

Computational analysis of 1,3-propanediol operon transcriptional regulators: insights into *Clostridium* sp. glycerol metabolism regulation

Carlos Eduardo Barragán¹✉, Andrés Julián Gutiérrez-Escobar², Dolly Montoya Castaño¹

Abstract

We designed a strategy for the sequencing and bioinformatical characterization of the 1,3-propanediol operon regulator genes from the Colombian *Clostridium* sp. strain IBUN13A, which is taxonomically related to *Clostridium butyricum*. Three genes are proposed to be involved in the operon's transcriptional activity, the *dbaS* and *dbaA* genes through a two-component system and the third gene named *dbaY*, which encodes a putative transcriptional regulator similar to the domains of the *dbaS/A* system. Phylogenetic analyses indicated that the predicted proteins had a modular structure consisting of domains homologous to different signal transduction systems, but had significant differences concerning their conserved residues, pointing to the possibility that they constitute ancestral domains. In accordance with the prediction of functions, we propose a mechanism of regulation of the proteins studied of the 1,3-propanediol operon of the native strain, as a response to the presence of glycerol in the medium, which provides valuable information on the overall regulation of the glycerol metabolism in *Clostridium* sp.

Keywords: *Clostridium* sp.; glycerol metabolism; 1,3-propanediol; regulator genes; two-component system

Edited by Alberto Acosta ✉

1. Grupo de Bioprocesos y Bioprospección, Instituto de Biotecnología, Universidad Nacional de Colombia, Bogotá, Colombia

2 Grupo de Investigaciones Biomédicas y de Genética Humana Aplicada, Universidad de Ciencias Aplicadas y Ambientales - UDCA, Bogotá, Colombia

Received: 16-06-2014 **Accepted:** 10-07-2014

Published on line: 26-09-2014

Citation: Barragán CE, Gutiérrez-Escobar AJ, Montoya-Castaño D (2015) Computational analysis of 1,3-propanediol operon transcriptional regulators: Insights into *Clostridium* sp. glycerol metabolism regulation. *Universitas Scientiarum* 20(1): 129-140 doi: 10.11144/Javeriana.SC20-1.capo

Funding: COLCIENCIAS; Universidad Nacional de Colombia.

Electronic supplementary material: Suppl 1,2,3.

Introduction

Biodiesel is a fatty acid methyl-ester obtained from renewable sources such as oil palm, soybean and sunflower seeds; it is one of the most promising alternatives to fossil fuels as it can be used in diesel motors and other combustion systems (Papanikolaou et al. 2008). Crude glycerine is the main subproduct produced during biodiesel synthesis; it can be used in anaerobic fermentation as a carbon source to produce 1,3-propanediol (1,3-PD) (Barbিরato et al. 1998, Paulo da Silva et al. 2009, Ayoub & Abdullah 2012). This 1,3-PD is used in the textile industry, in the production of adhesives, lubricants, solvents, resins, antifreeze and pharmaceutical products (Zeng & Biebl 2002) and in the synthesis of polyesters, polyurethanes



and especially polytrimethylene terephthalate (PTT), which is more elastic and has better tension recovery properties than other polymers (Nakamura & Whited 2003, Saxena et al. 2009).

Additionally, 1,3-PD is produced by chemical synthesis using two different courses: acrolein hydration or ethylene oxide hydroformylation (Paulo da Silva et al. 2009). However, for the last few years, research efforts have been focused on its biological production as it is a more environmentally, friendly alternative (Saxena et al. 2009, Yao & Shimizu 2013). Biotechnological 1,3-PD production from glycerol as the sole carbon source has been characterized in *Clostridium butyricum*, *Klebsiella pneumoniae* and *Cytrobacter freundii* (Barbirato et al. 1998, Saxena et al. 2009, Kubiak et al. 2012). *C. butyricum* is considered one of the best producers and is characterized because it does not require coenzyme B₁₂ to be added during fermentation, making this species an ideal biological model for diol research and production (Papanikolaou et al. 2000, González-Pajuelo et al. 2005).

The Colombian strain *Clostridium* sp. IBUN 13A is closely related to *C. butyricum* (Jaimes et al. 2006, Montoya et al. 2001), and it has shown similar 1,3-PD yields to those obtained by reference strains such as *C. butyricum* DSM 523 and *C. butyricum* DSM 2478 (Cárdenas et al. 2006). The genes involved in 1,3-PD synthesis have been molecularly characterized recently, and the structure of the 1,3-PD operon for this strain has been determined (GenBank accession code DQ901408) (Montoya 2008, Quilaguy et al. 2010).

Due to the similarity with the *C. butyricum* VPI1718 operon, it has been postulated that the regulator genes of the 1,3-PD operon of *Clostridium* sp. strain IBUN 13A are located upstream of the *dbaB1* gene as described for *C. butyricum* strain VPI1718 (Raynaud et al. 2003). The *dbaS* and *dbaA* genes are proposed to be involved in the *C. butyricum* operon transcriptional regulation by means of a two- component signal transduction system whose regulation could have a mechanism similar to the *dbaR* protein found in other 1,3-PD producing microorganisms (Sun et al. 2003).

This study referred to the *in silico* prediction of the modular structure of regulator proteins and their

phylogenetic history from *dbaS* and *dbaA* sequences, and to a putative transcriptional regulator gene identified upstream of the 1,3-PD operon genes of *Clostridium* sp. strain IBUN 13A, called *dbaY* in this research. Consequently, possible transcriptional regulation mechanisms could be proposed for a signal transduction system on the 1,3-PD operon.

Materials & Methods

Sequencing and sequence assembling

Clostridium sp. strain IBUN 13A, obtained from the strain bank of the Institute of Biotechnology of the Universidad Nacional de Colombia (Montoya et al. 2000), and *C. butyricum* strain, DSM 2478 (positive control), were used in this study. The strains were activated following the methodology described by Montoya et al. (2000). The microorganisms were grown in strict anaerobic conditions at 37 °C in modified TGY medium (16g/L tryptone, 5 g/L glucose, 5g/L yeast extract, 5g/L NaCl, and 0.5g/L L-cysteine) supplemented with 0.05 mg/mL resazurin for chromosomal DNA extraction. Chromosomal DNA was extracted using the methodology described by Jaimes et al. (2006).

