J. Biochem. 2010;148(3):299–308 doi:10.1093/jb/mvq064

JB THE JOURNAL OF BIOCHEMISTRY

A novel thermostable carboxylesterase from *Geobacillus thermodenitrificans*: evidence for a new carboxylesterase family

Received April 6, 2010; accepted June 14, 2010; published online June 29, 2010

David M. Charbonneau[1,2], Fatma Meddeb-Mouelhi[1,3] and Marc Beauregard[1,2,*]

[1]Département de chimie-biologie, Université du Québec à Trois-Rivières, 3351 Boul. Des Forges, C.P. 500 Trois-Rivières (Québec) G9A 5H7, [2]PROTEO, Université Laval, 2705 Boul. Laurier, Ste-Foy (Québec) G1V 4G2 and [3]Buckman Laboratories of Canada, 351, Joseph-Carrier Vaudreuil-Dorion (Québec) J7V 5V5, Canada

*Marc Beauregard, Département de chimie-biologie, Université du Québec à Trois-Rivières, C.P. 500, Trois-Rivières (Québec), G9A 5H7, Canada, Tel: 819 376 5011 (Ext. 3354), Fax: 819 376 5084, email: marc.beauregard@uqtr.ca

A novel gene encoding an esterase from *Geobacillus thermodenitrificans* strain CMB-A2 was cloned, sequenced and functionally expressed in *Escherichia coli* M15. Sequence analysis revealed an open reading frame of 747 bp corresponding to a polypeptide of 249 amino acid residues (named EstGtA2). After purification, a specific activity of 2.58 U mg^{-1} was detected using *p*-NP caprylate (C8) at 50°C and pH 8.0 (optimal conditions). The enzyme catalyses the hydrolysis of triglycerides (tributyrin) and a variety of *p*-nitrophenyl esters with different fatty acyl chain length (C4–C16). The enzyme has potential for various industrial applications since it is characterized by its activity under a wide range of pH, from 25 to 65°C. Using *Geobacillus stearothermophilus* Est30 esterase structure as template, a model of EstGtA2 was built using ESyPred3D. Analysis of this structural model allowed identifying putative sequence features that control EstGtA2 enzymatic properties. Based on sequence properties, multiple sequence comparisons and phylogenetic analyses, this enzyme appears to belong to a new family of carboxylesterases.

Keywords: Esterase/*Geobacillus thermodenitrificans*/*p*-NP caprylate/salt bridges/thermal stability/tributyrin.

Abbreviations: DMSO, dimethyl sulfoxide; IPTG, isopropyl-β-D-1-thiogalactopyranoside; LB, Luria broth; MSA, multiple sequence alignment; MWCO, molecular weight cut-off; OD, optical density; ORF, open reading frame; PCR, polymerase chain reaction; PDB, protein data bank; *p*-NP, *para*-nitrophenol; RMS, root-mean-square; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside.

Lipolytic enzymes are widely distributed in nature and are abundantly found in animals, plants and microorganisms (*1*). Carboxylesterases (EC 3.1.1.1) are hydrolases that catalyse the hydrolysis or synthesis of ester bounds in short chain aliphatic or aromatic esters (≤10 carbon atoms), in contrast with triacylglycerol lipases (EC 3.1.1.3) which can be defined as carboxylesterases that catalyse the hydrolysis or synthesis of long chain acylglycerol (≥10 carbon atoms) (*2*, *3*).

Microbial carboxyesterases have broad substrate specificity, exhibit high regio- and stereospecificity, require no cofactor and are active in organic solvents. They have important physiological, industrial and medical applications, namely in the synthesis of stereospecific compounds including the metabolic processing of drugs and antimicrobial agents (*4*, *5*). Carboxylesterases from thermophiles also show thermal stability at elevated temperatures and have become enzymes of special interest in biotechnological applications including: detergent formulation, food processing, cosmetics, fine chemical and pharmaceutical industry (*6*, *7*).

Crystallographic structure investigations have shown that all carboxylesterases and lipases belong to the α/β hydrolase fold family (*8*) and bear a catalytic triad formed by Ser-Asp/Glu-His. Microbial carboxylesterases have been classified into eight families based on their sequences and biological properties (*9*). Here we describe the cloning and characterization of a novel thermostable carboxylesterase from *Geobacillus thermodenitrificans* named EstGtA2. We suggest that EstGtA2 is a member of a new family of carboxylesterases which diverge from families suggested by the classification of Arpigny and Jaeger (*9*).

Materials and Methods

Isolation of bacteria from environmental samples

Geobacillus thermodenitrificans CMB-A2 was isolated in manure compost from a local distributor (Fafard et frère, St-Bonaventure, Québec, Canada). Compost was incubated at 60°C for 16 h and used to grow microbial colonies on Nutrient-agar plates. Colonies were grown at 60°C on emulsified minimal medium supplement with tributyrin 1% (v/v). Isolated colonies were identified by 16S rRNA gene sequencing as *G. thermodenitrificans* CMB-A2 (GenBank accession number GQ293454). Pure culture from single colony was cultivated in LB medium for 16 h at 60°C with agitation (230 r.p.m.) and preserved as suspension in 10% (v/v) glycerol at −80°C.

Strains, plasmids and culture conditions

Geobacillus thermodenitrificans strain CMB-A2 was used in this study. *Escherichia coli* XL1 Blue: endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F′[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17($r_K^- m_K^+$) (Stratagen®) was used for the cloning of EstGtA2 gene. *Escherichia coli* M15[pREP4] [NaI^S, Str^S, Rif^S, Thi$^-$, Lac$^-$, Ara$^+$, Gal$^+$, Mtl$^-$, F$^-$, RecA$^+$, Uvr$^+$, Lon$^+$] (QIAexpress®) was used for expression of the recombinant protein. The resulting expression system named *E. coli* M15[pREP4]/pQE31-EstGtA2 was cultured in Luria Bertani broth (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl), pH 7 containing ampicillin (100 μg/ml) and kanamycin (25 μg/ml) at 37°C and 230 r.p.m.

Preparation of genomic DNA

Genomic DNA was isolated from CMB-A2 using a standard method (*10*, *11*). The quality of the preparation was evaluated by spectrophotometry and by agarose gel electrophoresis.

Polymerase chain reaction

Genomic DNA was used as template for polymerase chain reaction (PCR). The 16S rRNA gene was amplified using the following primers: 27F: 5′-AGATTTGATCCTGGCTCAG-3′ and 1522R: 5′-AAGGAGGTGATCCARCCGCA-3′ (specific for universally conserved region of bacterial 16S rRNA gene sequences). The locus tag GTNG_1710 (Gene ID 49666446) identified in *G. thermodenitrificans* NG80-2 genome (*12*, *13*) was used to design oligonucleotides for directional cloning. The following primers were synthesized: est*Gt-Kpn*I F: 5′ TAATTAGGTACCTATGAAAGAACGATATC CTGTACTT 3′ and est*Gt-Hind*III R: 5′ TATTAAAGCTTTCAAG CATGTTTGGCGAA 3′ and were used for PCR. PCR mixtures were prepared in 50 μl containing 1× Taq Buffer^{TM}, 200 μmol l^{-1} dNTP, 1.25 units Taq DNA polymerase (New England Biolab) and 0.4 μmol l^{-1} of the respective primers. Initial denaturation was performed at 95°C for 4 min. Amplification was established in 5 cycles of 30 s at 95°C, 30 s at 50°C and 2 min at 72°C. followed by 30 cycles of 30 s at 95°C, 30 s at 58°C and 2 min at 72°C. Final extension step was achieved at 72°C for 10 min. PCR products were purified using Minelute PCR purification kit (Qiagen®) and analysed on 1% agarose gel electrophoresis.

