US010385327B2

(12) **United States Patent**
Baerga-Ortiz et al.

(10) **Patent No.: US 10,385,327 B2**
(45) **Date of Patent: Aug. 20, 2019**

(54) **ENHANCED *E. COLI* FOR THE PRODUCTION OF FATTY ACIDS AND METHOD OF PRODUCING THE SAME**

(75) Inventors: **Abel Baerga-Ortiz**, San Juan, PR (US); **Delise Oyola-Robles**, San Juan, PR (US)

(73) Assignee: **University of Puerto Rico**, San Juan, PR (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **13/344,062**

(22) Filed: **Jan. 5, 2012**

(65) **Prior Publication Data**

US 2013/0177989 A1 Jul. 11, 2013

(51) **Int. Cl.**
C12N 9/88 (2006.01)
C12N 15/70 (2006.01)

(52) **U.S. Cl.**
CPC ***C12N 9/88*** (2013.01); ***C12N 15/70*** (2013.01)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2005/0100995 A1* 5/2005 Weaver et al. 435/134
2007/0054358 A1* 3/2007 Blattner et al. 435/69.1

OTHER PUBLICATIONS

Udwary et al. 2002 (A Method for Prediction of the Locations of Linker Regions within Large Multifunctional Proteins, and Application to a Type I Polyketide Synthase; JMB, 323:585-598).*
Allen et al. 2002 (Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium *Photobacterium profundum* strain SS9; Microbiology; 148:1903-1913).*
Oyola-Robles et al. 2013 (Identification of novel protein domains required for the expression of an active dehydratase fragment from a polyunsaturated fatty acid synthase; Protein Science 22:954-963).*
Oyola-Robles et al. 2014 (Expression of dehydratase domains from a polyunsaturated fatty acid syntase increase the production of fatty acids in *Escherichia coli*; Enzyme and Microbial Technology, 55:133-139).*

* cited by examiner

Primary Examiner — Mary Maille Lyons
(74) *Attorney, Agent, or Firm* — Hoglund & Pamias, PSC; Roberto J. Rios

(57) **ABSTRACT**

The invention analyzed a protein sequence using the Udwary-Merski algorithm and identified a tetradomain fragment (DH1-DH2-UMA) which consists of two predicted DH-like domains and two pseudodomains N-terminal to them. This arrangement of domains and pseudodomains is fundamentally the opposite of what is typically observed in the DH cassettes of actinobacterial polyketide synthases or mammalian fatty acid synthases, both of which feature C-terminal pseudodomains. The invention modified *E. coli* by over expressing DH1-DH2-UMA in *E. coli* resulting in an increase in the overall production of all the fatty acids normally present in the *E. coli* fatty acid profile.

7 Claims, 6 Drawing Sheets

Specification includes a Sequence Listing.