GenBank sequences of *C. butyricum* VPI 1718 (accession codes AY112989.1 and AY138581) were used for designing the primers (Suppl. 1) due to the similarity found between the 1,3-PD operon of *Clostridium* sp. strain IBUN 13A and *C. butyricum* VPI1718 (Montoya, 2008).

Amplification reactions were done in 100 µL tubes using 25 µL final volumes at the following concentrations: 0.25 µM primers (synthesized by Integrated DNA Technologies, Inc), 0.07U/µL Taq polymerase (GoTaq, Promega), 2 mM MgCl₂, 200 µM dNTPs, 1X buffer and 50 ng chromosomal DNA, and a multigene thermocycler was used (Labnet International, Inc). Amplification products were confirmed using conventional electrophoresis on 1.5% agarose gels with 0.5 X TBE.

Amplified fragments were sent for purification and sequencing to MacroGen Inc. Sequence quality was assessed using PHPH software (Togawa & Brigido 2003) [<http://asparagin.cenargen.embrapa.br/phph/>] and assembled using Cap3 software.

Gene annotation

GeneMark.hmm PROKARYOTIC software (Besemer & Borodovsky 2005) [http://opal.biology.gatech.edu/GeneMark/genemark_prok_gms_plus.cgi] was used to find open reading frames (ORF) and ribosome binding sites (RBS). *Clostridium acetobutylicum* was the model organism selected to make predictions. BPROM [<http://linux1.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup>] software was used to predict -10 and -35 boxes from the promoter region.

The presence of rho-independent terminators was determined using the EMBOSS Palindrome application, and the parameters defined by Lesnik et al. (2001) for this kind of transcriptional terminator. The sequence predicted by Palindrome software was evaluated by OligoAnalyzer 3.0 to find its respective ΔG value.

Sequence analysis

Using the PSI-BLAST tool from the prediction made by using GeneMark, we performed a search for homologous protein sequences in the Swiss-Prot database; the search was restricted to bacterial taxa and 10^{-5} E- value. The process was iterated until convergence was obtained. A search was then made in UniProt with the domains identified in SMART [<http://smart.embl-heidelberg.de/>] and SCAN-PROSITE [<http://prosite.expasy.org/scanprosite/>] (Sigrist et al. 2013) for each protein, obtaining 100% cluster groupings.

The partial and redundant sequences obtained were debugged until we obtained a population of 35 protein clusters homologous with a domain present in the *dhaS* or *dhaY* proteins. Similarly, we obtained a population of 44 protein clusters that were homologous with a domain present in the *dhaA* and *dhaY* proteins (Suppl. 2). These sequences were classified according to the superfamily to which they belonged; *C. butyricum* AAM54726.1 and AAM54727.1 sequences were used as a reference.

Secondary structure was predicted using PSIPRED software [<http://bioinf.cs.ucl.ac.uk/psipred/>]. PSORTB 5.0 [<http://www.psort.org/psortb/index.html>] software was used to determine possible *dhaS*,

dhaA and *dhaY* protein cellular localization (Yu et al. 2010), and obtaining prediction for Gram-positive eubacteria proteins. TMpred [http://www.ch.embnet.org/software/TMPRED_form.html] and TopPred [<http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred>] software were used to find hypothetical transmembrane regions. T-Coffee software was used for the multiple alignments, using default parameters and manual editing.

Phylogenetic analysis

Three groups of alignments were selected, representing the three protein domains determined from the Uniprot cluster analysis. The protein sequence alignments obtained using T-Coffee software were used to determine the conserved residues traced by Evolutionary Trace Server software [<http://mordred.bioc.cam.ac.uk/~jiye/evoltrace/evoltrace.html>]. They were then manually edited in Jalview and analyzed in ProtTest software (Abascal et al. 2005) [http://darwin.uvigo.es/software/prottest_server.html] to identify which evolutionary model explained each alignment.

Phyml software [http://www.phylogeny.fr/version2/cgi/one_task.cgi?task_type=phyml] was then used for phylogenetic reconstruction using 1,000 bootstrap repeats as statistical support. Njplot software was used to visualize phylogenetic trees, and then, Tajima's test was used to determine which selection worked on the three groups of domains using MEGA 5.05 software (Kumar et al., 2008).

Results

Our sequencing strategy amplified two fragments of 2633 bp and 1590 bp, upstream of the 1,3-PD operon of *Clostridium* sp. IBUN 13A. The use of PHPH software verified the high quality of the sequences obtained, meaning that after the data depuration, 2,565 bp and 1,508 bp contigs, respectively, could be assembled without having any dissimilarity. Two ORFs were found in the assembled 2,565 bp sequence, called *dhaS* and *dhaA* according to Raynaud et al. (2003) 's description in *C. butyricum* and an ORF was found in the assembled 1,508 bp sequence.

The first ORF identified as *dbaS* had 1,218 bp. Potential -35 (5'-TTCATA-3') and -10 (5'-TAAAT-3') promoter sites were predicted to be upstream of the start codon. The GeneMark software did not produce a predicted RBS; this is why the TGGTGA sequence was proposed as a putative site in this work; however, this site has not been predicted in other *Clostridium* species. No sequence was identified for this gene that could form a typical rho-independent transcriptional terminator.

The *dbaA* gene start codon was located 9 bp from the *dbaS* gene stop codon and consisted of 1,056 bp. The software did not identify RBS; however, the AGGAAG sequence located 5 bp upstream from the start for this gene, is proposed as its RBS. We identified a possible hairpin (characteristic of a rho-independent transcriptional terminator). The obtained sequence was annotated in GenBank as an update of the 1,3-PD operon sequence of *Clostridium* sp. IBUN 13A (accession code DQ 901408.6).

Upstream of the *dbaS* gene, a third 1,059 bp ORF was found in the assembled 1,508 bp region that encodes a putative transcriptional regulator. It was named the *dbaY* gene because the regulator protein encoded by this ORF is not homologous to the DhaR protein described in *K. pneumoniae* and *C. freundii* (Sun et al. 2003). Potential -35 (5'-TTCATA-3') and -10 (5'-TTTAT-3') promoter sites were located upstream of the start codon. GeneMark software did not predict an RBS; therefore, GGAATA sequence has been proposed as being a putative site in this work. However, this site has not been predicted in other *Clostridium* species. We identified no sequence for this gene, which could form a typical rho-independent transcriptional terminator. The sequence obtained was noted in the GenBank as a new *Clostridium* sp. IBUN 13A sequence (accession code JF827037).