Cloning of PCR products

The purified PCR product of the EstGtA2 ORF was digested with *Kpn*I and *Hind*III restriction endonucleases for 16 h at 37°C, purified using Minelute PCR purification kit (Qiagen®) and then ligated (4 h at 16°C) into *Hind*III/*Kpn*I-digested pQE31 expression vector (Qiagen®). The pQE31-EstGtA2 recombinant plasmid was transferred by electroporation in *E. coli* XL1 Blue (electrocompetent cells). Transformed cells were selected on LB-agar with ampicillin 100 μg/ml, containing X-Gal and IPTG. Plasmid DNA isolation was performed using QIAprep^{TM} (Qiagen®) and the integrity of the insert was tested by colony PCR using the EstGtA2 primers set and by double digestions (*Kpn*I and *Hind*III) before the plasmid was used for sequencing of the insert.

DNA sequencing and bioinformatics analysis

Both strands of the recombinant EstGtA2 gene were sequenced using an ABI Prism 3700 by the Biomolecular analysis platform (Université Laval). Sequence edition and analysis were performed with Clone Manager Professional Suite version 7.03. Sequences for comparative study were retrieved from protein and nucleotides database on the NCBI Entrez at http://www.ncbi.nlm.nih.gov/Entrez/. Similarity searches were performed with BLAST 2.0 program (*14*) in the GenBank database (*15*) and with BLASTP (*16*). The conserved domains present in the EstGtA2 amino acid sequence were analysed using the NCBI Conserved Domain Database (CDD) (*17*). Multiple sequence alignments were performed with Clone Manager Professional Suites version 7.03 using a BLOSUM62 scoring matrix. The program MEGA 4.1 (*18*) was used for multiple sequences alignments using CLUSTAL W (*19*) and for construction of a phylogenetic tree using the neighbour-joining method (*20*). The stability of the relationships was assessed by a bootstrap analysis based on 1,000 resamplings of the neighbour-joining (*21*).

Structural modelling

The EstGtA2 3D structure was predicted by homology modelling using the EsyPred3D program (*22*). The crystal structure of carboxylesterase Est30 from *G. stearothermophilus* (PDB code 1TQH_A) was used as template (*5*). Image of the resulting 3D model was generated using Pymol (DeLano Scientific, Palo Alto, CA, USA) and the quality of the model regarding geometry and stereochemistry was evaluated by Whatcheck and PROCHECK analysis (*23*, *24*) using the Swiss-model workspace (*25*). The evaluation of salt bridges based on the EstGtA2 model was performed with the ESBRI interface (*26*).

Protein expression

The pQE31-EstGtA2 recombinant plasmid was transferred by electroporation in *E. coli* M15[pREP4] electrocompetent cells. Transformed cells were selected on LB-agar with ampicillin (100 μg/ml) and kanamycin (25 μg/ml). After selection, clones of *E. coli* M15[pREP4]/pQE31-EstGtA2 were cultured as described earlier. Expression of EstGtA2 was carried out at an $OD_{600\,nm}$ of 0.6 by addition of 1 mM IPTG for 150 min at 37°C (agitation 230 r.p.m.).

Protein purification and determination

Induced cells were centrifuged at 4°C for 30 min at 4,000*g* and the pellets were stored at −80°C. Cell samples (250 μl) were taken before and after induction for SDS–PAGE analysis. The His-tagged protein was purified under native conditions using Ni-NTA^{TM} nickel affinity resin (Qiagen®) according to manufacturer specifications at pH 8 using imidazole for elution. In order to remove salts and imidazole, purified EstGtA2 was dialysed twice against 2 l of phosphate buffer 50 mM, pH 8.0 for 48 h at 4°C using 3,500 MWCO Slide-A-Lyzer Dialysis Cassettes® (Thermo Scientific). Concentration of protein was determined by the bicinchonininc acid method (*27*) using bovine serum albumin as standard (Sigma-Aldrich, St-Louis, USA). The molecular weight of the proteins was determined by SDS–PAGE and proteins were detected by Coomassie blue staining following standard procedures. A low molecular range protein standard solution (Bio-Rad Laboratories Inc) was used as molecular weight marker.

Enzyme assays

Carboxylesterase activity was determined spectrophotometrically by measuring the amount of *p*-nitrophenol (*p*-NP) released after hydrolysis of different *p*-NP esters. The *p*-nitrophenyl esters were prepared in DMSO. The standard reaction (0.2 ml) contained 50 μM p-NP ester, 50 mM of appropriated buffer with 14 μg of the purified enzyme. The reaction was performed in 96-well microplates using VWR heating blocks. The production of *p*-NP was monitored at 405 nm using a Thermo spectrophotometer. Specific standard curves using known amount of *p*-NP were measured for each set of conditions (pHs, temperatures) in order to quantify the amount of *p*-NP released during the enzymatic reactions for each condition. One unit of activity was defined as the amount of enzyme releasing 1 μmol of *p*-NP per min under assay conditions.

Chain length specificity, optimal pH and temperature

Chain length specificity was determined by hydrolysis of different *p*-NP esters: *p*-NP butyrate (C4); *p*-NP caprylate (C8); *p*-NP decanoate (C10); *p*-NP laurate (C12); *p*-NP myristate (C14); *p*-NP palmitate (C16). For temperature and pH studies, the standard reaction (0.2 ml) was carried out with *p*-NP caprylate (C8) as substrate since EstGtA2 activity was maximal on this substrate. The effect of temperature was studied from 25 to 80°C. Activity was measured from pH 6 to 12 using the following buffers: sodium phosphate 50 mM was used from pH 6 to 8, Tris–HCl 50 mM was used at pH 9, CAPS 50 mM at pH 10 and sodium dibasic (Na_2HPO_4) 50 mM at pH 11 and 12 at 50°C (optimal temperature). Enzymatic activity on glyceryl tributyrate (tributyrin) was detected on agar-plate (1.5%) containing glyceryl tributyrate (1%) by appearance of a clear zone after incubation at 50°C from pH 6 to 12.

Thermal stability

Enzyme (1.5 ml) was incubated in 50 mM phosphate buffer (pH 8) in 2 ml tubes using a Nestlab-RT111 bath. The enzyme was incubated at the specified temperatures for different period of time and then the

activity was measured under standard condition at 50°C using *p*-NP caprylate (C8) as substrate.

Results

Sequence of EstGtA2

A thermophilic *Bacillus* strain CMB-A2 growing in the range of 50–65°C was isolated from manure compost and identified as *G. thermodenitrificans* based on 16S rRNA gene sequence [GenBank accession no. GQ293454, 99.5% identity to *G. thermodenitrificans* NG80-2 strain (*12*)]. The coding sequence of a putative esterase named EstGtA2 was amplified by PCR and cloned in *E. coli* M15 using the expression vector pQE31. Translation of DNA sequence revealed a 249 amino acid sequence comprising a pentapeptide formed by Gly-Leu-Ser-Met-Gly that matched the conserved pentapeptide Gly-X-Ser-X-Gly found in most esterases and lipases. Homology comparison revealed that EstGtA2 shows 99.2% identity with the amino acid sequence of a putative esterase/lipase (YP_001125819) predicted in the genome of *G. thermodenitrificans* NG80-2 (*13*). This enzyme corresponds to the putative gene identified in the related genome, and used for designing PCR primers (Materials and Methods section). Among the conserved domains identified in the EstGtA2 amino acid sequence we identified pfam00561 (α/β hydrolase fold family), COG1647 (esterases/lipases), and COG2267 (lysophospholipase).