Figure 1

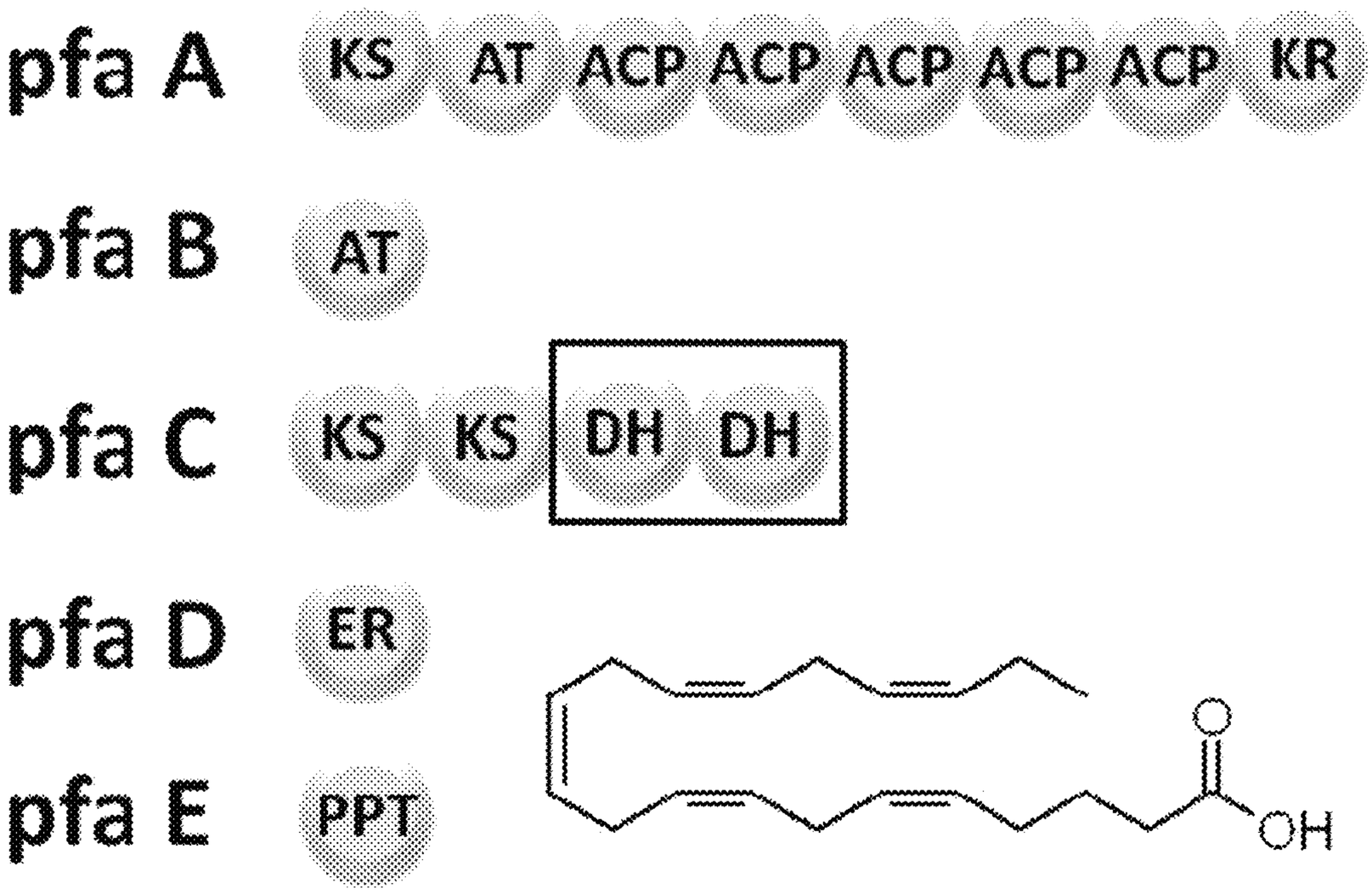

pfa A
KS
AT
ACP
ACP
ACP
ACP
ACP
KR
pfa B
AT
pfa C
KS
KS
DH
DH
pfa D
ER
pfa E
PPT
OH

