



## Structure–function relationship in cyclodextrin glycosyltransferase from *Bacillus circulans* DF 9R

Hernán Costa<sup>a</sup>, Sergio del Canto<sup>b</sup>, Susana Ferrarotti<sup>a</sup>, Mirtha Biscoglio de Jiménez Bonino<sup>b,\*</sup>

<sup>a</sup> Departamento de Ciencias Básicas, Universidad Nacional de Luján, Rutas 5 y 7 (6700) Luján, Buenos Aires, Argentina

<sup>b</sup> Departamento de Química Biológica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Instituto de Química y Físicoquímica Biológicas, Junín 956 (1113) Buenos Aires, Argentina

### ARTICLE INFO

#### Article history:

Received 27 May 2008

Received in revised form 19 September 2008

Accepted 19 September 2008

Available online 27 September 2008

#### Keywords:

*Bacillus circulans*

Cyclodextrin glycosyltransferase

Histidine modification

Molecular cloning

Structure–function relationship

### ABSTRACT

Cyclodextrin glycosyltransferases (CGTases E.C.2.4.1.19) catalyze cyclomaltooligosaccharides (cyclodextrins) production, an important industrial process. We herein report structural features of *Bacillus circulans* DF 9R cyclodextrin glycosyltransferase including its sequence and several aspects of enzyme structure–function relationship. Protein ethoxyformylation, under our experimental conditions, indicated that only one out of the 13 enzyme histidines was modified leading to a drastic drop in cyclizing and hydrolytic activity. Besides, tryptic digestion of the <sup>14</sup>C ethoxyformylated protein and studies of the peptide mixture showed that histidine 233 is the most reactive histidine residue. This is the first cyclodextrin glycosyltransferase with a known primary structure and a glutamine instead of glycine residue at position 179 in the highly conserved –6 subsite, shown to be involved in substrate binding. The presence of glycine at that position was considered as a requirement for such binding following the induced-fit mechanism already proposed. Moreover, the enzyme has all the features previously described for an  $\alpha$ - or  $\alpha/\beta$ -cyclodextrin producer.

© 2008 Elsevier Ltd. All rights reserved.

### 1. Introduction

Cyclodextrin glycosyltransferases (CGTases; E.C.2.4.1.19) are members of the GH13 family, also known as the  $\alpha$ -amylase family<sup>1</sup> which catalyze cyclomaltooligosaccharides (cyclodextrins, CDs) production from starch and other related carbohydrates through intramolecular transglycosylation. Besides, the enzyme catalyzes coupling and disproportionation reactions by intermolecular transglycosylation. Moreover, CGTase possesses a weak starch hydrolyzing activity.

Cyclomaltooligosaccharides are cyclic molecules formed by  $\alpha$ -(1→4)-linked D-glucopyranosyl units with an apolar central cavity and a hydrophilic outer surface. The most common CDs are  $\alpha$ ,  $\beta$  and  $\gamma$ , with six, seven and eight D-glucopyranose units, respectively. CGTases are mainly used for CD industrial production<sup>2</sup> since they can be employed for transglycosylation from donor substrates to several compounds such as sugar, alcohols, vitamins, glycosides, polyols and flavonoids to improve their properties.<sup>3,4</sup> The major microorganisms CGTase producers are members of the *Bacilli* class, although *Anaerobranca*, *Klebsiella*, *Thermococcus*, *Thermoanaerobacter* and *Thermoanaerobacterium* are also sources of CGTases.<sup>5</sup>

In all known CGTases—except that from *Klebsiella*—five domains (A–E) can be recognized. Domain A consists of two segments, A<sub>1</sub>

and A<sub>2</sub>, involving amino acid residues 1–138 and 203–406, respectively (*Bacillus circulans* 251 mature CGTase numbering. Accession P43379). Domain A contains a catalytic ( $\beta/\alpha$ )<sub>8</sub> barrel with a loop protruding at the third  $\beta$ -strand (B domain, 139–202). Subsites +2 to –7 forming the catalytic site come from domains A and B.<sup>6</sup> Three His residues (H140, H233 and H327) are conserved in most members of the GH13 family and in all known CGTases and located at the active center in conserved regions.<sup>7,8</sup>

We previously reported the isolation and partial characterization of a CGTase from *B. circulans* DF 9R<sup>9</sup> producing mainly  $\alpha$ - and  $\beta$ CDs.<sup>10</sup> We now report the modification of such CGTase with diethyl pyrocarbonate (DEP) and the influence of such modification on enzyme activity. Besides, the protein sequence was obtained by molecular cloning, Edman degradation and molecular mass determination, and compared with protein sequences of CGTases from different origins—special emphasis being cast on product specificity determinants.

### 2. Results and discussion

#### 2.1. Cysteine oxidation degree, protein molecular mass and N-terminal amino acid sequence

Incorporation of 4-vinylpyridine (4-VP) was measured before and after protein reduction indicating the presence of one disulfide bridge, a common feature in *Bacilli* CGTases. The molecular mass of

\* Corresponding author. Tel.: +54 11 4964 8290; fax: +54 11 4962 5457.  
E-mail address: [mbiscoglio@hotmail.com](mailto:mbiscoglio@hotmail.com) (M. B. de Jiménez Bonino).