The *dbaY* gene encodes a 353 amino acid residue-long protein (*DbaY* GenBank accession code AEH42432), the *dbaS* gene encodes a 405 amino acid residue-long protein (*DbaS* GenBank accession code ADP02232.1), and the *dbaA* gene encodes a 351 amino acid residue-long protein (*DbaA* GenBank accession code ACT78697.2). The cytoplasmic locations for the *dbaY*, *dbaS* and *dbaA* proteins were inferred in all applications.

The search for proteins homologous to the predicted *dbaY*, *dbaS* and *dbaA* proteins revealed the existence of domains belonging to two-component signal transduction systems with a modular design (**Figure 1**). Because we obtained no functional information from experimental data, proteins were annotated as suggested by Galperin (2006), based on their domain composition rather than any given group of protein sequence similarities.

Histidine kinase sensor (N-terminal region, first 179 residues), histidine kinase (amino acids 205 to 287), and HATPase domains (C-terminal region, last 110 residues) have been identified for the *dbaS* protein, thereby corroborating that this protein is a histidine protein kinase. PSI-BLAST software revealed a 55% similarity of the N-terminal region with the PocR superfamily. The predicted secondary structure did not show any pattern to classify this sensor as bi-functional or mono-functional according to criteria of Alves & Savageau (2003).

The response regulatory domain -REC- (N-terminal region, first 118 residues) and HTH_AraC DNA binding domain (C-terminal region, amino acids 247 to 345) were identified for the *dbaA* protein (common in a two-component response regulator protein). The *dbaY* protein had hybrid *dbaS* and *dbaA* protein modular organization, and the PocR sensor domain (N-terminal region) as HTH_AraC DNA binding domain (C-terminal region) were identified.

Because of the heterogeneity of the superfamilies identified (more than 70,000 reported sequences), UniProt cluster analysis was used to debug the alignments until protein sequences of the same pattern of motifs as the *dbaY*, *dbaS* and *dbaA* proteins were obtained. The 34 sequences homologous to the *dbaY* or *dbaS* proteins were classified within the PocR and histidine kinase superfamilies, and the 44 sequences homologous to the *dbaY* and *dbaA* proteins were classified within the REC superfamily and the HTH_AraC superfamily (**Suppl. 2**). The groups of sequences for each superfamily have been referred to as Hiskin, REC and HTH to facilitate the analysis.

The final population of sequences studied was the following: 15 consensus sequences for the histidine kinases, REC/25 consensus sequences, and

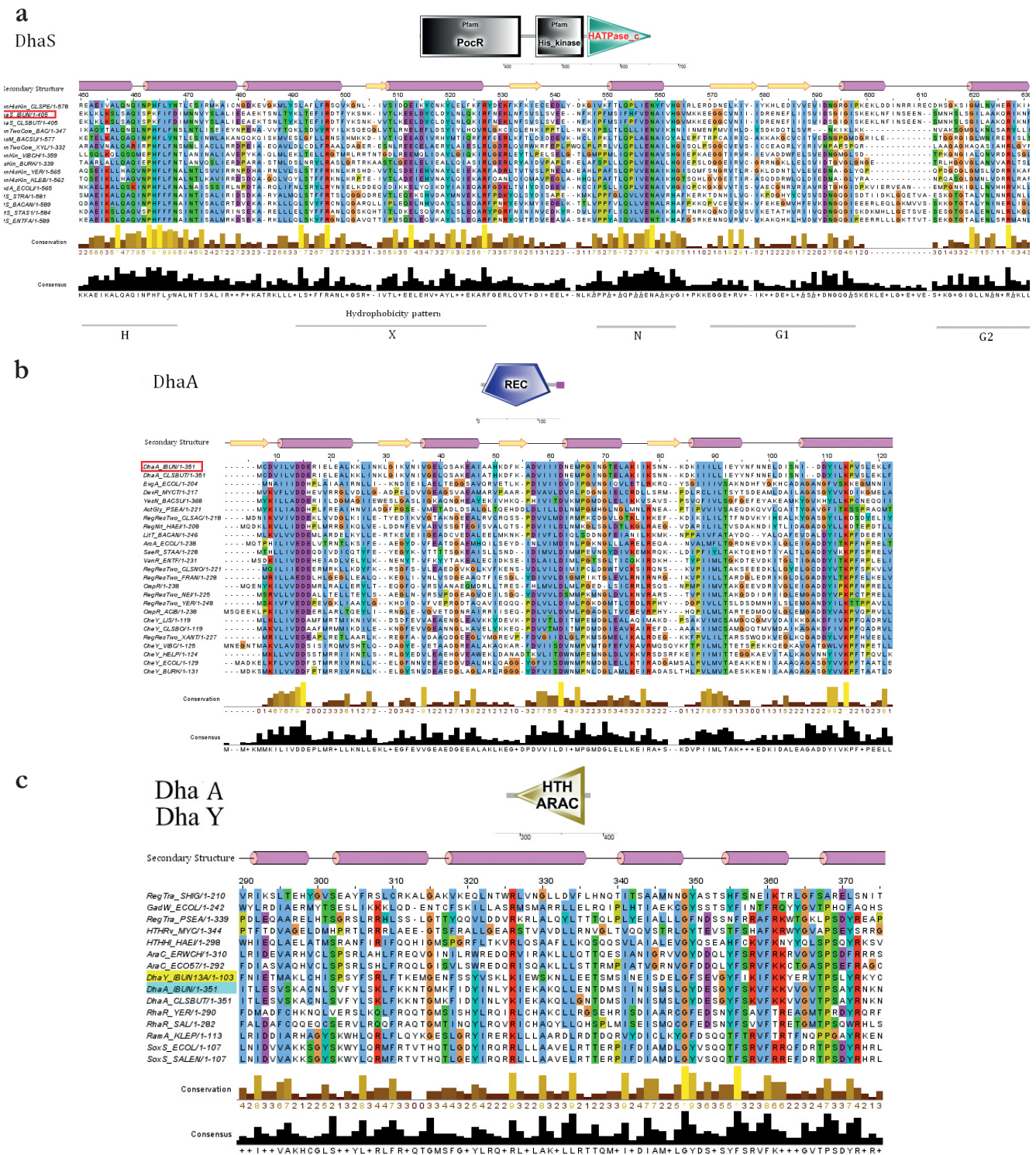


Fig. 1. Sequence analysis. Multiple sequence alignment. At the top, domain found in the SMART application and the secondary structure predicted by Psi-Pred. In the middle, multiple cluster alignment developed with Jalview (sequences are noted according to the name found in Suppl. 2). At the bottom are the conserved amino acids and also the consensus sequences (a) dhaS protein. The H-X-N-G1-G2 motifs found are shown below the consensus sequence. The nomenclature defined by Grebe & Stock (1999) has been used in the consensus sequence, hydrophobic residues 'h' (ALICVMYF) and 'y' means any structurally similar residue: H, F or Y. (b) The dhaA protein REC domain. The residues identified in the consensus sequence are those described by Stock & Da Re (2000). (c) The dhaA and dhaY protein HTH domain. Residues identified in the consensus sequence are those described for araC family / XylS transcriptional regulators by Gallegos et al. (1997).