Figure 2

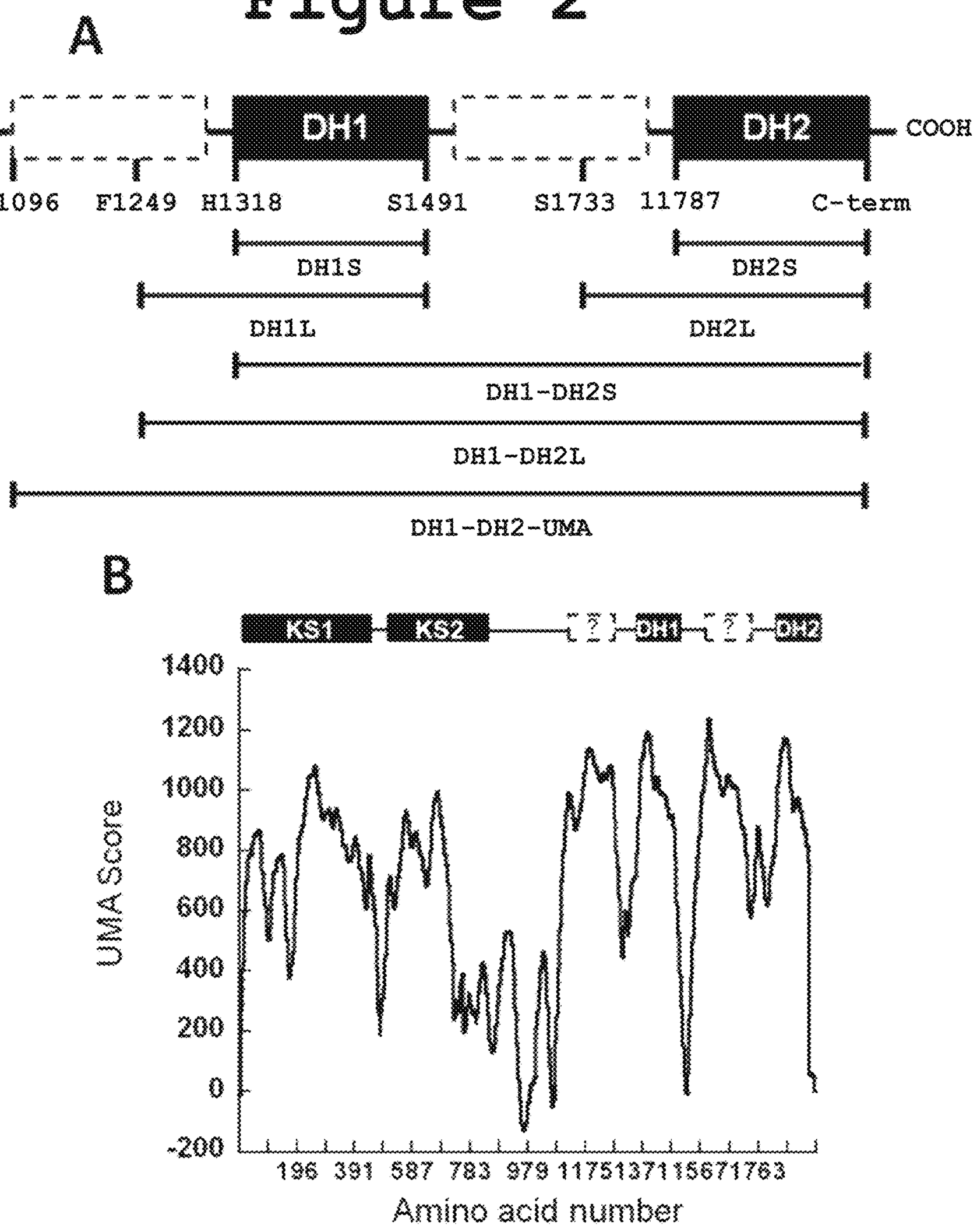

A
DH1
DH2
COOH
I1096
F1249
H1318
S1491
S1733
I1787
C-term
DH1S
DH2S
DH1L
DH2L
DH1-DH2S
DH1-DH2L
DH1-DH2-UMA
B
KS1
KS2
?
DH1
?
DH2
UMA Score
1400
1200
1000
800
600
400
200
0
-200
196
391
587
783
979
1175
1371
1567
1763
Amino acid number