**Table 1**

N-terminal sequence of the native protein and N-terminal sequence and molecular mass of selected tryptic peptides from the EF-CGTase from *B. circulans* DF 9R

Number	Th. mass <sup>a</sup>	(M+H) <sup>+</sup>	Edman sequence <sup>b</sup>
1	944.1	945.4	<sup>1</sup> APDTSVLNKG/
2	—	—	<sup>1</sup> APDTSVLNKGQSTSDVIYQIVT <sup>22</sup> c...
3	2240.4	2346.9 <sup>d</sup>	<sup>25</sup> FADGDASNPNPSGAASFPGCTNLK <sup>47</sup> /
4	2644.9	2645.8	<sup>108</sup> TNPAFGSMSDFQELIDTAHAHNK <sup>131</sup> /
5	2685.9	2686.2	<sup>132</sup> VIIDFAPNHTSPASETQPSFAENGR <sup>156</sup> /
6	3229.4	3963.2	<sup>157</sup> LYENGSLIAGYTGDNGIFHHNQGTDFSSL <sup>186</sup> ...
7	1288.5	1304.7 <sup>e</sup>	<sup>217</sup> LWLDMGIDGIR <sup>227</sup> /
8	1523.8	1596.5 <sup>f</sup>	<sup>228</sup> VDAVKHMPQGWQK <sup>240</sup> /
9	1011.2	1083.8 <sup>f</sup>	<sup>233</sup> HMPQGWQK <sup>240</sup> /
10	2574.8	4291.1	<sup>295</sup> DNTDMNYGLNSMLAETAAEYTHI <sup>317</sup> ...
11	1015.1	1016.5	<sup>324</sup> IDNHMDMR <sup>331</sup> /
12	2003.2	2345.9	<sup>354</sup> GVPAIYGTGEQYMGAGNDP <sup>372</sup> ...
13	948.1	949.4	<sup>376</sup> AFMSSFSTA <sup>384</sup> /
14	1225.4	1226.6	<sup>382</sup> STASTAYQVIGK <sup>393</sup> /
15	966.1	967.4	<sup>385</sup> STAYQVIGK <sup>393</sup> /
16	1420.5	1421.5	<sup>400</sup> SNPAIAYGNTQER <sup>412</sup> /
17	1484.6	1485.6	<sup>413</sup> WINNDVYIYER <sup>423</sup> /
18	1004.1	1005.7	<sup>511</sup> AGTTLTIDGR <sup>520</sup> /
19	621.7	622.7	<sup>521</sup> GFGDVK <sup>526</sup> /
20	828.9	829.3	<sup>527</sup> GVFFGT <sup>534</sup> /
21	2861.2	3203.6	<sup>535</sup> AVSDSQILQWEDSQIIVPAAAPGSH <sup>561</sup> ...
22	3102.4	3231.5	<sup>591</sup> FIVNNASTVYGESVYLTGNTAELGNWAPN <sup>619</sup> ...
23	2777.1	3524.0	<sup>621</sup> AIGPFNQIITAYPSWYDVSVAG <sup>645</sup> ...
24	406.5	ND <sup>g</sup>	<sup>652</sup> FIK <sup>654</sup> /
25	2060.2	3158.2	<sup>656</sup> NGTAVTWEGGANHSFTAPVSG <sup>676</sup> ...

Numbering corresponds to that of CGTase from *B. circulans* 251.

<sup>a</sup> The value includes only those amino acid residues assigned by Edman degradation.

<sup>b</sup> The peptide sequence is already completed (/); the peptide is still longer (...).

<sup>c</sup> It was obtained from the native protein.

<sup>d</sup> The molecular mass difference corresponds to the peptide modification with 4-VP.

<sup>e</sup> The molecular mass difference corresponds to Met oxidation.

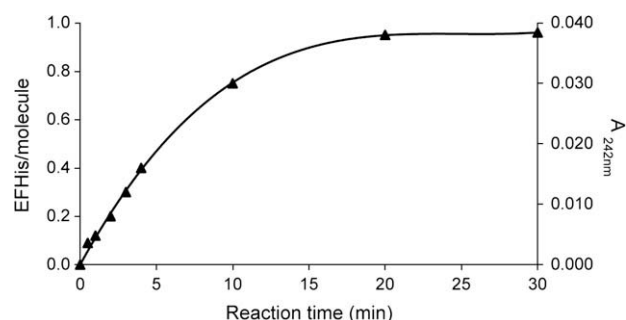
<sup>f</sup> The molecular mass difference corresponds to the incorporation of an ethoxyformyl group.

<sup>g</sup> Not determined.

74,470 Da, obtained by mass spectrometry, agrees with that of 78,000 Da previously reported as obtained by SDS–PAGE.<sup>11</sup> The N-terminal sequence of the native protein is included in Table 1, 2.

## 2.2. CGTase ethoxyformylation kinetics

Protein ethoxyformylation with DEP is frequently used for protein structure–function relationship studies.<sup>12,13</sup> The CGTase ethoxyformylation kinetics in a 100 mM phosphate buffer, pH 7.4, with a DEP/His molar ratio of 2 was determined (Fig. 1). Under the experimental conditions, only one out of the enzyme 13 His residues was modified. Circular dichroism and fourth-derivate



**Figure 1.** Time course of the reaction of DEP ( $2.1 \times 10^{-4}$  M final concentration) with His in the CGTase from *B. circulans* DF 9R ( $6.3 \times 10^{-6}$  M final protein concentration) at 25 °C.

spectra (not shown) indicated that the protein conformation was conserved. However, the modification led to a drastic decrease in the  $\alpha$ CD (63%) and  $\beta$ CD (75%) formation and in hydrolytic activity (75%). When the His modification was performed in the presence of 12.5 mM  $\alpha$ CD, the decrease in  $\beta$ CD formation and hydrolytic activity was only 35% and 25%, respectively, and 15% for both activities in the presence of 12.5 mM  $\gamma$ CD thus indicating a significant CD protective effect. Although it was also attempted to use  $\beta$ CD as protective agent during modification, its presence interferes with activity determination by the phenolphthalein assay.