Expression and purification of EstGtA2

The recombinant EstGtA2 was overexpressed in *E. coli* M15 as a 6× His tagged fusion protein and purified under native conditions as described in 'Materials and Methods' section. A new single band at the expected molecular weight of about 30 kDa was detected after purification (Fig. 1). The recombinant EstGtA2 was functionally expressed in *E. coli* M15 at a yield of 30 mg/l of culture.

Characterization of EstGtA2

The optimum pH and temperature were determined using a standard assay (using *p*-NP caprylate since EstGtA2 showed highest activity on this substrate, described in Materials and Methods). The enzyme has its maximal activity at pH 8 and displayed activity over a wide range of pH values, with 62% of relative activity observed at pH 6 and 36% at pH 11 (Fig. 2). EstGtA2 showed a temperature optimum of 50°C. A maximal activity (specific activity) of 2.58 U mg^{-1} was observed at 50°C and pH 8. The enzyme retained over 90% of its maximal activity at 55 and 60°C. However, a decrease in activity was observed at 65°C (83% left), while only 15% of relative activity was observed at 75°C (Fig. 3).

Substrate specificity was tested using *p*-NP esters of different chain length at pH 8 and 50°C. As shown in Fig. 4, EstGtA2 exhibited activity toward *p*-nitrophenyl esters with a broad range of acyl chain lengths. Maximal activity was observed toward *p*-NP caprylate (C8). Activity decreased when the chain length increased from C8 to C16, suggesting that EstGtA2 is an esterase. Relative activity toward other substrates tested was compared to activity on *p*-NP caprylate. EstGtA2 showed moderate activity toward *p*-NP decanoate (65%), *p*-NP laurate (58%), *p*-NP myristate

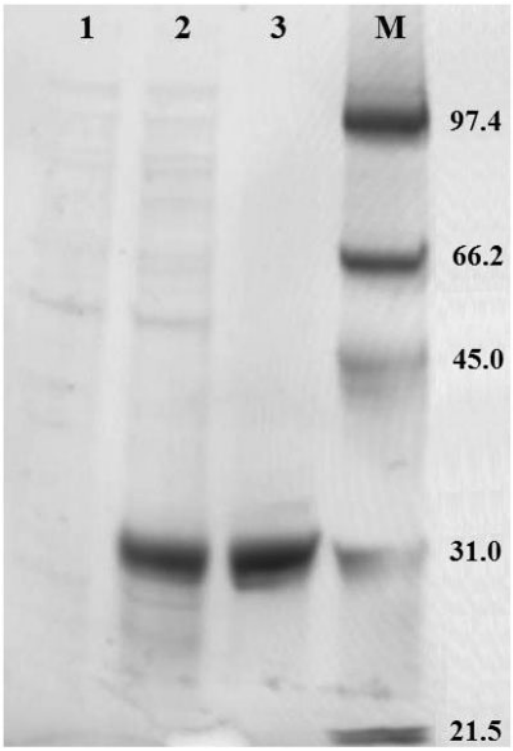

Fig. 1 SDS–PAGE of the recombinant EstGtA2. Lane 1: *E. coli* M15[pREP4]/pQE31-EstGtA2 total proteins extract before induction of EstGtA2; lane 2: total protein extract after induction 2.5 h with 1 mM IPTG; lane 3: purified 6× His tagged EstGtA2; M: low range molecular weight marker (97.4, 66.2, 45, 31 and 21.5 kDa).

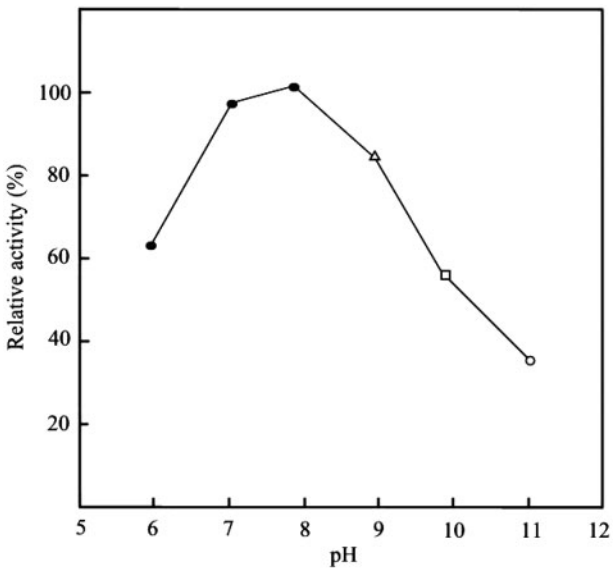

Fig. 2 Activity of EstGtA2 at different pH values. Activity of EstGtA2 was assayed with 14 μg of protein and 50 μM *p*-NP caprylate at 50°C in appropriate buffers. Closed circles: 50 mM sodium phosphate, pH 6–8; open triangle: 50 mM Tris–HCl, pH 9; open squares: 50 mM CAPS, pH 10; open circles: sodium phosphate dibasic, pH 11. Activity at pH 8 was defined as 100% relative activity.

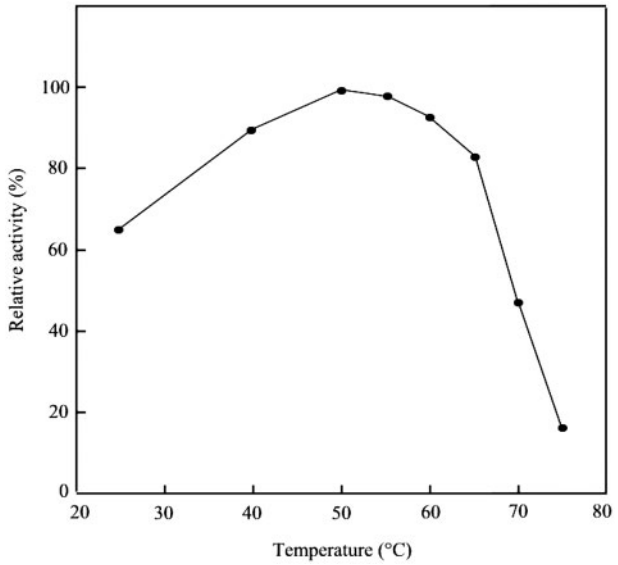

Fig. 3 Activity of EstGtA2 at different temperatures. Activity of EstGtA2 was assayed with 14 μg of protein and 50 μM *p*-NP caprylate in 50 mM phosphate buffer, pH 8 at specified temperatures. Activity at 50°C was defined as 100% relative activity.