Figure 3

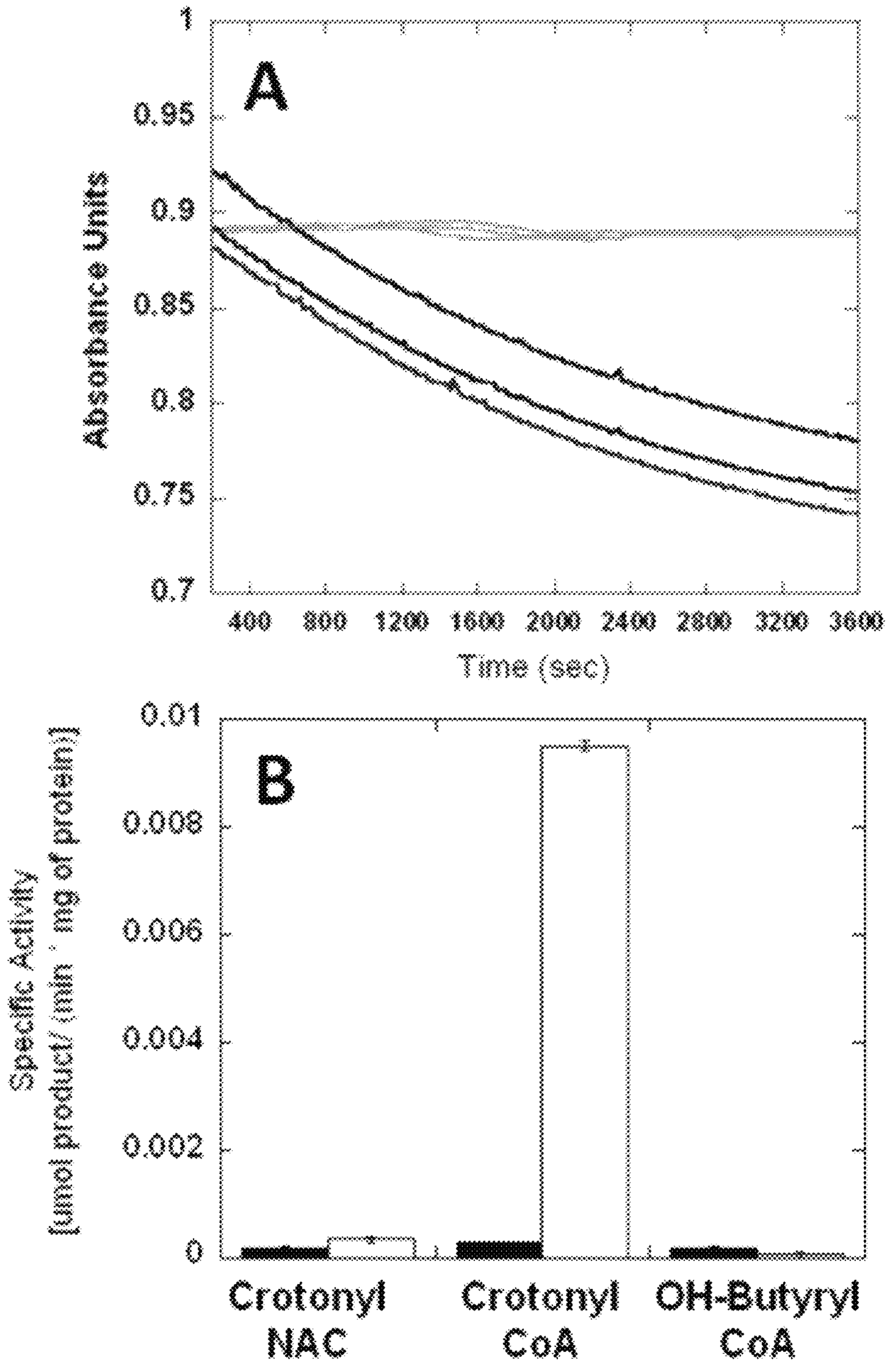

A
Absorbance Units
1
0.95
0.9
0.85
0.8
0.75
0.7
400
800
1200
1600
2000
2400
2800
3200
3600
Time (sec)
B
Specific Activity
[umol product/ (min * mg of protein)]
0.01
0.008
0.006
0.004
0.002
0
Crotonyl NAC
Crotonyl CoA
OH-Butyryl CoA

Figure 4

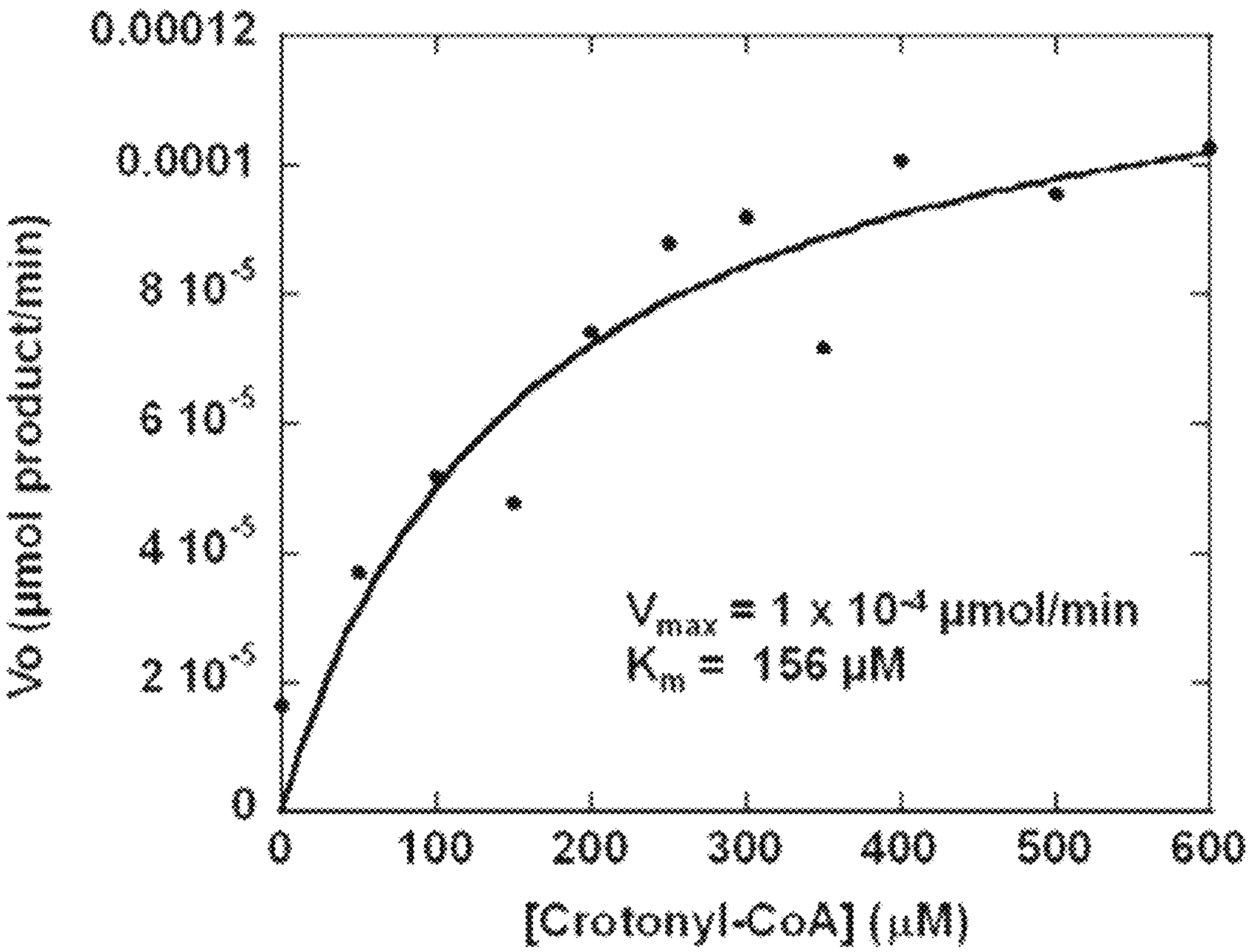

0.00012
0.0001
8 10-5
6 10-5
4 10-5
2 10-5
0
0
100
200
300
400
500
600
[Crotonyl-CoA] (μM)
Vmax = 1 x 10-4 μmol/min
Km = 156 μM