## 2.3. Mass spectrometry analysis and sequencing of tryptic peptides

To elucidate which of the three conserved His residues described as relevant for enzyme activity was the most reactive against DEP and mainly accountable for the drop in catalytic activity, the  $^{14}$ C ethoxyformylated (EF)-enzyme was submitted to tryptic digestion and RP–HPLC resolution of the peptide mixture was performed (Fig. 2). Edman sequencing of the two radioactive fractions indicated that they contained H233, and molecular mass determination showed the increase of 72 Da corresponding to the incorporation of the ethoxyformyl group to an imidazolic nitrogen of the His residue (Table 1, 8 and 9).

On the other hand, Edman sequencing and molecular mass of an unlabelled fraction showed a peptide containing H140 (Table 1, 5). Another unlabelled fraction (Table 1, 10), contained a peptide starting at D295 which, as judged by its molecular mass, extends up to R331 thus including H327; that is why peptide 10 was digested with chymotrypsin, which in turn yielded sequence 11. Overall results indicated that H233 is the most reactive of the three His residues (140, 327 and 233) that are considered important for enzyme activity. We were able to find nine unmodified and one modified His residue out of the 13 in the protein. As there is no other radioactive fraction, the other three remained unmodified and we were able to localize them by further studies of molecular biology.

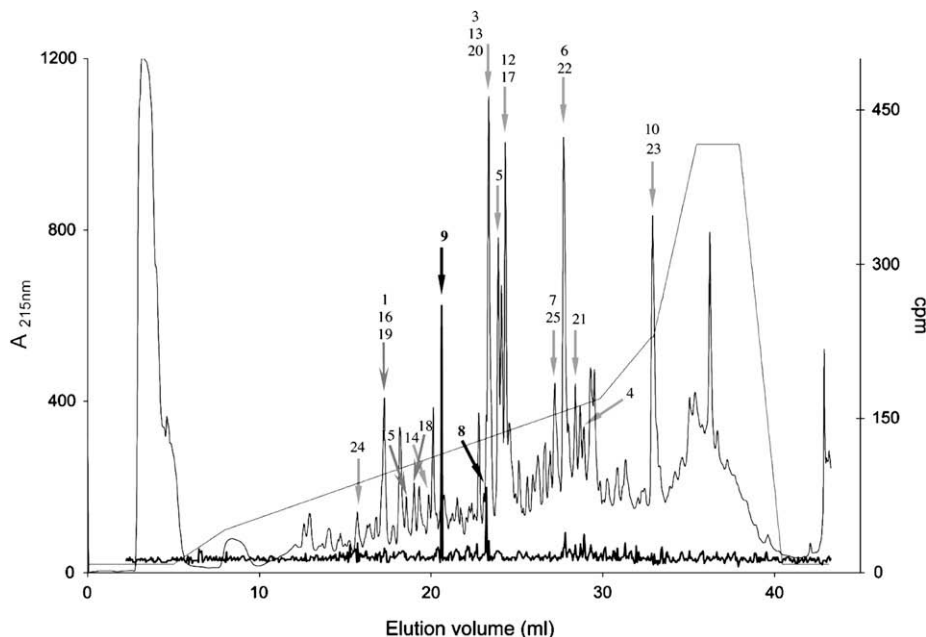
As described above, ethoxyformylation of H233 brought about a drastic drop in starch hydrolysis and  $\alpha$ - and  $\beta$ CD formation, this agreeing with results of Nakamura et al.<sup>8</sup> and Mattsson et al.<sup>14</sup> who showed that site-directed mutagenesis of each of those His leads to an important decrease in both starch hydrolysis and cyclizing activity. Moreover, Ishii et al.<sup>15</sup> studied the crystalline structure of the native and mutated (H233N) CGTase from *Bacillus* sp. 1011, and showed that hydrogen bonds between H233 and the substrate are important for  $\alpha$ CD formation.

We also took advantage of the high resolution obtained by RP–HPLC of the tryptic peptide mixture to further extend the knowledge of enzyme primary structure. All tryptic fractions were subjected to mass spectrometry; those containing only one molecular species were sequenced while simple mixtures were first rechromatographed and new fractions were also analyzed by ESIMS (Table 1). Sequence alignment showed that the protein has unique characteristics as regards subsite –6 involved in the catalytic site since Gln has replaced Gly at position 179. This led to the completion of our determination of the protein primary structure.

## 2.4. Cloning and sequencing

A nucleotide fragment (1855 bp, accession number EU644086) corresponding to a 90% of the mature CGTase-coding sequence was amplified, cloned and sequenced.

Translation of the DNA allowed us to obtain the 32–645 sequence, which fully agrees with all assignments made by Edman sequence and mass spectrometry.