Table 1. Natural Selection Test for HTH, REC and Hiskin clusters. MEGA5.10 software was used to estimate Tajima's test. All positions with gaps and missing data were eliminated from the dataset (complete deletion option). The abbreviations used were as follows: m = number of sites, S = number of segregating sites, Ps = S/m, T = Ps /a1 and Θ = nucleotide diversity. D is Tajima's test score.

Cluster	m	S	Ps	T	Θ	D
HTH	14	101	0.980583	0.308346	0.732956	6.144532
REC	26	106	0.972477	0.254845	0.721581	7.174758
HISKIN	15	246	0.964706	0.296690	0.731877	6.498782

Sequences having the PocR domain were not multiply aligned due to the reduced number of homologous sequences found. However, *dbaS* Hiskin sensor and *dbaY* sensor domain similarity with the PocR protein was demonstrated by alignment; this lead to the identification of three highly conserved cysteine residues (Suppl. 3). The previous suggests that they likely make up the interior of the surface contacting the ligand (Anantharaman & Aravind 2005).

The HisKin *dbaS* catalytic domain (corresponding to a class I histidine protein kinase) is grouped within the group 8 HisKin subfamily; this according to the classification by Grebe & Stock (1999). Motifs common by these kinases were identified in the alignment (Figure 1-a).

According to ProfTest software, the WG and LG evolutionary models explained cluster alignments (Table 1). The three phylogenetic trees showed base nodes having significant bootstrap values (77% HTH, 56% REC and 100% Hiskin) and topology that was correlated to the structural analysis mentioned below (Figure 2).

Discussion

The first characterization study of genes involved in the anaerobic metabolism of glycerol revealed that the *dba* regulon organization is present in different microorganisms capable of using glycerol as a carbon source. Although heterogeneity has been shown in the organization of these genes, even within organisms belonging to the same genus, a global system of transcriptional regulation controlled by response regulator proteins has been proposed (Sun et al. 2003). The distinctive characteristic of Clostridia is their two-component system regulator organization; this is demonstrated in the *C. butyricum* 1,3-PD operon characterization (Raynaud et al. 2003, 2011). Characterizing the 1,3-PD operon in the Colombian strain *Clostridium* sp. IBUN 13A has indicated an organization similar to that of *C. butyricum* (Montoya 2008, Quilaguy et al. 2010).

It is likely that the genes sequenced in this study (called *dbaS* and *dbaA*) comprise an operon, due to the promoter region identified upstream of the first gene, the reduced distance separating them in the same DNA chain (only 9 bp), and the single rho-factor-independent transcriptional terminator downstream of both genes. This hypothesis must be confirmed by primer extension or 5'-RACE, which could identify a two-component system characterized by its structural genes organized in operons (Mascher et al. 2006). A two-component system may regulate the glycerol metabolism in the Colombian strain, inversely to *K. pneumoniae* or *C. freundii* (Sun et al. 2003). Clostridia genomes have shown that two-component signal transduction protein-encoding regions frequently occur (Cheung et al. 2005).

Because *C. butyricum* is the closest species to the Colombian strain *Clostridium* sp. IBUN 13A (Montoya et al. 2001, Jaimes et al. 2006), the similarity between their 1,3-PD operon regulator genes could suggest an ancestral event involving horizontal transfer of this genomic region. It has been proposed that two-component transcriptional regulation systems become diversified in prokaryotic organisms because of gene duplication (Hoch 2000).

The domains found in the deduced proteins exhibit the versatile modular organization pattern