Figure 5

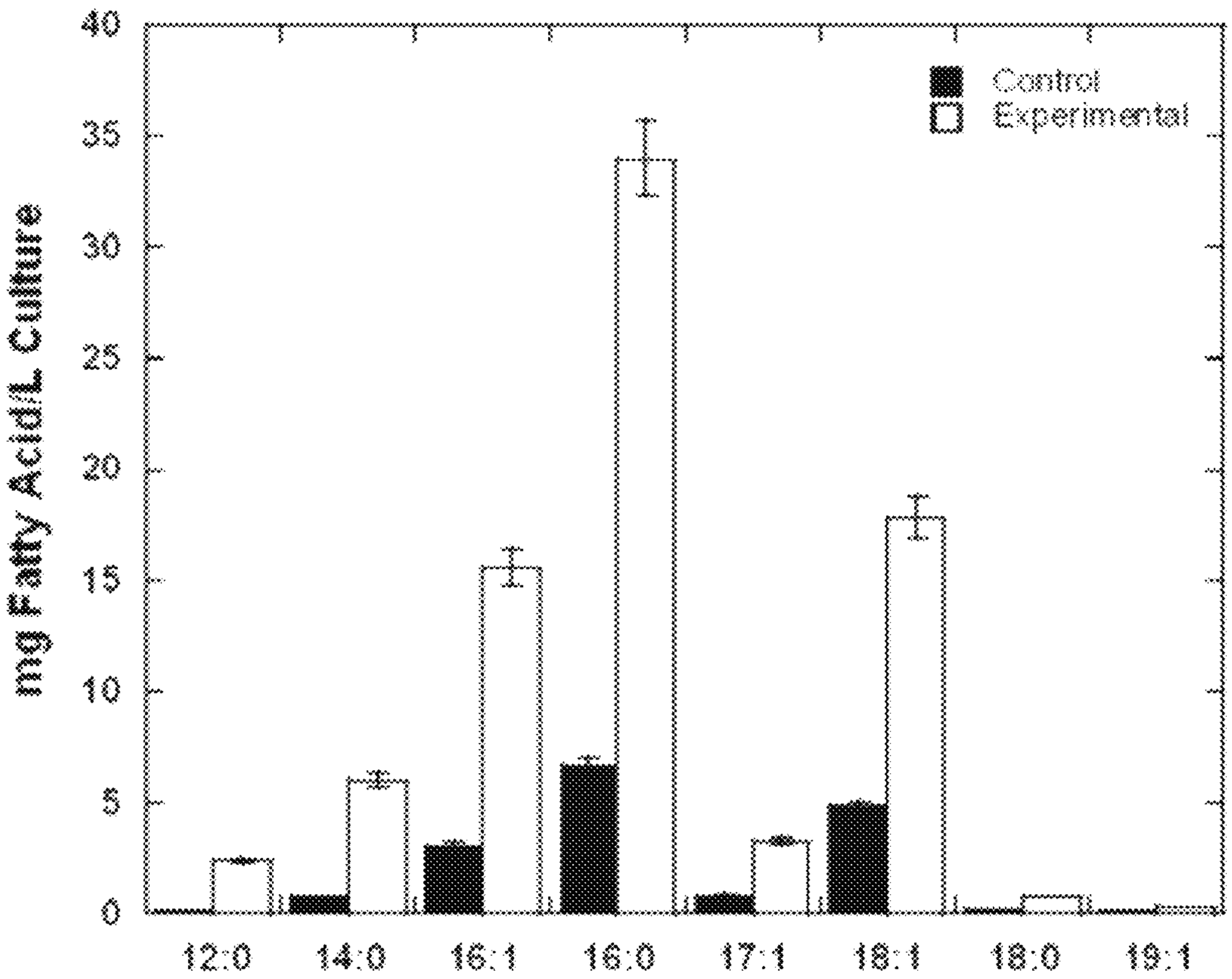
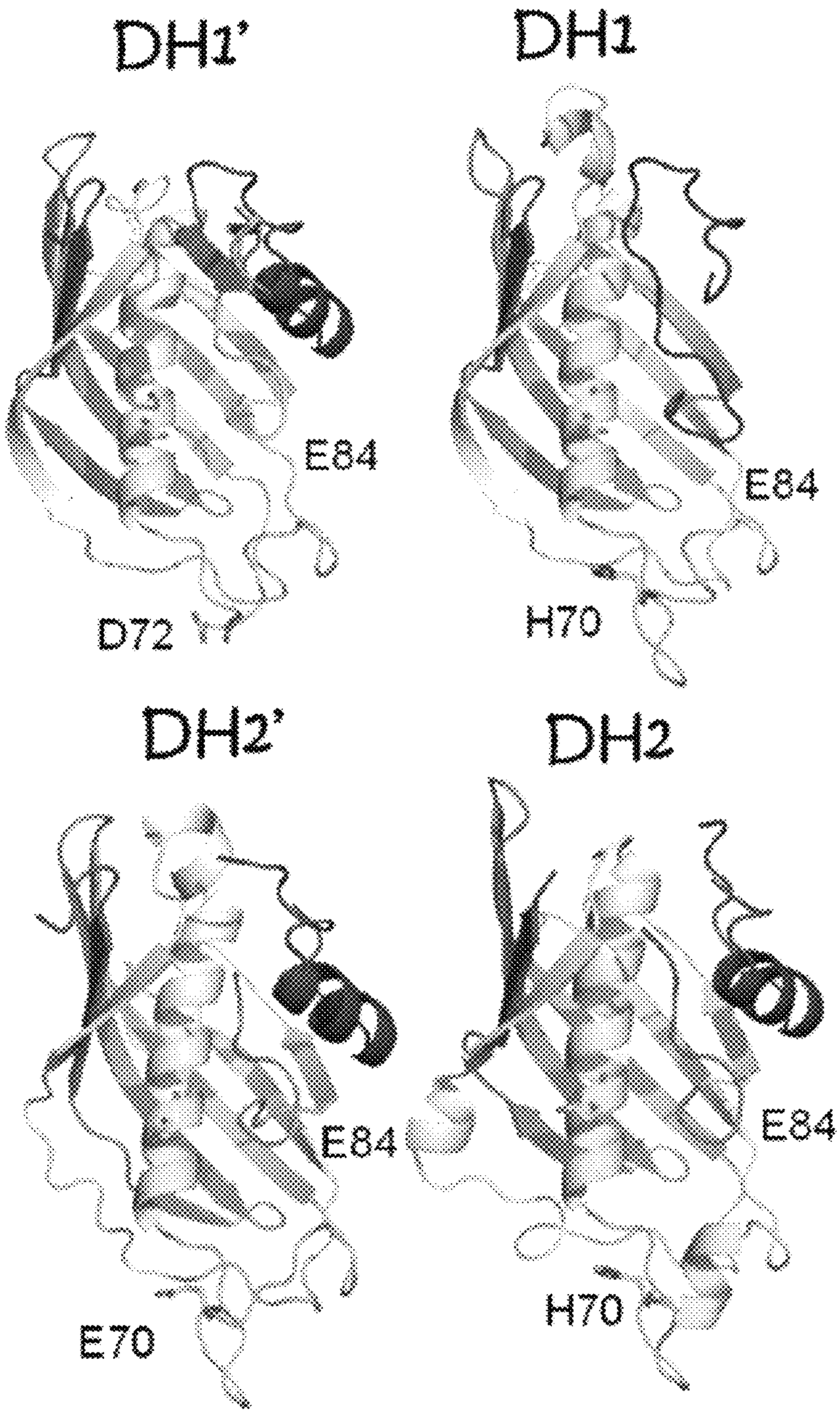