**Figure 2.** RP-HPLC elution pattern (Vydac C18, 250 × 4.6 mm) of the tryptic digest of the  $^{14}\text{C}$  EF-CGTase. Flow rate 0.5 mL/min. Arrows indicate fractions yielding peptide sequences, those in black show radioactive fractions. Numbers correspond to those of purified peptides obtained from the peak (Table 1).

## 2.5. Primary structure

Figure 3 summarizes results obtained by Edman sequencing, mass spectrometry and gene cloning and sequencing, the last from N32 to G645. On the other hand, Asp and Arg residues were assigned to positions 23 and 24, respectively, since they are conserved in all CGTases reported up to now; with regard to R24, trypsin specificity was also considered. The sequence  $^{646}\text{Q-K}^{651}$  was obtained on the basis of the molecular mass of fragment 23 and that of fragment  $^{621}\text{A-G}^{645}$  whose sequence was determined by Edman methodology, plus searches for sequence identity and similarity (BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>); the  $^{646}\text{Q-K}^{651}$  sequence was found to be identical with that of *B. circulans* A11–649EFK<sup>651</sup> being conserved in all sequences of *Bacilli* CGTases. Lys was assigned to position 655 by homology and trypsin specificity. Finally, the sequence of the last 10 amino acids was proposed on the basis of the difference between the molecular mass of fragment 25 (Table 1) and that of the sequence  $^{656}\text{N-G}^{676}$ , plus searches in BLAST; 8 out of those 10 residues are fully conserved in all *B. circulans* CGTases with an already-known primary structure; such C-terminal sequence was found to be identical with that of *B. circulans* 8 CGTase. We are working on the determination of this 10-residue sequence unambiguously by Edman degradation and/or molecular biology to undertake future experiments of site-directed mutagenesis and to build a molecular model.

## 2.6. CGTase alignment and product specificity

*B. circulans* DF 9R sequence was aligned with the already reported CGTase sequences (not shown). The following points must be considered: (i) the –6 subsite (Y167, G179, G180, N193 and D196) is fully conserved (Fig. 4). It was suggested that G179 and G180 are conserved as the absence of side chains is a requirement for substrate binding favoring the induced-fit mechanism already proposed.<sup>16</sup> In this work, we describe the presence of Gln instead of Gly at position 179. This is the first change reported for subsite –6 in a natural CGTase so far—such change being determined by both Edman degradation and DNA sequencing. Moreover, Leemhuis et al.<sup>16</sup> obtained the mutant G179L of the CGTase from *B. circulans*

251 and reported that the  $\alpha$  cyclizing activity was reduced to 50% of the wild-type activity; (ii) the presence of Arg or Lys at position 47 is related to the production of  $\alpha$ - or  $\beta$ /CDs;<sup>17</sup> *B. circulans* DF 9R CGTase has Lys at that position (Table 1,3) and produces  $\alpha$ / $\beta$ CDs in a 10/7 ratio;<sup>10</sup> (iii) Kim et al.<sup>18</sup> already reported that the mutated N94S CGTase from *Bacillus* sp. I-5 produces about twice as much  $\alpha$ CD as compared to the wild type. *B. circulans* DF 9R CGTase has S94 while enzymes from other *B. circulans* having N94 are producers of  $\beta$ CD and (iv) van der Veen et al.<sup>6</sup> described that CGTases producing similar amounts of  $\alpha$ - and  $\beta$ CDs have the stretch  $^{144}\text{ASET}^{147}$ . CGTases producing predominantly  $\beta$ CD have either  $^{144}\text{ASSD}^{147}$  or  $^{144}\text{AMET}^{147}$ . The enzyme studied in this work is the only CGTase possessing  $^{144}\text{ASET}^{147}$  out of those from all strains of *B. circulans*.

In summary, this work is a contribution to the knowledge of CGTase structure–function relationship, particularly to features involved in product specificity; the mechanism by which the enzyme determines the size of the synthesized CD is still unclear up to now. Besides, findings in this paper could be the basis for further protein engineering studies leading to the production of a single CD type.