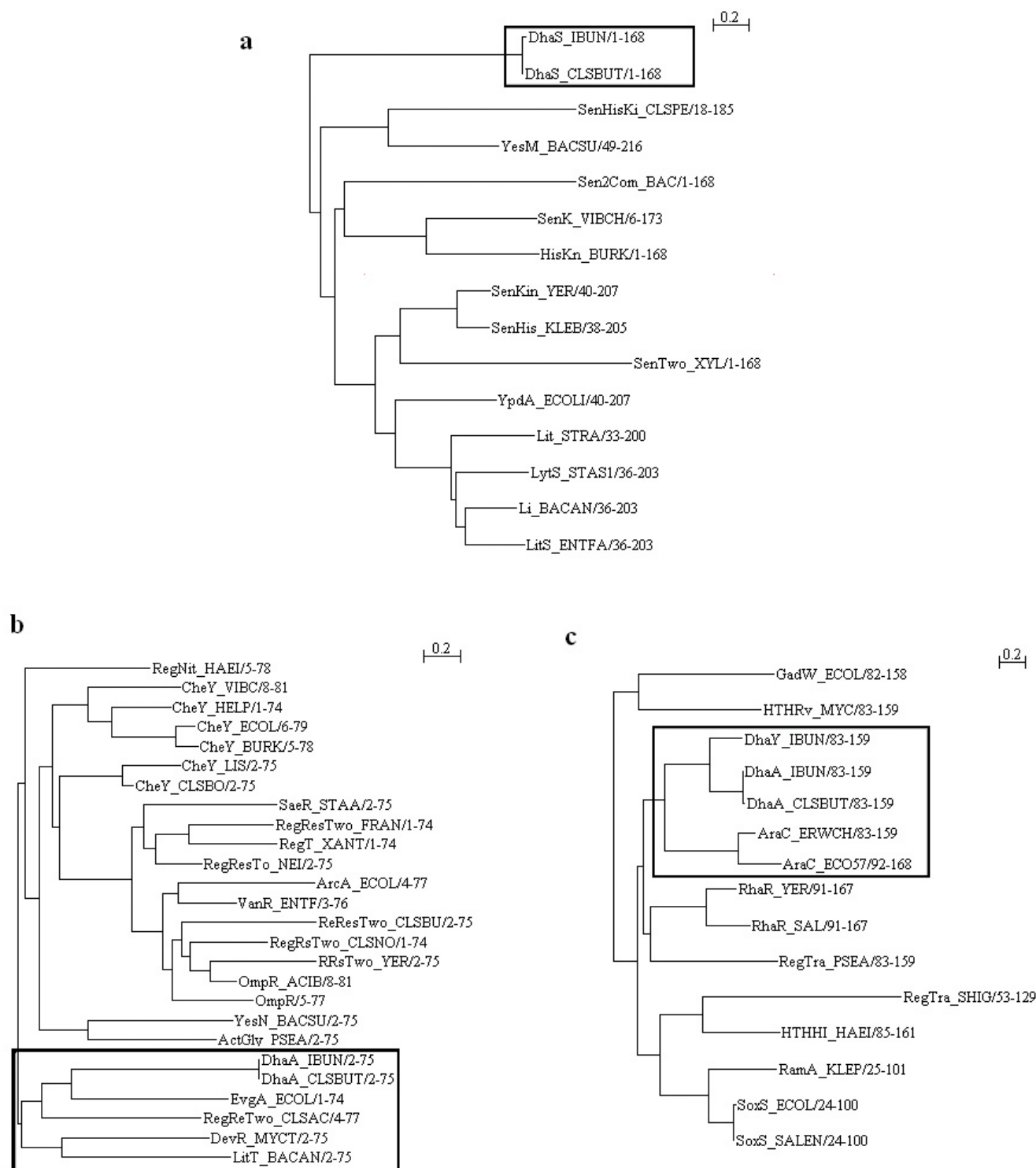


Fig. 2. Phylogenetic analysis. The phylogenetic tree was constructed using Phym software and the evolutionary model according to the results obtained by ProtTest, sequence groups were aligned by T-Coffee. Phylogenetic reconstruction required 1,000 bootstrap repetitions for statistical robustness. Tajima's test was used to determine which type of selection operated on three groups of domains using MEGA 5.05. (a) HisKin cluster sequences using 13 consensus sequences representing 125 individual sequences. The box identifies the ancestral clade to which the Colombian strain belongs. (b) REC cluster using 24 consensus sequences representing 566 individual sequences. The box identifies the ancestral clade to which the Colombian strain belongs. (c) HTH cluster using 12 consensus sequences representing 191 individual sequences. The Colombian strain's HTH domains are shown in the box, grouped with araC proteins.

characteristic of signal transduction proteins. This property confers the organism that harbors them great adaptability (West & Stock 2001, Jung et al. 2012). Our phylogenetic analysis supported this evolutionary trend. Our data suggested a functional purifying selection model (Table 1), following the Lego principle; most domains could be easily recognized and associated with particular biochemical functions (Galperin 2004, 2006). There is currently no generalized classification for all response regulator proteins, because they vary widely in their sequence, membrane topology, composition, and domain arrangement (Mascher et al. 2006).

According to the predictions, the PocR domain binds to intracellular ligands, thereby inducing a conformational change that is then transmitted to a catalytic signal transduction domain, such as the HisKin domain (Anantharaman & Aravind 2005). It has been experimentally shown that PocR protein activity is regulated *in vitro* by 1,2-propanediol molecules (Rondon & Escalante 1996). In turn, this has led to the proposal that PocR should recognize simple hydrocarbon derivatives such as 1,2-propanediol or acetate, and, therefore, plays a role in the detection of the substrates required for microbial growth (Anantharaman & Aravind 2005).

Due to its importance in 1,3-PD operon transcriptional regulation, it is probable that *dbaY* and *dbaS* sense the presence of glycerol (ligand) as a substrate in the cytoplasm, and can activate the promoter transcription because of a two-component system of regulation of nutrient uptake and its metabolism that frequently occurs (Tetsch & Jung 2009). However, little information is available on complementary regulation between a two-component system, and a transcriptional regulator made up of homologous domains (Townsend et al. 2013), which may be the case with a *dbaY* protein and *dbaS*/A system. Future studies are required to experimentally determine whether different ligands such as glycerol, 1,2-propanediol or even 1,3-propanediol can generate a conformational change in the HisKin *dbaS* and *dbaY* sensor domains of the Colombian strain.

Evidence shows that the sensors could be coupled to transport proteins acting as co-sensors and be responsible for substrate translocation through the

membrane (Tetsch & Jung 2009, Kobir et al. 2011). For that reason, further experimental assays must establish whether HisKin *dbaS* is coupled to the glycerol GlpF transport facilitator in *Clostridium* sp. IBUN 13A, because of its cytoplasmic location.

The study of the histidine kinase sensor and the response regulator effector domains should be highlighted in sequence analysis since catalytic and REC domains are highly conserved and, as a result, do not provide conclusive information (Mascher 2006). Nevertheless, the analysis of the catalytic domain in *dbaS* allowed us to predict the absence of the F-box. The H-box also revealed the substitution of the canonical histidine residue by lysine (K-216) (Figure 1-a residue 464); this has been previously identified in *C. butyricum* (Raynaud et al. 2003). Even though a histidine is not the phosphorylated residue in different bacterial protein homologues (Foussard et al. 2001), only two proteins were identified in our review in which histidine was substituted by arginine (slr1414_ *Synechocystis* sp. and mth1260_ *Methanobacterium* sp.).

The identification of this substitution in multiple alignments suggests that the lysine could be phosphorylated to transfer the phosphate group to the response regulator, making the histidine residue substitutable by another residue, this can be observed in the study of conserved traces (ETS) (Figure 1-a). The P and F residues surrounding the histidine residue seemed to have a greater conservation because, perhaps, they play an important role in phosphoryl group interaction with the phosphorylatable residue.

The thermodynamic differences in phosphoryl group transfers from phosphorylated lysine to aspartic residues from the response regulator should be studied. The phosphorylated histidine phosphoimidazole bond seemed to be chemically ideal for this reaction (West & Stock 2001). Trace analysis did not otherwise show glycine residues as being the most conserved in the G-boxes (Figure 1-a). As this motif plays an important role in phosphotransfer (Grebe & Stock 1999), perhaps arginine and aspartic acid residues could have a more direct interaction with adenine nucleotide phosphates.