mg Fatty Acid/L Culture
40
35
30
25
20
15
10
5
0
Control
Experimental
12:0
14:0
16:1
16:0
17:1
18:1
18:0
19:1

Figure 6



ENHANCED *E. COLI* FOR THE PRODUCTION OF FATTY ACIDS AND METHOD OF PRODUCING THE SAME

GOVERNMENT INTEREST

The claimed invention was made with U.S. Government support under grant number CHE-0953254 awarded by the US National Science Foundation (NSF). The government has certain rights in this invention.

SEQUENCE LISTING

The sequence listing submitted via EFS, in compliance with 37 CFR § 1.52(e) (5), is incorporated herein by reference. The sequence listing text file submitted via EFS contains the file "Sequence Listing 13344062", created on Jun. 25, 2012, which is 32,350 bytes in size.

BACKGROUND OF THE INVENTION

Long-chain polyunsaturated fatty acids (PUFAs) have been implicated in human brain development as well as in the maintenance of cardiovascular health. Although animals have the enzymes necessary to form long-chain PUFAs through the elongation of plant-derived PUFAs, this oxygen-dependent process is not efficient. An efficient pathway for the biosynthesis of PUFAs in deep-sea bacteria utilizes a polyketide synthase-like (PKS-like) multienzyme complex. A total of five genes from this pathway have been found to be sufficient for the production of polyunsaturated fatty acids in an otherwise non-producing *Escherichia coli*. These genes are pfaA, pfaB, pfaC, pfaD, encoding PUFA synthases containing enzyme domains for acyl tranferases (AT), keto-acyl synthase (KS), acyl carrier protein (ACP), keto-acyl reductase (KR), enoyl reductase (ER) and dehydratase (DH) activities and also pfaE, which encodes a required phospho-pantetheine transferase (PPTase) essential for the activation of ACP domains through chemical modification as shown in FIG. **1**. While some of the required enzymatic activities are housed in independent stand-alone proteins (pfaB, pfaD and pfaE: FIG. **1**) others are assembled into multidomains (pfaA and pfaC: FIG. **1**). No thioesterase activity has been observed in the PUFA synthase cluster and no dedicated thioesterase protein from the producing organism is required for heterologous production of PUFAs in *E. coli*.