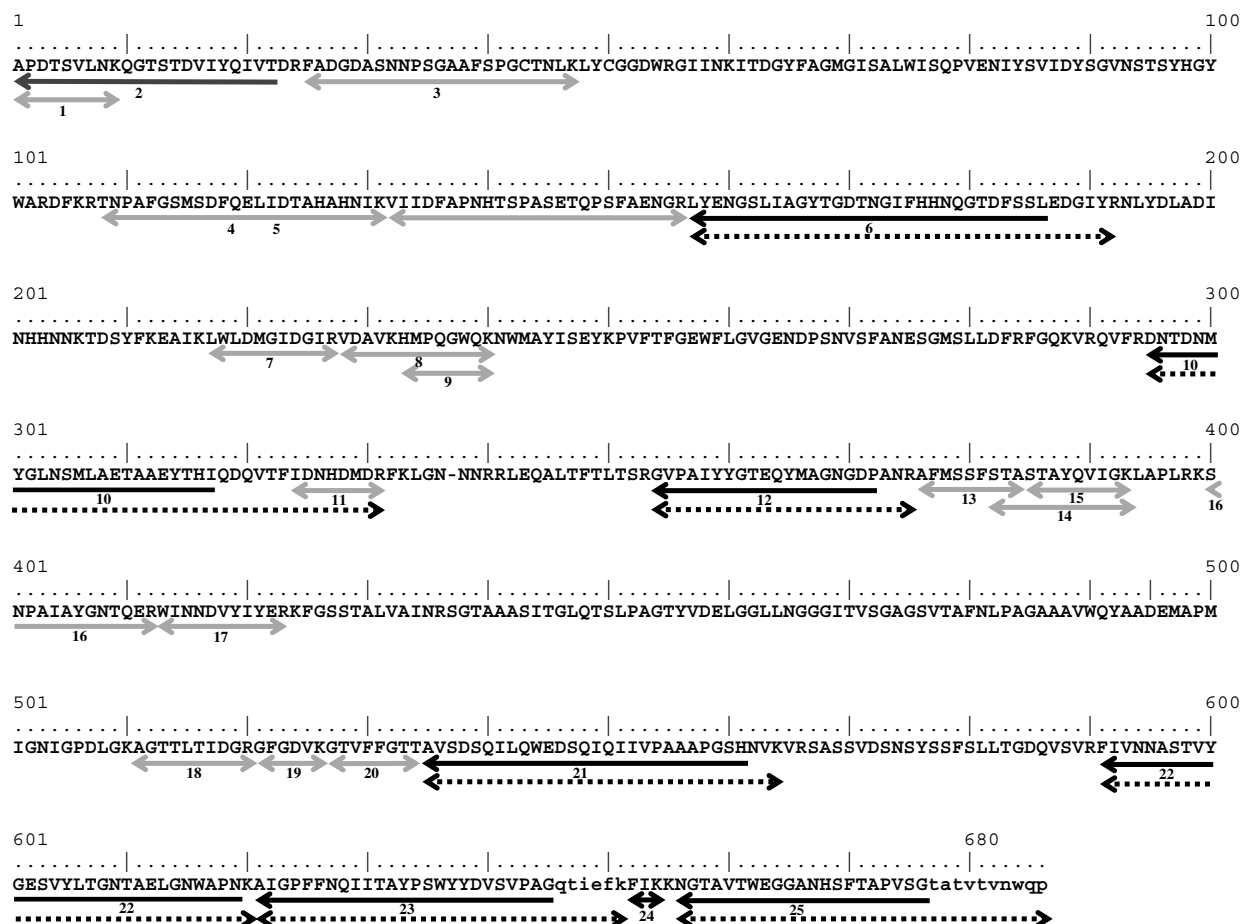
## 3. Experimental

### 3.1. Materials

Diethyl pyrocarbonate,  $\alpha$ -,  $\beta$ - and  $\gamma$ CD, 4-vinylpyridine were from Sigma Chemical Co., MO, USA.  $^{14}\text{C}$ CO DEP was from American Radiolabelled Chemicals, Inc., MO, USA. Sequencing grade trypsin and chymotrypsin were from Promega, Madison, WI, USA. All other chemicals were of AR grade. Reagents for cloning and sequencing—molecular biology grade—were from Invitrogen, CA, USA.

### 3.2. Enzyme assays

The amylolytic activity was determined by measuring the reduction in the intensity of the blue colour of the starch–iodine complex using 0.6% soluble potato starch as the substrate in 50 mM acetate buffer, pH 5.5, at 40 °C.<sup>19</sup>



**Figure 3.** *B. circulans* DF 9R CGTase amino acid sequence. Numbering corresponds to that of CGTase from *B. circulans* 251 and peptide numbers to those in Table 1. Results from DNA sequencing: N32–G645; in all cases they agree with those from Edman sequencing (peptides 1–25). D23, R24 and K655 were positioned by homology (searches for identity and similarity, BLAST) and enzyme specificity. Amino acids positioned by molecular mass determination and homology studies are in small letters. Black arrows indicate that the sequence was determined only by Edman degradation and open-ended ones denote that the sequence continues; grey arrows indicate Edman sequencing plus molecular mass determination and dotted arrows show that the sequence was determined by molecular mass and homology studies.

The CGTase  $\beta$ -cyclizing activity was determined by the method of Goel and Nene<sup>20</sup> employing 2% soluble potato starch as the substrate in 50 mM Tris–HCl buffer, pH 7.0, at 55 °C.  $\beta$ CD obtained was detected on account of its ability to form a stable colourless inclusion-complex with phenolphthalein. The decrease in absorbance at 550 nm is proportional to the  $\beta$ CD concentration.

The  $\alpha$ CD cyclizing activity was assayed by the decrease in absorbance at 507 nm caused by the formation of a methyl orange- $\alpha$ CD complex using as substrate 1% soluble potato starch in 50 mM phosphate buffer, pH 7.0, at 40 °C.<sup>21</sup>

### 3.3. SDS-PAGE

It was performed according to Laemmli.<sup>22</sup>

### 3.4. Protein reduction and alkylation

After protein reduction<sup>23</sup>, alkylation with 4-VP (1  $\mu$ L/100  $\mu$ g of protein) was performed under N<sub>2</sub> at room temperature, for 30 min, in the dark. Reagent excess was removed by RP-HPLC.

### 3.5. Histidine modification

It was carried out in the cell of a Perkin Elmer lambda 6 spectrophotometer at 25 °C. Seven  $\mu$ L of DEP in EtOH (final concentration  $2.1 \times 10^{-4}$  M) was added to 50  $\mu$ L enzyme soln in 100 mM phosphate buffer, pH 7.4 ( $6.3 \times 10^{-6}$  M final protein concentration).