Even though the residues located in the X-box were not as conserved as in the other boxes, because this motif plays a structural role (Grebe & Stock 1999), four trace residues were identified in the alignment within the group 8 HQ family (L-492, R-497, E-513, R-527) (Figure 1-a).

The *dbaA* protein was confirmed as being the two-component system response regulator because the characteristic REC and HTH domains were identified. Five conserved residues were identified in the REC domain: D-14, D-15 (located after the first β/α loop), D-62, K-114 and Y-111 (Figure 1-b). However, trace analysis and multiple alignments revealed another difference in the REC's domain: the most conserved residues were located in the protein's N-terminal region; this demonstrates the particularity of the proteins of *Clostridium* sp. IBUN 13A. The threonine residue located in the active region, important in propagating conformational change following HisKin phosphorylation (West & Stock 2001), was not observed (I-92). Also, the formation of four α -helices interleaved between β -sheets (Alves & Savageau 2003, Casino et al. 2010) was not observed due to the lack of $\beta 5$ (Figure 1-b).

The conserved threonine hydroxyl group moves away from its position when proteins are phosphorylated; this allows the space left by such movement to become filled by a conserved aromatic residue (F/Y) moving from an exposed position to a buried position on the $\alpha 4/\beta 5/\alpha 5$ surface, thereby inducing conformational change (Stock & Da Re 2000). Threonine substitution for isoleucine in the *dbaA* protein might not affect its functionality, despite the absence of a hydroxyl group. Consequently, these residues were not observed in the trace analysis, contrarily to the three aspartic residues and the lysine residue, without which signal transduction could be deleteriously affected.

There was no discrepancy in conserved trace residues in multiple *araC* type proteins in the *dbaY* and *dbaA* transcriptional regulator domains. The presence of the G-hydrophobic residue in the small-residue triad could be observed in the HTH DNA-binding domain (Aravind et al. 2005), located in the $\alpha 4/\alpha 5$ loop (G-349, F-350, S-351); comparable, to the canonical hydrophobic residues in $\alpha 3$ (L-334) and

$\alpha 5$ (F-356) (Figure 1-c). DNA binding motifs, from the HTH domain, were organized to facilitate DNA-protein interface binding to $\alpha 3$ residues (recognition helix) and the presence of hydrophobic residues in $\alpha 1$ and $\alpha 3$, stabilizing the domain (Iyer & Aravind 2012). It was also confirmed that these domains presented a tetra-helicoidal conformation, characterized by an additional C-terminal helix (Aravind et al. 2005).

The phylogenetic analysis presented a defined topology for multiple genetic duplication events that were correlated to structural analysis. Such gene duplication-based distribution defines structural evolutionary domains for this highly promiscuous type of protein. The separation of the clade formed by the Colombian strain IBUN 13A and *C. butyricum* was observed in the group of Hiskin sequences (protein *dbaS*) (Figure 2), possibly due to the particular organization of residues from the conserved H-box motif in which canonical histidine had been substituted for the lysine residue. This substitution was not found in any other histidine kinase and suggested the probability of a constituted ancestral domain.

It is likely that the domains developed according to a strong purifying selection model (Table 1) suggest that current populations have precise defined functional niches for each domain without having polymorphism. This is intriguing because of the enormous number of current representatives for the groups.

Conclusion

Our study has reported the sequencing, identification and *in silico* characterization of 1,3-propanediol operon transcriptional regulators. The *Clostridium* sp. IBUN 13A strain *dbaS* and *dbaA* genes were identified, leading to the proposal of a two-component signal transduction system regulation mechanism as described previously for *C. butyricum*. However, the most important finding was the identification of a third gene named *dbaY*, which encodes a putative regulator protein. Annotating and studying these deduced protein domains has led to the description of their molecular evolution and to the elucidation of such signal transduction system physiological roles. According to these

findings, we propose that the *dbaY* protein is implicated in the regulation of glycerol metabolism with the *dbaS/dbaA* two-component system. The identification of these three genes constitutes a first approach to overall glycerol metabolism regulation and will prompt genetic manipulation strategies to improve the fermentation process for *Clostridium* sp. 1,3-PD production.

Acknowledgments

This study was carried out within the '1,3-propanediol production from glycerol obtained during biodiesel production, using Colombian *Clostridium* sp. strains: research on the operon, fermentation parameters and its economic viability' project financed by COLCIENCIAS (Departamento Administrativo de Ciencia, Tecnología e Innovación) (project code 1101-1217848) and the Universidad Nacional de Colombia (project codes 20101007337 and 20101006947). We would like to thank Jason Garry (Instituto de Biotecnología, Universidad Nacional de Colombia) for translating the manuscript and José David Montoya for his contributions to improve this document.

Conflict of Interest

The authors declare that they have no conflict of interest regarding this publication. All the authors participated in the research and article preparation, and the final version has been approved by all of them.