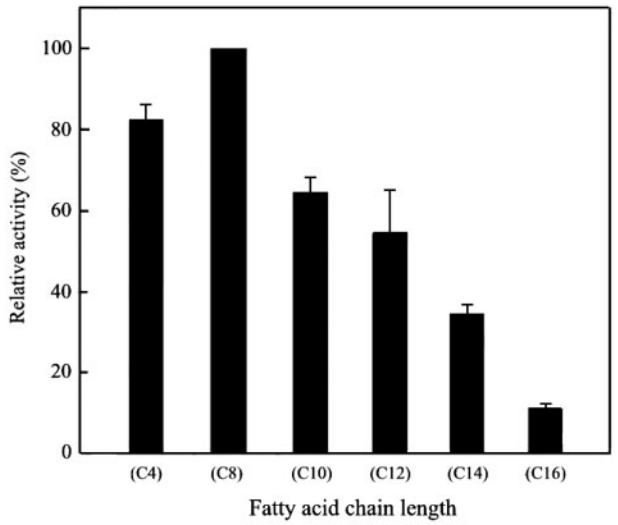

Fig. 4 Activity of EstGtA2 toward *p*-NP esters of different fatty acids chain length. Activity of EstGtA2 was assayed with 14 μg of protein on specified *p*-NP esters (50 μM) at 50°C in 50 mM phosphate buffer pH 8. Activity using *p*-NP caprylate (C8) as substrate was used as 100% relative activity.

(36%), and *p*-NP palmitate (12%). The substrate specificity was assayed from 25 to 60°C. Interestingly, relative activity toward long chain *p*-nitrophenyl esters (*p*-NP myristate and *p*-NP palmitate) was increased when the temperature increased to 50 and 60°C when compared with 25–40°C (data not shown). The enzyme showed also activity on glyceryl tributyrate (tributyrin) over a wide range of pH value (pH 6–12), but did not show any activity on longer acyl chain triglycerides (data not shown).

Thermal stability was investigated and the results showed that the enzyme was stable for >2 h at 60°C. The enzyme stability profile was similar from 50 to

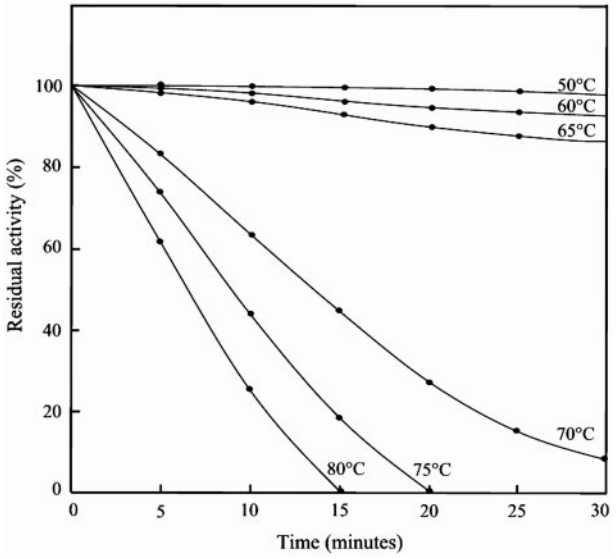

Fig. 5 Stability of EstGtA2 at different temperatures. Thermal stability was assayed at pH 8. Enzyme was incubated for different period of time at 50, 60, 65, 70, 75 and 80°C and then activity was measured under the standard reaction conditions at 50°C.

65°C, however an important decrease in stability was observed at 70°C, where the enzyme retained 63% of its initial activity, down to 43% at 75°C and further to 25% at 80°C after 10 minutes (Fig. 5).

Sequence comparison of EstGtA2

Relatedness of EstGtA2 to known carboxylesterases and lipases was investigated with a BLAST analysis which indicated high sequence similarity for the esterase/lipase (YP_001125819) of *G. thermodenitrificans* strain NG80-2 (99.2% identity, gene used for designing the PCR primers), monoacylglycerol lipase (P82597) from *Bacillus. Sp* H-257 (89% identity) (*28*, *29*) and acylglycerol lipase (YP_147178) from *G. kaustophilus* HTA426 (86% identity).

A multiple sequence alignment (MSA) was created for EstGtA2 with its closest homologues (99–86%): YP_001125819 from *G. thermodenitrificans* strain NG80-2; P82597 from *Bacillus. Sp* H-257; YP_147678 from *G. kaustophilus* HTA426 and its closest homologue (32% identity) for which a crystal structure was known, Est30 from *G. stearothermophilus* (PDB code 1TQH, hereafter named Est30Gs). The MSA shown in Fig. 6 also contains closest homologues to Est30Gs, namely EstA from *G. thermoleovorans*, GkCE from *G. kaustophilus* HTA426 (*30*) and Est30 from *G. thermodenitrificans* NG80-2 (another putative esterase from *G. thermodenitrificans* NG80-2 also named Est30) (*13*).

The MSA suggests two distinct groups of carboxylesterases. The first group which includes EstGtA2 (top sequences) shares ~30% identity with the second group including Est30Gs (lower half sequences). Despite the low sequence identity between these two groups, conserved blocks are identified in all the sequences of the MSA (Fig. 6). The putative catalytic triad has been identified in EstGtA2 based on comparison with Est30Gs and GkCE sequences for which