The DEP/His molar ratio was 2. Absorbance at 242 nm was monitored.<sup>24</sup> The modification extent was calculated through the molar extinction coefficient of  $3.9 \times 10^3$  M<sup>-1</sup> cm<sup>-1</sup>. De-ethoxyformylation was achieved by adding a 2 M KOH solution to the EF-enzyme solution up to pH 8.4. When identification of the modified His/s was required, <sup>14</sup>C-DEP was used. Ethoxyformylation was also performed in the presence of CDs by incubating the enzyme (final concentration  $8.1 \times 10^{-6}$  M) in the presence of  $12.5 \times 10^{-3}$  M  $\alpha$ - or  $\gamma$ CD for 10 min in a 100 mM phosphate buffer, pH 7.4.

### 3.6. Protein conformation

- Circular dichroism measurements were performed in the 205–250 nm range at 25 °C in a Jasco J-810-150S spectropolarimeter with a cell provided with a circular window, pathlength 1 mm and calibrated with camphorsulfonic acid. Protein concentration was 0.47 mg/mL in 100 mM phosphate buffer, pH 7.4.
- Fourth-derivate spectra of native or EF-CGTase were determined at room temperature in a Perkin Elmer lambda 6 spectrophotometer with a slit width of 0.7 nm.

### 3.7. Tryptic digestion and HPLC peptide purification

The native or EF-enzyme (140  $\mu$ g) was digested with trypsin (enzyme/protein ratio 1:20) for 18 h at 37 °C in a 100 mM phosphate buffer, pH 7.4, with 3 M urea (final volume 300  $\mu$ L).



		170	180	190
EU644086	<i>B. circulans</i> DF 9R	AGYTGDT	NGIFHHNQGT	DFSSLEDGIY
P43379	<i>B. circulans</i> 251	GGYTNDT	QNLFFHHNGGT	DFSTTENGIIY
P30920	<i>B. circulans</i> 8	GGYTNDT	NGYFHHNGGS	DFSSLENGIIY
Q9F5W3	<i>B. circulans</i> A11	GGYTNDT	QNLFFHHYGGT	DFSTIENGIIY
AX356717	<i>B. agaradhaerens</i> DSM8721	GSYSDDS	DLFLY-NGGT	DFSNIYEDEIY
AX453023	<i>B. agaradhaerens</i> DSM9948	SSYSDNS	DLFLY-NGGT	DFSTYEDEIY
Q7X3T0	<i>B. agaradhaerens</i> LS-3C	GSYSDDS	DLFLY-NGGT	DFSNIYEDEIY
Q8L3E0	<i>B. clarkii</i> 7364	GHYSTDA	NNYFYNYGGS	DFSDEYENSIY
Ref <sup>25</sup>	<i>B. firmus</i> var <i>alkalophilus</i>	GNYSNDQ	QNLFQHNGGT	DFSSEYKSIY
P14014	<i>B. licheniformis</i>	GGYTNDT	NGYFHHNGGS	DFSTLENGIIY
P27036	<i>B. ohbensis</i>	GNYSNDP	NNLFHHNGGT	DFSSEYEDSIY
L25256	<i>B. sp</i> Q CK104	GGYTNDT	NGYFHHNGGS	DFSSLENGIIY
O82984	<i>B. sp</i> A25a	GNYSNDP	NNLFHHNGGT	DFSSEYEDSIY
P05618	<i>B. sp</i> 1011	GGYTNDT	QNLFFHHYGGT	DFSTIENGIIY
A3F9M7	<i>B. sp</i> BL-31	GSYSDDS	DLFLY-NGGT	DFSNIYEDEIY
A18991	<i>B. sp</i> 290-3	GHYSNDS	EDYFYTNNGS	DFSSEYEDSIY
P09121	<i>B. sp</i> 38-2	GGYTNDT	QNLFFHHYGGT	DFSTIENGIIY
P17692	<i>B. sp</i> B1018	GGYTNDT	QNLFFHHNGGT	DFSTTENGIIY
P30921	<i>B. sp</i> 17-1	GGYTNDT	HNLFFHHNGGT	DFSTTENGIIY
P31746	<i>B. sp</i> 1-1	GNYSNDQ	QNLFFHHNGGT	DFSSEYEDSIY
P31747	<i>B. sp</i> 6.6.3	GGYTNDT	NGYFHHNGGS	DFSSLENGIIY
Q5U9V9	<i>B. sp</i> G1-2004	GNYSNDQ	QNLFFHHNGGT	DFSSEYEDSIY
Q5U9W0	<i>B. sp</i> TS1-1	GNYSNDQ	QNLFFHHNGGT	DFSSEYEDSIY
Q6S3E3	<i>B. sp</i> I-5	GGYTNDT	QNLFFHHYGGT	DFSTIENGIIY
Q25CB6	<i>B. sp</i> G-825-6	GHYSNDN	EDYFYTNNGS	DFSSEYEDSIY
Q197W1	<i>B. sp</i> N-227	GGYTNDT	QNLFFHHYGGT	DFSTIENGIIY
Q59239	<i>B. sp</i> KC201	GNYSNDQ	QNLFFHHNGGT	DFSSEYEDSIY
Z34466	<i>B. sp</i> E-1	GNYSNDQ	QNLFFHHNGGT	DFSSEYEDSIY
Q5ZEQ7	<i>A. gottschalkii</i>	ASYSNDL	NEIFYHFHGGT	DFSTYEDSIY
O30565	<i>Br. brevis</i> CD162	GNYSNDR	NKLFHHNGGT	DFSSEYEDSIY
Q9ZAQ0	<i>G. stearothermophilus</i> ET1	GGYTNDT	NSYFHHNGGT	TFSNLEDGIY
P31797	<i>G. stearothermophilus</i> No2	GGYTNDT	NMYFHHNGGT	TFSNLEDGIY
P31835	<i>P. macerans</i>	GKYSNDT	AGLFHHNGGT	DFSTTESGIY
O52766	<i>P. macerans</i> IB7	GAYSNDT	AGLFHHNGGT	DFSTIEDGIY
P04830	<i>P. macerans</i> IAM1243	GAYSNDT	AGLFHHNGGT	DFSTIEDGIY
P26827	<i>T. thermosulfurigenes</i> DSM3896	GGYTNDT	NGYFHHYGGT	DFSSEYEDGIY
Ref <sup>26</sup>	<i>Thermoanaerobacter</i> sp 501	GGYTNDT	NGYFHHYGGT	NFSSYEDGIY
Z35484	<i>Thermoanaerobacter</i> ATCC53627	GGYTNDT	NGYFHHYGGT	NFSSYEDGIY
	Clustal Consensus	. *: :	: *	** *. ** :***:***