References

- Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105. doi: 10.1093/bioinformatics/bti263
- Alves R, Savageau M (2003) Comparative analysis of prototype two-component systems with either bifunctional or monofunctional sensors: differences in molecular structure and physiological function. *Molecular Microbiology* 48:25–51. doi: 10.1046/j.1365-2958.2003.03344.x
- Anantharaman V, Aravind L (2005) MEDS and PocR are novel domains with a predicted role in sensing simple hydrocarbon derivatives in prokaryotic signal transduction systems. *Bioinformatics* 21:2805–2811. doi: 10.1093/bioinformatics/bti418
- Aravind L, Anantharaman V, Balaji S, Babu M, Iyer L (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* 29:231–262. doi: 10.1016/j.fmrre.2004.12.008
- Ayoub M, Abdullah AZ (2012) Critical review on the current scenario and significance of crude glycerol resulting from biodiesel industry towards more sustainable renewable energy industry. *Renewable & Sustainable Energy Reviews* 16:2671–2686. doi: 10.1016/j.rser.2012.01.054
- Barbirato F, Himmi E.H, Conte T, Bories A (1998) 1,3-propanediol production by fermentation: An interesting way to valorize glycerin from the ester and ethanol industries. *Industrial Crops and Products* 7:281–289. doi: 10.1016/S0926-6690(97)00059-9
- Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research* 33:W451–W454. doi: 10.1093/nar/gki487
- Cárdenas D, Pulido C, Aragón O, Aristizabal F, Suárez Z, Montoya D (2006) Evaluación de la producción de 1,3 –Propanodiol por cepas nativas de *Clostridium* sp. mediante fermentación a partir de Glicerol USP y glicerol industrial subproducto de la producción de Biodiesel. *Revista Colombiana de Ciencias Químico-Farmacéuticas* 35:120–137. (On-line) <http://www.revistas.unal.edu.co/index.php/rccquifa/article/viewFile/1600/2270>
- Casino P, Rubio V, Marina, A (2010) The mechanism of signal transduction by two-component systems. *Current opinion in structural biology* 20.6: 763-771. doi: 10.1016/j.sbi.2010.09.010
- Cheung J K, McGowan S, Rood JI (2005) Two-Component signal transduction systems in Clostridia. In: Dürre P (ed) Handbook on CLOSTRIDIA. Taylor & Francis Group, pp 545–560
- Foussard M, Cabantous S, Pedelacq J, Guillet V, Tranier S, et al. (2001) The molecular puzzle of two-component signaling cascades. *Microbes and Infection* 3:417–424. doi: 10.1016/S1286-4579(01)01390-9
- Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos J (1997) AraC/XylS family of transcriptional regulators. *Microbiology and Molecular Biology Reviews* 61:393–410
- Galperin M (2004) Bacterial signal transduction network in a genomic perspective. *Environmental Microbiology* 6:552–567. doi: 10.1111/j.1462-2920.2004.00633.x
- Galperin M (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *Journal of Bacteriology* 188:4169–4182. doi: 10.1128/JB.01887-05
- González-Pajuelo M, Andrade JC, Vasconcelos I (2005) Production of 1,3-Propanediol by *Clostridium butyricum* VPI 3266 in continuous cultures with high yield and productivity. *Journal of Industrial Microbiology and Biotechnology* 32:391–396. doi: 10.1007/s10295-005-0012-0
- Grebe TW, Stock J (1999) The histidine protein kinase superfamily. *Advances in Microbial Physiology* 41:139–227. doi: 10.1016/S0065-2911(08)60167-8

- Hoch J (2000) Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology* 3:165–170. doi: 10.1016/S1369-5274(00)00070-9
- Iyer L, Aravind L (2012) Insights from the architecture of the bacterial transcription apparatus. *Journal of Structural Biology* 179:299–319. doi: 10.1016/j.jsb.2011.12.013
- Jaimes C, Aristizabal F, Bernal M, Suárez Z, Montoya D (2006) AFLP fingerprinting of colombian *Clostridium* strains, multivariate data analysis and its taxonomical implications. *Journal of Microbiological Methods* 67:64–69. doi: 10.1016/j.mimet.2006.03.002
- Jung K, Fried L, Behr S, Heermann R (2012) Histidine kinases and response regulators in networks. *Current Opinion in Microbiology* 15:118–124. doi: 10.1016/j.mib.2011.11.009
- Kobir A, Shi L, Boskovic A, Grangeasse C, Franjevic D, Mijakovic I (2011) Protein phosphorylation in bacterial signal transduction. *Biochimica et Biophysica Acta* 1810:989–994. doi: 10.1016/j.bbagen.2011.01.006
- Kubiak P, Leja K, Mysza K, Celińska E, Spychala M, et al. (2012) Physiological predisposition of various *Clostridium* species to synthesize 1,3-propanediol from glycerol. *Process Biochemistry* 47:1308–1319. doi: 10.1016/j.procbio.2012.05.012
- Kumar S, Nei M, Dudley J, Tamura K. (2008) MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9:299–306. doi: 10.1093/bib/bbn017
- Lesnik E, Sampath R, Levene H, Henderson T, McNeil J, Ecker D (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*. *Nucleic Acids Research* 23:3583–3594. doi: 10.1093/nar/29.17.3583
- Mascher T (2006) Intramembrane-sensing histidine kinases: a new family of cell envelope stress sensors in Firmicutes bacteria. *FEMS Microbiology Letters* 264:133–144. doi: 10.1111/j.1574-6968.2006.00444.x
- Mascher T, Helmann J, Uden G (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiology and Molecular Biology Reviews* 70:910–938. doi: 10.1128/MMBR.00020-06
- Montoya D, Arévalo C, González S, Aristizabal F, Scharwz WH (2001) New solvent producing *Clostridium* sp. strains hydrolyzing a wide range of polysaccharides, are closely related to *Clostridium butyricum*. *Journal of Industrial Microbiology and Biotechnology* 27:309–335. doi: 10.1038/sj.jim.7000193
- Montoya D, Spitia S, Silvia E, Scharwz WH (2000) Isolation of mesophilic solvent-producing clostridia from colombian sources; Physiological characterization, solvent production and polysaccharide hydrolysis. *Journal of Biotechnology* 79:117–126. doi: 10.1016/S0168-1656(00)00218-2
- Montoya JD (2008) Determinación de la secuencia de genes putativamente involucrados en la producción de 1,3-Propanodiol en la cepa nativa colombiana *Clostridium* IBUN 158B. Master thesis. Faculty of Sciences, Universidad Nacional de Colombia, Colombia
- Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* 14:454–459. doi: 10.1016/j.copbio.2003.08.005
- Papanikolaou S, Aggelis G, Ficka, Chevalot I, Galiotou-Panayotou M, et al. (2008) Biotechnological valorisation of raw glycerol discharged after bio-diesel (fatty acid methyl esters) manufacturing process: Production of 1,3-propanediol, citric acid and single cell oil. *Biomass and Bioenergy* 32:60–71. doi: 10.1016/j.biombioe.2007.06.007
- Papanikolaou S, Ruiz-Sánchez P, Pariset B, Blanchard F, Fick M (2000) High production of 1,3-propanediol from industrial glycerol by a newly isolated *Clostridium butyricum* strain. *Journal of Biotechnology* 77:191–208. doi: 10.1016/S0168-1656(99)00217-5
- Paulo da Silva G, Mack M, Contiero J (2009) Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnology Advances* 27:30–39. doi: 10.1016/j.biotechadv.2008.07.006
- Quilagay D M, Montoya J D, Suárez Z, Bernal M, Montoya D (2010) Analysing the *dhbT* gene in Colombian *Clostridium* sp. (Clostridia) 1,3-propanediol-producing strains. *Universitas Scientiarum* 15:17–26
- Raynaud C, Sarçabal P, Meynial-Salles I, Croux C, Soucaille P (2003) Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proceedings of the National Academy of Sciences* 100:5010–5015. doi: 10.1073/pnas.0734105100
- Raynaud C, Jieun L, Sarçabal P, Croux C, Meynial-Salles I, Soucaille P (2011) Molecular characterization of the glycerol-oxidative pathway of *Clostridium butyricum* VPI 1718. *Journal of Bacteriology* 193:3127–3134. doi: 10.1128/JB.00112-11
- Rondon M, Escalante J (1996) in vitro analysis of the interactions between the PocR regulatory protein and the promoter region of the cobalamin biosynthetic (*cob*) operon of *Salmonella typhimurium* LT2. *Journal of Bacteriology* 178:2196–2203
- Saxena R.K, Anand P, Saran S, Isar J (2009) Microbial production of 1,3-propanediol: Recent developments and emerging opportunities. *Biotechnology Advances* 6:895–913. doi: 10.1016/j.biotechadv.2009.07.003
- Sigrist C, Cerutti L, de Castro E, Cuche BA, Hulo N, et al. (2013) New and continuing developments at PROSITE. *Nucleic Acids Research* 41: D344 - D347. doi: 10.1093/nar/gks1067