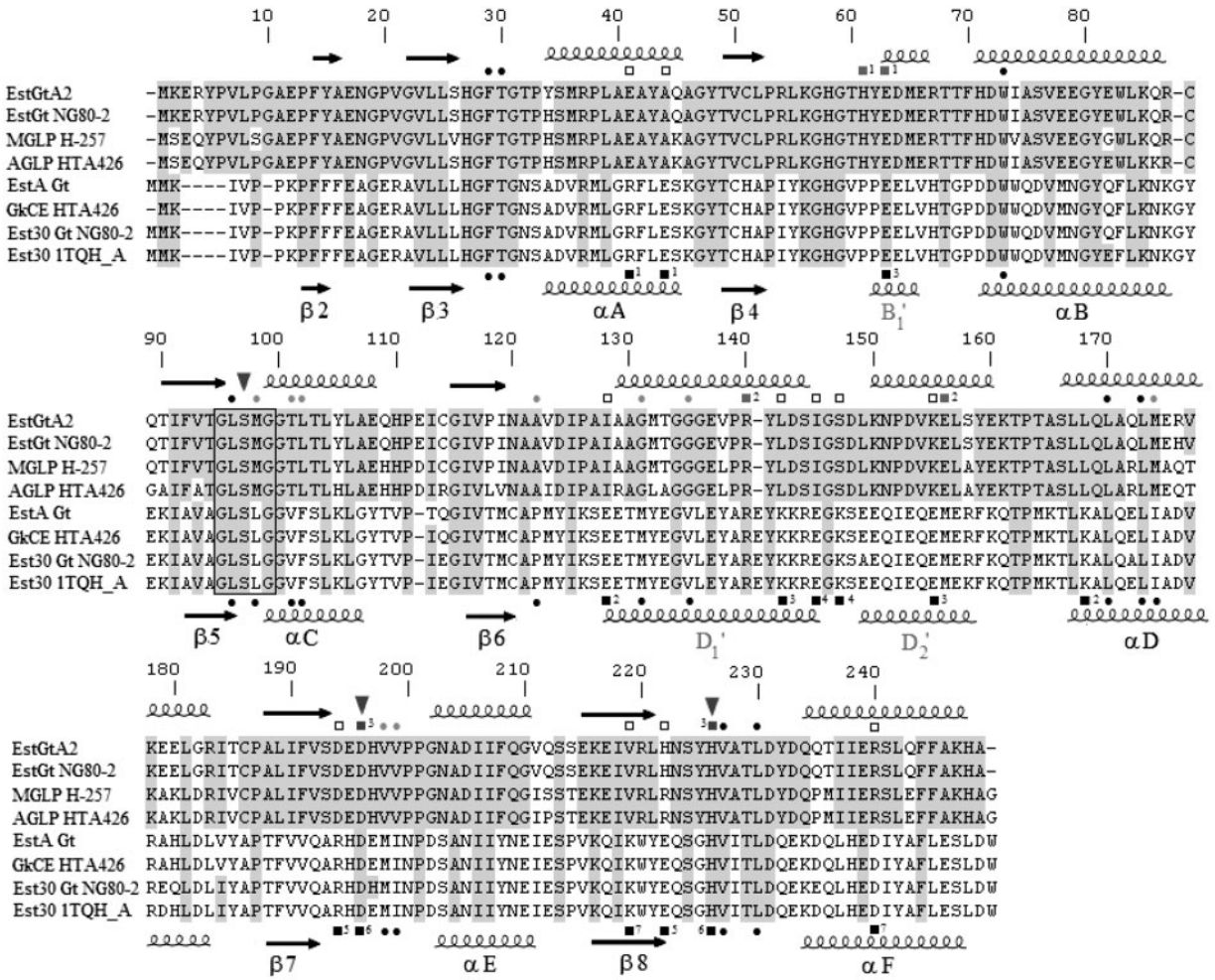

Fig. 6 Multiple sequence alignment for EstGtA2 from *G. thermodenitrificans* and related *Geobacillus* esterases. Sequences are grouped according to similarity. Closest homologues of EstGtA2 found in BLAST analysis (first three sequences) were EstGt from *G. thermodenitrificans* NG80-2, MGLP from *Bacillus sp.* H-257, AGLP from *G. kaustophilus* HTA426. Sequences related to Est30Gs from *G. stearothermophilus* were EstA from *G. thermoleovorans*, GkCE from *G. kaustophilus* HTA426 and Est30 from *G. thermodenitrificans*. Conserved residues are embedded in a solid background. Secondary structures of Est30Gs (as calculated from X-ray resolved structure) are indicated below the MSA, and secondary structures modelled for EstGtA2 are indicated above the MSA. Amino acids of the catalytic triad are identified with arrows. The putative salt bridges in the EstGtA2 and Est30Gs structures were analysed with ESBRI. Amino acids involved in salt bridges are identified by closed black squares and annotated with respective numbers in pairs (1–3 for EstGtA2 and 1–7 for Est30Gs) above the MSA for EstGtA2, and underneath for Est30Gs. Open squares above the MSA represent potential lost of salt bridges in EtGtA2 compared with Est30Gs. The residues of the hydrophobic specificity pocket identical between Est30Gs and EstGtA2 are indicated with black closed circles. Grey closed circles above the MSA represent differences in EstGtA2 specificity pocket compared with Est30Gs. MSA positions are numbered using EstGtA2 amino acid sequence.

residues of the catalytic triad have been determined (*5*, *30*). In Est30Gs, the nucleophilic serine residue is located at position 94 in the pentapeptide (Gly-Leu-S94-Leu-Gly) which corresponds to position Ser97 in EstGtA2 (located in the corresponding pentapeptide Gly-Leu-Ser97-Met-Gly), the catalytic aspartate and histidine residues, located at position Asp193 and His223 in Est30Gs, correspond to Asp196 and His226 in EstGtA2. These residues are conserved in all the sequence analysed in the MSA (Fig. 6).

Despite low sequence identity, structural similarity is suggested for EstGtA2 and Est30Gs. Secondary structures prediction for EstGtA2 suggests high conservation in comparison with those calculated from Est30Gs resolved structure. The alpha helices (αA, to αF) and the beta strands (β2 to β8) in Est30Gs appear to be highly conserved in EstGtA2. The three helices, (B_1', D_1' and D_2'), involved in the formation of a cap structure in Est30Gs are also predicted in EstGtA2 sequence (Fig. 6).

Structure modelling of EstGtA2

A 3D model of EstGtA2 was generated by homology modelling using Est30Gs (access 1TQH_A) as template. The Ramachandran plot (not shown) indicated that 94.1% of the residues had dihedral angles predicted in the core and 5.4% in allowed regions. Bond lengths, bond angles and torsion angles in the model were evaluated using WHAT IF program. A RMS z-score for a normally restrained data set is expected to be around 1.0. The bond length and bond angles

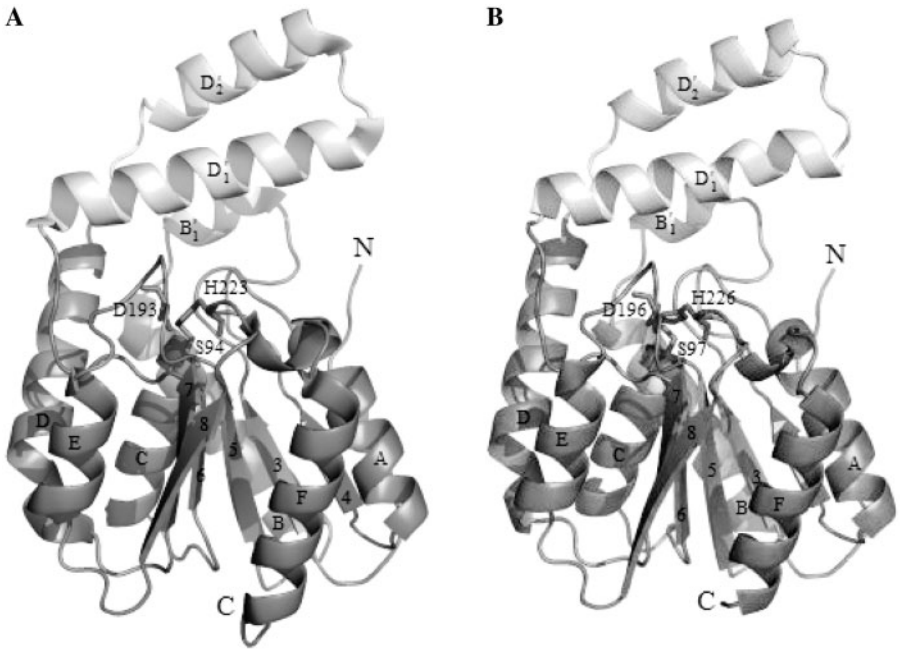

Fig. 7 3D model of EstGtA2. Ribbon structure of Est30Gs (A) and modelled structure of EstGtA2 (B). Amino acids of the catalytic triad are identified and shown in the centre of the respective structure. Alpha helices are identified (from A to F) and beta strands (*2*–*8*). The lid helices are identified (D1′, D2′ and B1′) and are protruding above the enzyme, and shown light grey.

were found to deviate normally from standards values (RMS z-score of 0.93 and 1.265) respectively.

The model shown in Fig. 7 suggests that EstGtA2 can adopt a 3D structure which is very similar to Est30Gs. It also suggests the existence of two distinct domains: A larger domain resembling the classical α/β hydrolase fold; and a smaller domain formed by three helices: a small helix (B_1') with two larger helix (D_1' and D_2'), involved in the formation of a cap structure which lies over the N-terminal end. The 3D model shows that the amino acid residues of the catalytic triad (Ser97, Asp196 and His226, positions suggested on the basis of the MSA, Fig. 6) are indeed located in close proximity and are arranged in catalytic triad-like configuration (Fig. 7). This catalytic triad is located at the interface of the two domains with Ser97 located at a sharp turn between the strand β5 and helix αC, called the nucleophile elbow. Asp196 and His226 are positioned on loops between β7-αE and β8-αF, respectively.