**Figure 4.** Alignment of the –6 subsite (black bars) of 38 CGTases. Accession numbers and producer microorganisms are indicated. B.: *Bacillus*; A.: *Anaerobranca*; Br.: *Brevibacillus*; G.: *Geobacillus*; P.: *Paenibacillus*; T.: *Thermoanaerobacterium*. (See above-mentioned references for further information.)

For peptide purification, an AKTA Purifier (Amersham Biosciences), with a C18 Vydac 218TP54 column was used. Elution was performed with a 2–45% acetonitrile (ACN) gradient in 50 min followed by a 45–55% ACN gradient in 6 min and a 55–100% ACN gradient in 5 min, all of them in 0.1% trifluoroacetic acid. Flow rate was 0.5 mL/min, elution was monitored at 215 nm and by radioactivity. For the latter, a 10 vol % of the collected HPLC tryptic fractions was diluted in 2 mL scintillation soln, and radioactivity was measured on a 1214 Rackbeta liquid scintillation counter equipped with the GENTERM V2,B software.

All tryptic fractions were submitted to mass spectrometry; those containing only one molecular species were then sequenced, while simple mixtures were first rechromatographed by changing the column and/or the elution gradient; new fractions were also analyzed by ESIMS.

### 3.8. Liquid chromatography–mass spectrometry

Either the native protein or the tryptic fractions were loaded onto a Brownlee C8 or a Brownlee C18, respectively (1 × 30 mm) column and gradient eluted into a Thermo Finnigan LCQ DUO-trap mass spectrometer (Thermo\_Quest, San José, CA, USA). Positive ions were analyzed using electrospray ionization and MS data acquired in the full-scan mode. The ESI sheath gas was operated at 40 units, the heated capillary temperature at 200 °C and the spray voltage at 4.5 kV. An XCALIBUR 1.3 software was utilized.

### 3.9. Chymotryptic digestion

An enzyme–peptide ratio 1:20 was used; digestion was performed for 15 h at 30 °C in 100 mM ammonium hydrogencarbonate, pH 8.0 (final vol 100 µL). Fractions molecular mass was determined as described above.

### 3.10. Peptide sequencing

It was performed at the *Facility for Peptide and Protein Sequencing* (Universidad de Buenos Aires). A Procise 492 Applied Biosystems sequencer was used.

### 3.11. Amplification of the CGTase gene

DNA was extracted and purified from *B. circulans* DF 9R pellets. Bacteria were lysed by lysozyme digestion (10 mg/mL) at 37 °C for 60 min followed by treatment with proteinase K (2 mg/mL) and SDS 0.6% overnight at 56 °C. Nucleic acids were extracted with 25:24:1 phenol–CHCl<sub>3</sub>–isoamyl alcohol, and the aqueous phase was incubated with RNase A (10 µg/µL) for 30 min at 37 °C, followed by CHCl<sub>3</sub> extraction. DNA was precipitated from the aq phase with 0.1 vol sodium acetate 3 M, pH 5.5 and 1 vol isopropanol. The resulting chromosomal DNA was washed with 70° EtOH and dissolved in nuclease free water.

The CGTase gene was partially amplified by PCR using 1 (5' GGACGGCAATCCSKCAACAA 3') and 12com (5' GCCTGCCGWA

CGCTGAC ATCATA 3') as primers. The 50 µL final reaction mixture contained 50 mM KCl; 20 mM Tris-HCl pH 8.4; 1.5 mM MgCl<sub>2</sub>; 0.2 mM each dNTP; 0.2 µM primers and 1 U Taq DNA polymerase. Amplification conditions: 2 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 30 s at 50 °C and 2 min at 72 °C. Amplification products were evidenced on 1% agarose gels.

### 3.12. Cloning and sequencing

Amplification products were purified by a Promega column and cloned into a pCR2.1-TOPO Cloning vector with the TOPO TA cloning kit (Invitrogen). *Escherichia coli* DH5α transformants were selected on LB agar plates with 50 µg/mL kanamycin and 50 µg/mL X-Gal. The recombinant plasmids were isolated by the alkaline-lysis method using a Promega kit, digested with *Eco*RI and visualized on agarose gels to confirm the presence of an inserted fragment. DNA fragment was sequenced by the dideoxynucleotide chain-termination method in a Applied Biosystems 3130xl Genetic Analyzer. Sequencing primers: M13 Reverse (5' CAGGAAACAGC TATGAC 3'), M13 (-21) Forward (5' TGTAAAACG ACGGCCAGT 3'), int1 (5' TCGACGGCATCCGGGTAGAT 3'), int1com (5' ATCTACC CGGATG- CCGTCGA 3'), int2 (5' TTCGGCAGCAGCACAGCGCT 3') and int2com (5' AGCGCTGTGCTGCTGCCGAA 3').