- Stock J, Da Re S (2000) Signal transduction: Response regulators on and off. *Current Biology* 10 (11): R420–R424. doi: 10.1016/S0960-9822(00)00509-1
- Sun J, van den Heuvel J, Soucaille P, Qu Y, Zeng A-P (2003) Comparative genomic analysis of *dba* regulon and related genes for anaerobic glycerol metabolism in bacteria. *Biotechnology progress* 19:263–272. doi: 10.1021/bp025739m
- Tetsch L, Jung K (2009) The regulatory interplay between membrane-integrated sensors and transport proteins in bacteria. *Molecular Microbiology* 73:982–991. doi: 10.1111/j.1365-2958.2009.06847.x
- Togawa RC, Brigido MM (2003) PHPH: Web based tool for simple electropherogram quality analysis. Poster Presented at the 1st International Conference on Bioinformatics and Computational Biology. May 14–16, 2003. Ribeirão Preto, Brasil. Available in: http://asparagin.cenargen.embrapa.br/phph/ribeirao_preto_poster.pdf, [Retrieved 05/07/2014]
- Townsend GE, Raghavan V, Zwir I, Groisman EA (2013). Intramolecular arrangement of sensor and regulator overcomes relaxed specificity in hybrid two-component systems. *Proceedings of the National Academy of Sciences* 110(2):E161–E169. doi: 10.1073/pnas.1212102110
- West A, Stock A (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences* 26:369–376. doi: 10.1016/S0968-0004(01)01852-7
- Yao R, Shimizu K (2013) Recent progress in metabolic engineering for the production of biofuels and biochemicals from renewable sources with particular emphasis on catabolite regulation and its modulation. *Process Biochemistry* 48:1409–1417. doi: 10.1016/j.procbio.2013.02.032
- Yu N, Wagner J, Laird M, Melli G, Rey S, et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26:1608–1615. doi: 10.1093/bioinformatics/btq249
- Zeng A-P, Biebl H (2002) Bulk chemicals from biotechnology: The case of 1,3-Propanediol production and the new trends. *Advances in Biochemical Engineering/Biotechnology* 74:239–259. doi: 10.1007/3-540-45736-4_11

Análisis computacional de los reguladores transcripcionales del operón 1,3-propanodiol: indicios sobre la regulación del metabolismo del glicerol en *Clostridium* sp.

Resumen. Se diseñó una estrategia de amplificación, secuenciación y caracterización bioinformática de los genes reguladores del operón 1,3-propanediol (1,3-PD) de la cepa nativa colombiana *Clostridium* sp. IBUN 13A, relacionada taxonómicamente con *Clostridium butyricum*. Se identificaron tres genes que pueden estar involucrados en la regulación transcripcional de dicho operón: los genes *dbaS* y *dbaA* -a través de un sistema de transducción de señales de dos componentes- y un tercer gen que se denominó *dbaY*, que codifica para un regulador transcripcional putativo, similar a los dominios presentes en las proteínas del sistema DhaS/A. Los análisis filogenéticos indican que las proteínas predichas presentan una estructura modular con dominios homólogos a diferentes sistemas de transducción de señales, pero muestran diferencias importantes en los residuos conservados, lo que sugiere que podrían ser estos los dominios ancestrales. La predicción de funciones postula un mecanismo de regulación de las proteínas estudiadas sobre el promotor del operón 1,3-PD de la cepa nativa como respuesta a la presencia de glicerol en el medio, lo cual aporta información importante sobre la regulación global del metabolismo del glicerol en *Clostridium* sp.

Palabras clave: *Clostridium* sp.; metabolismo del glicerol; 1,3-propanodiol; genes reguladores; sistema de dos componentes

Análise computacional dos reguladores transcripcionais do operon do 1,3-Propanodiol: Panorama da regulação do metabolismo do glicerol no *Clostridium* sp.

Resumo. Nesta pesquisa foi feita uma estratégia para a amplificação, sequenciamento e caracterização bioinformática dos genes reguladores do operon 1,3 propanodiol (1,3-PD) da cepa colombiana *Clostridium* sp. IBUN 13A, relacionada taxonomicamente com o *Clostridium butyricum*. Têm sido identificados três genes que podem estar envolvidos na regulação transcripcional do operon. Os genes *dbaS* e *dbaA* por meio de um sistema de dois componentes e o terceiro gene nomeado de *dbaY*, que codifica para um regulador transcripcional putativo, parecido com os domínios presentes nas proteínas do sistema DhaS/A. A análise filogenética mostra que estas proteínas apresentam uma estrutura modular com domínios homólogos a diferentes sistemas de tradução de sinais, mas pressupõem diferenças importantes nos resíduos conservados, indicando provavelmente que possam constituir os domínios ancestrais. De acordo com a predição de funções, é postulado um mecanismo de regulação do sistema DhaS/A, DhaY sobre o promotor do operon 1,3-DP da cepa nativa, como resposta à presença de glicerol no meio, aportando informações importantes da regulação global do metabolismo do glicerol no *Clostridium* sp.

Palavras-chave: *Clostridium* sp.; metabolismo do glicerol; 1,3-propanodiol; genes reguladores; sistema de dois componentes