Phylogenetic tree of carboxylesterases

A phylogenetic tree of esterases and lipases (amino acid sequences) featuring some lipase families suggested from the classification of Arpigny and Jaeger (*9*) is shown in Fig. 8. Sequences used for MSA analysis (Fig. 6) formed two clusters which emerge from the canonical Arpigny and Jaeger families: the cluster including Est30Gs was named (N), and the cluster including EstGtA2 was named (N′). It should be noted that clusters N and N′ include the sequences used to build the MSA and show around 30% identity between each other. Sequences from the new cluster N′ show ∼15% identity with the sequences of the other families (I–VIII).

Discussion

In this study we characterized a new lipolytic enzyme (EstGtA2) cloned from a *Geobacillus thermodenitricans* strain that we named CMB-A2. This strain screened from compost under high temperature (60°C) was identified on the basis of a 16S rRNA analysis. The EstGtA2 sequence, structure and properties were compared with lipases and esterases identified in the classification of Arpigny and Jaeger (*9*) in order to assign EstGtA2 to a family (I–VIII). The novel EstGtA2 is somewhat similar to family VI. This family includes enzymes of a molecular mass in the range of (23–26 kDa), with no activity toward long chain triglycerides and ∼40% similarity with eukaryotic lysophospholipases. Sequence analysis shows that EstGtA2 is a relatively small esterase (27.4 kDa), which is active on short chain triglycerides (tributyrin) and contains a lysophospholipase domain, but did not display the expected similarity with eukaryotic lysophospholipases as suggest for esterases from family VI (*9*). Moreover, EstGtA2 has none of the conserved motifs identified for this family (*9*). We suggest that EstGtA2 is a member of a new family which diverges from the eight families of the classification by Arpigny and Jaeger (*9*).

It was already suggested that Est30Gs (from *G. stearothermophilus,* 1TQH) belonged to a new carboxylesterase family where all members of this new family should have a similar tertiary structure, including the three-helix cap of Est30 (*5*). EstGtA2 sequence shares 32% identity with Est30Gs, and structural similarity among these two enzymes was suggested on the basis of secondary structure alignment and 3D modelling. The 3D model of EstGtA2 suggests the presence of the three-helix cap structure of Est30 formed by helix B_1', D_1' and D_2'. Despite these similarities,

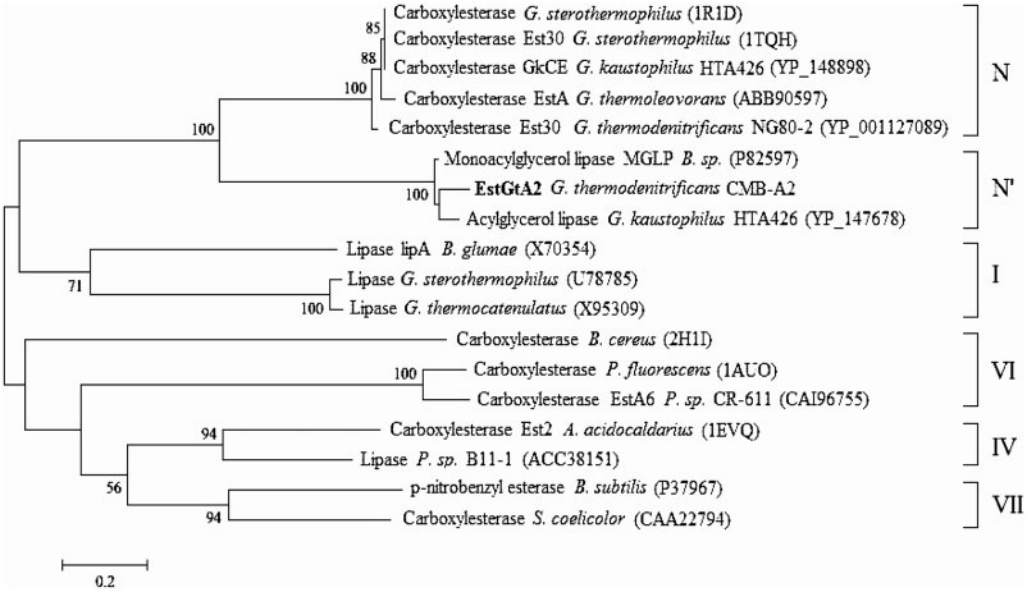

Fig. 8 Phylogenetic tree of EstGtA2 and related lipolytic enzymes. Phylogenetic tree representing EstGtA2 with related homologues and with lipolytic enzymes used in the classification of Arpigny and Jaeger (*9*). Roman numbers represent the lipolytic enzyme family according to the original classification. N and N′ represent the new clusters of carboxylesterases as suggested here.

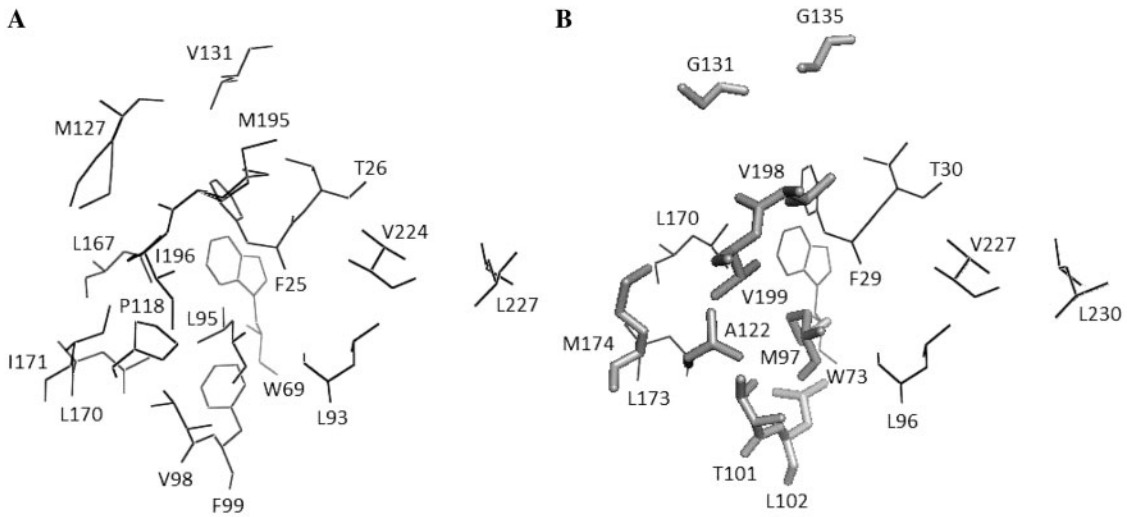

Fig. 9 Hydrophobic specificity pocket of Est30Gs and EstGtA2. Residues of the hydrophobic specificity pocket for substrate of Est30Gs (A), and as predicted by EstGtA2 modelling (B), are shown. Amino acid residue substitutions for EstGtA2 are shown as thick sticks.

important differences support the hypothesis that EstGtA2 belongs to another family.