Dehydratase (DH) domains are responsible for the formation of the cis double bonds in the structure of PUFAs. They can be easily identified by their sequence similarity to FabA and FabZ, the two DH enzymes involved in fatty acid biosynthesis in *E. coli*. FabA/Z catalyze the dehydration of 3Rhydroxyacyl-ACP via a syn elimination mechanism which has also been reported in the DH domain from the erythromycin PKS.

The structure of FabA, and more recently FabZ, revealed an obligate homodimeric arrangement in which both DH subunits contribute key residues to the active site. This distinct architectural feature has been found to extend to DH domains from the animal Fatty Acid Synthase (FAS), and more recently to the erythromycin PKS, although with the following variation on the *E. coli* arrangement. While the *E. coli* FabA and FabZ form homodimers of identical subunits, the DH domains from FAS and PKS systems form a heterodimeric double hotdog arrangement in which two contiguous pseudosubunits are housed within the same polypeptide and separated by a 25-residue amino acid stretch. Thus, the required dimerization of the DH domain in the context of a multienzyme complex does not necessarily involve interactions between different polypeptides, but rather within the same polypeptide.

In both the FAS and PKS DH, the protein region that is homologous to FabA is followed by a necessary C-terminal pseudodomain with no previously known function and no known sequence homologue. In the case of the FAS DH, the C-terminal pseudodomain was found to contribute to dehydratase activity in in vitro enzyme assays. The structure of the PKS DH showed that the Cterminal pseudodomain forms the other half of the double hotdog in the three-dimensional structure. In that work, the protein construct that was crystallized, and whose structure was determined, contained the pseudodomain but lacked dehydratase activity in vitro, although mutations made elsewhere did show an effect on overall polyketide production by the full-length multienzyme.

The PUFA synthase multienzyme contains two putative DH domains in tandem. They have been identified as DH domains based on their sequence similarity to FabA/Z, but their activity or specificity has not been confirmed biochemically. The tandem arrangement, while not previously observed in other biosynthetic enzyme systems, is a well-conserved feature of PUFA synthases. However, it is unknown how these tandem domains act to generate the combination of double and single C—C bonds in the final PUFA structure.

SUMMARY OF THE INVENTION

According to an aspect of the invention, a protein fragment consisting of the two tandem putative DH domains and the two corresponding pseudodomains from the PUFA synthase was designed using the Udwary-Merski Algorithm (UMA) developed at Johns Hopkins University.

According to another aspect of the invention, the resulting tetradomain fragment showed some dehydratase activity against an acyl-CoA soluble substrate. Examination of the three dimensional models for the individual domains reveal that while two domains contain all the conserved residues expected for a functional DH domain, the other two domains contain other residues present on other hot-dog proteins.

According to still another aspect of the invention, the analysis of the tetradomain sequence anticipates an "inverted" double hotdog arrangement in which the pseudodomain is actually located N-terminal to the FabA homology domain, thus providing an alternative topological solution which suggests evolutionary convergence of the DH architecture in PUFA synthase multienzymes.

BRIEF DESCRIPTION OF THE DRAWINGS

Further features and advantages of the invention will become apparent from the following detailed description taken in conjunction with the accompanying figures showing illustrative embodiments of the invention, in which:

FIG. **1** shows a gene cluster for the anaerobic production of eicosapentaenoic acid (EPA) in *Photobacterium profundum* according to the present invention.

FIG. **2** shows a scheme summarizing the construction of different protein fragments for the isolation of dehydratase activity according to the present invention.

FIG. **3** shows graphs for Dehydratase activity of DH1-DH2-UMA towards crotonyl-CoA and the specific activity of DH1-DH2-UMA toward crotonyl-NAC in the hydration reaction and towards β-hydroxybutyryl-CoA in the dehydration reaction according to the present invention.

FIG. **4** shows a saturation curve obtained by measuring the activity of DH1-DH2-UMA towards crotonyl-CoA at 235 nm according to the present invention.

FIG. **5** shows a graph illustrating the over-expression of DH1-DH2-UMA in *E. coli* resulting in an increase in the total production of free fatty acids in liquid bacterial culture according to the present invention.

FIG. **6** illustrates a comparison of the three-dimensional models obtained for the FabA-homology regions (DH1 and DH2) and for the uncharacterized pseudodomains (DH1.' and DH2.') according to the present invention.

Throughout the figures, the same reference numbers and characters, unless otherwise stated, are used to denote like elements, components, portions or features of the illustrated embodiments. The subject invention will be described in detail in conjunction with the accompanying figures, in view of the illustrative embodiments.