### 3.13. Sequence analysis

Nucleotide and amino acid sequences were edited using BIOEDIT software version 7.0.4.1 and multiple sequence alignments were obtained with the CLUSTALX program.

### Acknowledgements

We thank Dr. J. Delfino and Biochemists L. Curto and G. Gómez for performing circular dichroism; Drs. O. Taboga and A. Distefano for their supervision on cloning and sequencing; Dr. R. Rios for his kind supply of materials; S. Linskens, E. Dacci and C. Paván for their technical assistance; Professor R. Davis for language supervision.

The work was supported by grants from the Universidad Nacional de Luján, the Universidad de Buenos Aires and CONICET.

### References

1. Carbohydrate-Active enZyme (CAZy) database at <http://www.cazy.org/CAZY>.
2. Biwer, A.; Antranikian, G.; Heinzel, E. *Appl. Microbiol. Biotechnol.* **2002**, *59*, 609–617.
3. Plou, F. J.; Martín, M. T.; Gómez de Segura, A.; Alcalde, M.; Ballesteros, A. *Can. J. Chem.* **2002**, *80*, 743–752.
4. Qi, Q.; Zimmermann, W. *Appl. Microbiol. Biotechnol.* **2005**, *66*, 475–485.
5. Thiemann, V.; Dönges, C.; Prowe, S. G.; Sterner, R.; Antranikian, G. *Arch. Microbiol.* **2004**, *182*, 226–235.
6. van der Veen, B. A.; Uitdehaag, J. C. M.; Dijkstra, B. W.; Dijkhuizen, L. *Biochim. Biophys. Acta* **2000**, *1543*, 336–360.
7. Janeček, Š. *Biol. Bratislava* **2002**, *57*, 29–41.
8. Nakamura, A.; Haga, K.; Yamane, K. *Biochemistry* **1993**, *32*, 6624–6631.
9. Ferrarotti, S. A.; Rosso, A. M.; Maréchal, M. A.; Krymkiewicz, N.; Maréchal, L. R. *Cell. Mol. Biol.* **1996**, *42*, 653–657.
10. Szyman, N.; Schroh, I.; Rossi, A. L.; Rosso, A. M.; Krymkiewicz, N.; Ferrarotti, S. A. *Bioresour. Technol.* **2007**, *98*, 2886–2891.
11. Maréchal, L. R.; Rosso, A. M.; Maréchal, M. A.; Krymkiewicz, N.; Ferrarotti, S. A. *Cell. Mol. Biol.* **1996**, *42*, 659–664.
12. Lacorazza, H. D.; Otero de Bengtsson, M.; Biscoglio de Jiménez Bonino, M. J. *Neurochem. Int.* **1996**, *28*, 77–97.
13. Kaulpiboon, J.; Pongsawasdi, P. *J. Biochem. Mol. Biol.* **2003**, *36*, 409–416.
14. Mattsson, P.; Battchikova, N.; Sippola, K.; Korpela, T. *Biochim. Biophys. Acta* **1995**, *1247*, 97–103.
15. Ishii, N.; Haga, K.; Yamane, K.; Harata, K. *J. Mol. Recognit.* **2000**, *13*, 35–43.
16. Leemhuis, H.; Uitdehaag, J. C. M.; Rozeboom, H. J.; Dijkstra, B. W.; Dijkhuizen, L. *J. Biol. Chem.* **2002**, *277*, 1113–1119.
17. van der Veen, B. A.; Uitdehaag, J. C. M.; Dijkstra, B. W.; Dijkhuizen, L. *Eur. J. Biochem.* **2000**, *267*, 3432–3441.
18. Kim, Y. H.; Bae, K. H.; Kim, T. J.; Park, K. H.; Lee, S. M. *Biochem. Mol. Biol. Int.* **1997**, *41*, 227–234.
19. Rosso, A. M.; Ferrarotti, S. A.; Krymkiewicz, N.; Nudel, B. C. *Microb. Cell Factories* **2002**, *1*, 1–10.
20. Goel, A.; Nene, S. *Starch* **1995**, *47*, 399–400.
21. Lejeune, A.; Sakaguchi, K.; Imanaka, T. *Anal. Biochem.* **1989**, *181*, 6–11.
22. Laemmli, U. K. *Nature* **1970**, *227*, 680–687.
23. Matsudaira, P. T. In *A Practical Guide to Peptide and Protein Purification for Microsequencing*; Matsudaira, P. T., Ed.; Academic Press: New York, 1989; pp 17–30.
24. Choong, Y. S.; Shepherd, M. G.; Sullivan, P. A. *Biochem. J.* **1977**, *165*, 385–393.
25. Park, T. H.; Shin, H. D.; Lee, Y. H. *J. Microbiol. Biotechnol.* **1999**, *9*, 811–819.
26. Norman, B. E.; Jorgensen, S. T. *Denpun Kagaku* **1992**, *39*, 101–108.