A high conservation of the hydrophobic residues lining the substrate pocket was also defined as a criterion for members of the new family suggested for Est30Gs (*5*). However, ~50% of the residues of the hydrophobic specificity pocket are different in EstGtA2 compared to Est30Gs. As shown in Fig. 9, the residue L95(98) (first number corresponds to Est30Gs sequence, while number in brackets correspond to EstGtA2 sequence, i.e. as shown in Fig. 5) included in the pentapeptide GLSLG is replaced by M98 in EstGtA2 (leading to GLSMG). The residues V98(101), F99(102), P118(122), M127(131), V131(135), I171(174), M195(198) and I196(199) in Est30Gs are replaced by T101, L102, A122, G131, G135, M174, V198 and V199 in EstGtA2. Such differences may be responsible for EstGtA2 affinity for longer *p*-NP (residue substitutions indicated in Fig. 9). EstGtA2 was shown here to be active on C12-C14 and slightly active on C16, which are not hydrolysed by Est30Gs (*5*).

Another important difference is the optimal temperature of EstGtA2 (50°C) and Est30Gs (70°C) (*5*, *31*). It is generally recognized that salt bridges are involved in protein thermostability since thermophilic protein often display a high number of surface salt bridges in comparison with its mesophilic counterpart (*32*, *33*). Salt bridges were suggested in the EstGtA2 modelled structure and compared with Est30Gs, using

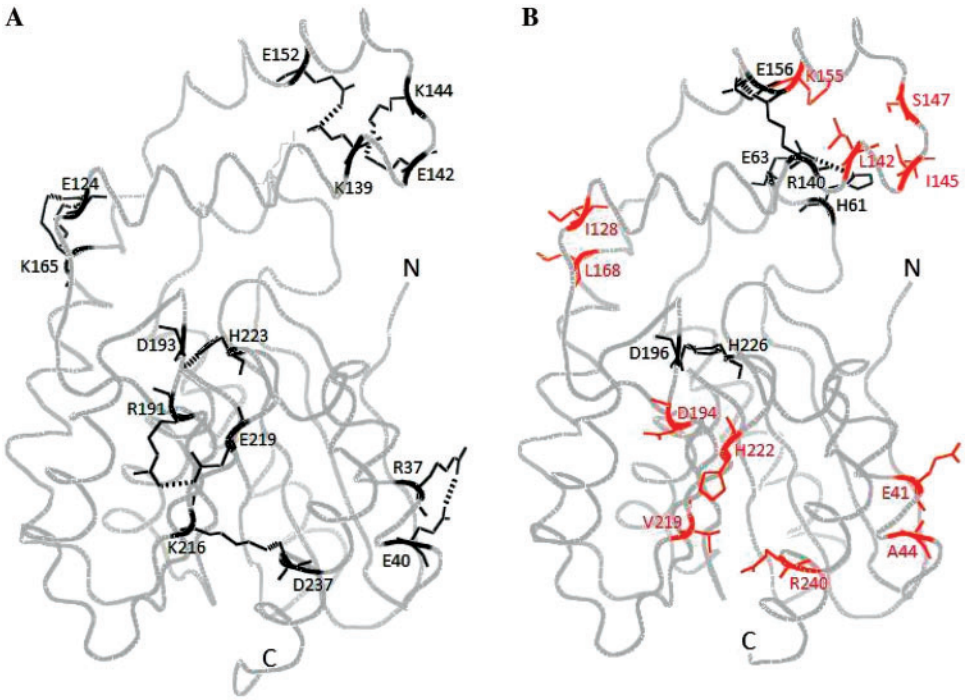

Fig. 10 Prediction of salt bridges in EstGtA2. Predicted salt bridges in Est30 (A) and EstGtA2 (B) structures are shown by dotted lines. Putative salt bridge residues (<3 Å) are shown in black, while amino acid substitutions leading to loss of bridges in EstGtA2 (compared to Est30Gs) are shown in red.

ESBRI interface (*26*) and the following criteria: salt bridges formation was inferred for a pair of oppositely charged (Asp or Glu with Arg, Lys or His) if one pair of Asp or Glu side-chain carboxyl oxygen atoms and side-chain nitrogen atoms of Arg, Lys or His are within a 4 Å distance (*34*). Analysis of EstGtA2 model (Fig. 10) shows only three salt bridges that met this criterion in comparison with seven for Est30Gs: a first bridge between H61 and E63 of the helix B_1' involved in the formation of the cap structure (3.83 Å); an inter-helices D_1'-D_2' salt bridge (3.66 Å) between R140 and E156, which seems to maintain these helices in close proximity and stabilize the cap domain; and finally between D196 and H226 (2.93 Å) in the putative catalytic triad.

As shown in Fig. 10, seven pairs of charged residues which met the criteria described above were found in the Est30Gs structure. The salt bridge D193(196)-H223(226) involved in the catalytic triad appears to be equivalent to D196-H226 in EstGtA2. The other six salt bridges are different in Est30Gs when compared with EstGtA2, and should contribute to its higher thermoresistance. A salt bridge (3.06 Å) links E142 (C terminal end of the helix D_1') to K144 (in the loop just before the N-terminal end of the helix D_2'). In addition, a salt bridge (2.94 Å) formed between K139 and E152 might link the helix D1′ to D2′, contributing to stabilizing the cap domain. It should be noted that K139 (D_1') may also form a salt bridge (3.43 Å) with E59 located in the helix B1′. Rotamer analysis would be useful for determining which putative salt bridge is most favoured. In one case or another, any salt bridge formed by K139 may contribute to the stabilization of the cap domain by linking the helix D_1' to D_2' or D_1' to B_1'. Moreover, the N-terminal end of the helix D_1' is bound to the N-terminal end of αD by a salt bridge (3.74 Å) formed between E124-K165. These interactions seem to confer a high stability of the cap domain in Est30 compared to EstGtA2 for which the cap structure was stabilized only by the inter-helices D_1'-D_2' salt bridge (R140-E156) and intra-helix B_1' salt bridge (H61-E63). Clearly, these salt bridges and consequent stabilization of the cap seem to be an important determinant of thermostability in carboxylesterase of this family (*G. stearothermophilus*).

Additional salt bridges were found to be potentially stabilizing for Est30Gs structure. An inter helix-strand, αF-β8 salt bridge formed between K216 and D237 (2.70 Å) might be highly important for the thermostability of Est30Gs since the N-terminal end of the strand β8 is also involved in a salt bridge (2.78 Å) with N-terminal end of the strand β7 (R191-E219) that might stabilize the residues D193 and H223 of the catalytic triad located on loops between β7-αE and β8-αF respectively. These stabilizing interactions are not observed in the EstGtA2 model since K216 was substituted for a valine (V219) and D237 substituted for an arginine (R240) in EstGtA2 (Fig. 10). In addition of these mutations, R191 is substituted for an aspartate (D194) and E219 for a histidine (H222) in EstGtA2. The pair D194-H222 formed by these substitutions in EstGtA2 is separated by a distance of 4.74 Å which does not meet the criterion of a 4 Å distance for salt bridge detection. Although an impact on thermostability from such variations in charged residue composition is expected, it remains that their proximity to the catalytic triad might also play a role in the

different pH activity curves recorded for the two enzymes compared here (optimal pH for EstGtA2 is 8, and 9 for Est30).

