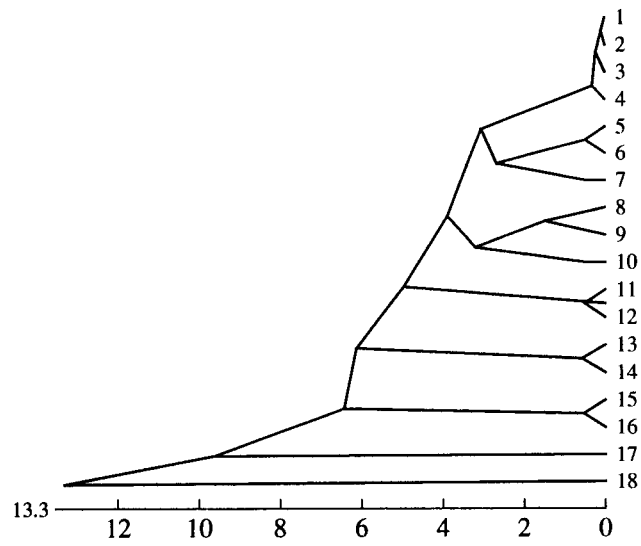


**Fig. 7.** Phylogenetic tree derived from the comparative analysis of the primary structure of 16S rRNA genes of microorganisms from the genera *Bacillus, Paenibacillus, Brevibacillus, Thermoanaerobacter*, and *Klebsiella*: (1) *B. licheniformis* B-6-4-J [AN D31739], (2) *B. licheniformis* DSM 13 [AN X68416], (3) *B. licheniformis* NCDO 1772 [AN 60623], (4) *Bacillus* sp. [AN X81131], (5) *B. circulans* NCDO 1775 [AN X60613], (6) *B. circulans* IAM 12462 [AN D78312], (7) *Bacillus* sp. DSM 3922 [AN X60602], (8) *Bacillus* sp. DSM 2349 [AN 26929], (9) *Bacillus* sp. DSM 8717 [AN X76441], (10) *Bacillus* sp. DSM 8718 [AN X76442], (11) *B. stearothermophilus* T10 [AN X57309], (12) *B. stearothermophilus* NCDO 1768 [AN X60640], (13) *B. brevis* NCIMB 9372 [AN 60612], (14) *B. brevis* JCM 2503 [AN D78457], (15) *P. macerans* NCDO 1764 [AN 60624], (16) *P. macerans* DSM 24 [AN X57306], (17) *Thermoanaerobacter* sp. Ab11 [AN U51198], and (18) *K. pneumoniae* DSM 30102 [AN X93214].

and *Thermoanaerobacter* from a common ancestor, whereas the CDGT of *Klebsiella pneumoniae* is independent and results from the convergence of different ancestors. The degree of homology of the leader peptides of CDGTs, which are evolutionary low-conservative, may serve as a criterion for establishing the intraspecies relatedness of CDGT-producing microorganisms.

## ACKNOWLEDGMENTS

This work was supported by the Russian State Scientific and Technical Program "Advanced Methods in Bioengineering" and by the State Program for the Promotion of Leading Scientific Schools, project no. 96-15-98061.

## REFERENCES

1. Mattsson, P., *The Structure-Function Relationships of Cyclomaltodextrin Glucanotransferase* from *Bacillus circulans* var. *alkaliphilus*, Turku: Turun Yliopisto, 1994, pp. 1–74.
2. Szejtli, J., *Cyclodextrins and Their Inclusion Complexes*, Budapest: Akademiai Kiado, 1982.
3. Svensson, B., Jespersen, H., Sierks, M.R., and MacGregor, E.A., Sequence Homology between Putative Raw-Starch Binding Domains from Different Starch-degrading Enzymes, *Biochem. J.*, 1989, vol. 264, pp. 309–311.
4. Jespersen, H.M., MacGregor, E.A., Sierks, M.R., and Svensson, B., Comparison of the Domain-Level Organization of Starch Hydrolases and Related Enzymes, *Biochem. J.*, 1991, vol. 280, pp. 51–55.
5. Klein, C. and Schulz, G.E., Structure of Cyclodextrin Glycosyltransferase Refined at 2.0 *A Resolution, J. Mol. Biol.*, 1991, vol. 217, pp. 737–750.
6. Usanov, N.G., Loginov, O.N., and Melent'ev, A.I., Method to Isolate *Bacillus macerans* from Soil, Patent no. 1703681, 1990.
7. Usanov, N.G., Loginov, O.N., and Melent'ev, A.I., *Bacillus* sp. Strain Producing β-Cyclodextrin Glucanotransferases, Patent no. 1738852, 1993.
8. Chemeris, A.V., Bikbulatova, S.M., and Vakhitov, V.A., Nucleotide Sequence of α-Cyclodextrin Glucanotransferase Gene from *Paenibacillus macerans, DDBJ/EMBL/GenBank*, 1998. Access number AFO47363.
9. Vakhitov, V.A., Akhmetzyanov, A.A., Bikbulatova, S.M., Usanov, N.G., and Chemeris, A.V., Molecular Cloning and Sequence Analysis of β-Cyclodextrin Glucanotransferase of *Bacillus* sp. Strain 6.6.3, *Biotekhnologiya*, 1999, no. 1, pp. 3–10.
10. Sanger, F., Nicklen, S., and Coulson, A.R., DNA Sequencing with *Chain-terminating Inhibitors, Proc. Natl. Acad. Sci. USA*, 1977, vol. 74, pp. 5463–5467.
11. Ludwig, W. and Schleifer, K.H., Bacterial Phylogeny Based on 16S and 23S rRNA Sequense Analysis, *FEMS Microbiol. Rev.*, 1994, vol. 15, pp. 155–173.
12. Nielsen, P., Rainey, F.A., Outtrup, H., Priest, F.G., and Fritze, D., Comparative 16S rRNA Sequence Analysis of Some Alkaliphilic Bacilli and the Establishment of a Sixth rRNA Group within the Genus *Bacillus, FEMS Microbiol. Lett.*, 1994, vol. 117, pp. 61–66.