The detailed analysis of MSA (Fig. 5) allowed identification of localized sequence features that are probably responsible for the important differences between EstGtA2 and Est30Gs enzymatic properties. Furthermore, these sequence features are well conserved in the cluster N when compared to the sequences related to EstGtA2 (cluster N′), and their mutations in EstGtA2 are also well conserved in the N′ family proposed here as shown in Fig. 8.

Analysis of CMB-A2 16S rRNA and EstGtA2 amino acid sequence indicates that CMB-A2 is a novel *G. thermodenitrificans* strain, closely related to strain NG80-2 identified recently in a deep oil reservoir from Northern China (*12*). The isolation and recovery of CMB-A2 from compost 'grown' in a part of the world devoid of any oil extraction activity (Mauricie region, Quebec) corroborates the hypothesis that NG80-2 originated from a soil environment. NG80-2 would have gained the capacity for alkane oxidization using plasmid pLW1071 in an oil contaminated soil before invading the oil reservoir (*13*). It will be interesting to investigate the potential ability of CMB-A2 to process longer chain alkanes (i.e. C36) and verify the presence of plasmid similar to pLW1071.

Conclusion

This study allowed identifying a novel carboxylesterase EstGtA2 from *G. thermodenitrificans* which have potential in various industrial applications considering its wide pH range and its thermostability (65°C). This study also suggests the evidence of the existence of a new family of carboxylesterases, tentatively named N and N′. Further work on the correlation between sequence features and enzymatic properties such as thermostability in EstGtA2 and other esterases and lipases from organisms related to the *Geobacillus* genus is in progress. Our ultimate goal is to develop an understanding of the determinants of properties and stability in this focused group of enzymes.

Acknowledgements

D.M.C. acknowledges a scholarship from the NSERC.

Funding
Natural Sciences and Engineering Research Council of Canada (NSERC).

Conflict of interest
None declared.

References

1. Okuda, H. (1991) Esterase in *A Study of Enzymes* (Kuby, S.A., ed.) Vol. 1, pp. 563–577, CRC press, Boca Raton, Florida
2. Verger, R. (1997) Interfacial activation of lipases: facts and artifacts. *Trends Biotechnol.* **15**, 32–38
3. Jaeger, K-E., Ransac, S., Dijkstra, B.W., Colson, C., Heuvel, M., and van Misset, O. (1994) Bacterial lipases. *FEMS Microbiol. Rev.* **15**, 29–63
4. Liu, P., Wang, Y.F., Ewis, H.E., Abdelal, A., Lu, C.D., and Weber, I.T. (2003) Crystallization and preliminary X-ray diffraction data for the carboxylesterase Est30 from *Bacillus stearothermophilus*. *Acta. Cryt.* **D59**, 1472–1473
5. Liu, P., Wang, Y.F., Ewis, H.E., Abdelal, A., Lu, C.D., Harrison, R.W., and Weber, I.T. (2004) Covalent reaction intermediate revealed in crystal structure of the *Geobacillus stearothermophilus* carboxylesterase Est30. *J. Mol. Biol.* **342**, 551–561
6. Bornscheuer, U.T. (2002) Microbial carboxyl esterases: classification, properties and applications in biocatalysis. *FEMS Microbiol. Rev.* **26**, 73–81
7. Gupta, M.N. and Roy, I. (2004) Enzymes in organic media. Forms, functions and applications. *Eur. J. Biochem.* **271**, 2575–2583
8. Ollis, D.L., Cheach, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., Sussman, J.L., Verschueren, K.H.G., and Goldman, A. (1992) The α/β hydrolase fold. *Protein Engin.* **5**, 197–211
9. Arpigny, J.L. and Jaeger, K-H. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem. J.* **343**, 177–183
10. Wilson, K. (1987) Preparation of genomic DNA from bacteria in Current protocol in *Molecular Biology* (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J. A., Seidman, J. G., and Struhl, K., eds.), pp. 2.4.1–2.4.5, Wiley, New York
11. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A laboratory Manual*., p. 134, Cold Spring Harbor Lab. Press, Cold Spring Harbor, New York
12. Wang, L., Tang, Y., Wang, S., Liu, R-L., Liu, M-Z., Zhang, Y., Liang, F-L., and Feng, L. (2006) Isolation and characterization of a novel thermophilic *Bacillus* strain degrading long-chain *n*-alkanes. *Extremophiles* **10**, 347–356
13. Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., Tang, Y., Liu, X., Han, W., Peng, X., Liu, R., and Wang, L. (2006) Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc. Natl Acad. Sci. USA* **104**, 5602–5607
14. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.M. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9
15. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2007) GenBank. *Nucleic Acids Res.* **36**, D25–D30
16. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402
17. Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., and Bryant, S.H. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* **30**, 281–283
18. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599

19. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**, 4673–4680
20. Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. **4**, 406–425
21. Swofford, D.L. and Sullivan, J. (2002) Phylogeny inference based on parsimony and other methods using PAUP in *The Phylogenetic Handbook, A Practical Approach to DNA and Protein Phylogeny* (Salemi, M. and Vandamme, A-M., eds.) Vol. 1, pp. 160–202, Cambridge University press, Cambridge
22. Lambert, C., Léonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of protein 3D structures. *Bioinformatics* **18**, 1250–1256
23. Hooft, R.W.W., Vriend, G., Sander, C., and Abola, E.E. (1996) Errors in protein structures. *Nature* **381**, 272
24. Laskowski, R.A., MacArthur, M.W., Moss, D., and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst*. **26**, 283–291
25. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201
26. Constantini, S., Colonna, G., and Facchianno, A.M. (2008) ESBRI: a web interface for evaluating slat bridges in proteins. *Bioinformation* **1**, 137–138
27. Smith, P.K., Krohn, R.I., Hermanson, G.T., Mailla, A.K., Gartner, F.H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., and Klenk, D.C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem*. **150**, 76–85
28. Imamura, S. and Kitaura, S. (2000) Purification and characterization of a monoacylglycerol lipase from the moderately thermophilic *Bacillus* sp. H-257. *J. Biochem*. **127**, 419–425
29. Kitaura, S., Suzuki, K., and Imamura, S. (2001) Monoacylglycerol lipase from moderately thermophilic *Bacillus* sp. strain H-257: molecular cloning, sequencing and expression in *Escherichia coli* of the gene. *J. Biochem*. **129**, 397–402
30. Montoro-Garcia, S., Martinez-Martinez, I., Nevarro-Fernadéz, J., Takami, H., Garcia-Carmona, F., and Sanchez-Ferrer, A. (2009) Characterization of a novel thermostable carboxylesterase from *Geobacillus kaustophilus* HTA426 shows the existence of a new carboxylesterse family. *J. Bacteriol*. **191**, 3076–3085
31. Ewis, H.E., Abdelal, A.T., and Lu, C.D. (2004) Molecular cloning and characterization of two thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *GENE* **329**, 187–195
32. Takano, K., Tsuchimori, K., Yamagata, Y., and Yutani, K. (2000) Contribution of salt bridges near the surface of a protein to the conformational stability. *Biochemistry* **39**, 12375–12381
33. Strop, P. and Mayo, S.L. (2000) Contribution of salt bridges to protein stability. *Biochemistry* **39**, 1251–1255
34. Kumar, S. and Nussinov, R. (1999) Salt bridge stability in monomeric proteins. *J. Mol. Biol*. **293**, 1241–1255