DETAILED DESCRIPTION OF THE INVENTION

Experimental Procedures

Cloning, Expression and Purification

Different DH fragments were cloned from fosmid 8E1. All restriction endonucleases, polynucleotide kinase, T4 DNA ligase, and alkaline phosphatase were purchased. The primers used to make the different fragments are summarized in Table 1 below. For cloning into pGEX4T-3 vector, the amplified DNA was phosphorylated using polynucleotide kinase and cloned into pUC19 which was previously digested with SmaI and treated with alkaline phosphatase. The ligation mixture was used to transform DH10B cells and clones were selected in LB-agar containing ampicillin (100 μg/mL). Insertion of the DH fragment into pUC19 was confirmed by agarose gel electrophoresis. The resulting plasmid pUC19:DH was digested with BamHI and SmaI and the resulting excised DNA fragment was cloned into the corresponding sites in pGEX4T-3.

TABLE 1

Dehydratase construct	Oligonucleotide sequence
DH1S	(SEQ.ID1) - Fwd: 5'-CATGCATGGGATCCAACTTGCTAGACGCAAATATCGCA-3'
	(SEQ.ID2) - Rv: 5'-CATGCATGCCCGGGTCATGATTCTTCTTTGATCATCACG-3'
DH1L	(SEQ.ID3) - Fwd: 5'-CACCTTCTCTTACGAATGTTTCGTTGGC-3'
	(SEQ.ID4) - Rv: 5'-CATGCATGCCCGGGTCATGATTCTTCTTTGATCATCACG-3'
DH2S	(SEQ.ID5) - Fwd: 5'-CATGCATGGGATCCAACTTACTGGATAAAGAAAGCCGTT-3'
	(SEQ.ID6) - Rv: 5'-TCAGGCTTCTTCAATACAGATTGC-3'
DH2L	(SEQ.ID7) - Fwd: 5'-CATGCATGGGATCCTTCAGCTTCGAACTCAGTACCGA-3'
	(SEQ.ID8) - Rv: 5'-TCAGGCTTCTTCAATACAGATTGC-3'
DH1-DH2-S	(SEQ.ID9) - Fwd: 5'-CACCAACTTGCTAGACGCAAATATCGCA-3'
	(SEQ.ID10) - Rv: 5'-TCAGGCTTCTTCAATACAGATTGC-3'
DH1-DH2-L	(SEQ.ID11) - Fwd: 5'-CACCTTCTCTTACGAATGTTTCGTTGGC-3'
	(SEQ.ID12) - Rv: 5'-TCAGGCTTCTTCAATACAGATTGC-3'
DH1-DH2 UMA	(SEQ.ID13) - Fwd: 5'-CACCCGCAAACCTTGTATCTGGGATTA-3'
	(SEQ.ID14) - Rv: 5'-TCAGGCTTCTTCAATACAGATTGC-3'

For the cloning of fragments into pET200TOPO SEQ.ID15, the amplified DNA was gel purified using a Gel Extraction Kit and incubated with pET200TOPO. The resulting clones were selected in LB-agar containing kanamycin (100 μg/mL). All resistant clones were introduced into *E. coli* strain SEQ.ID16 BL21-DE3-Codon Plus-RIL and grown in liquid LB at 37° C. until the OD600=0.4 at which time the temperature was decreased to 22° C.° until the OD600=0.6 at which time protein expression was induced with 1 mM IPTG. After 16 h, the cells were collected and resuspended in lysis buffer (50 mM Na3HPO4 pH 7.2, 150 mM NaCl, 1 mM DTT, 10% glycerol, 0.1 mg/mL lysozyme and DNAse) for 1 hr, sonicated and centrifuged at a speed of 14,000 rpm at 4° C. for 30 min in a J2-21 Beckman centrifuge in a JA17 rotor. Samples were collected for the total, supernatant and pellet to assess solubility of the protein products.

For His-tagged soluble proteins, the lysate was collected and poured through a column filled with Ni-NTA resin (Qiagen) equilibrated in 25.0 mM Tris pH 8.0, 150 mM NaCl, 10% glycerol, 1.0 mM DTT. The DH fragment was eluted with the same buffer containing 300 mM imidazole.

Eluted protein was infused into an ion exchanger column operated at room temperature and equilibrated in 25 mM Tris pH 8.0, 150 mM NaCl, 1.0 mM DTT and 10% glycerol. The proteins were eluted in a 40-minute gradient 0.15 M-2 M NaCl. The fraction containing the protein was concentrated and stored at −80° C. Typical yields for all proteins were 1.0 mg of protein per liter of culture, purity ~99% by 8% SDS-PAGE.

UMA Parameters.

The UMA program was used and UMA calculations were done as in (Udwary et al., 2002) using the sequence of SEQ.ID17 pfaC from *Photobacterium profundum* (GenBank Accession no. AF409100.1). A multiple alignment of homologues of pfaC was performed in CLUSTALW in ".pir" format and a secondary structure prediction for pfaC was performed using the PSIPRED Server (University College London). The output for the secondary structure prediction was used to generate an ".ss" file. Finally, both the "pir." alignment and the "ss." secondary structure prediction were used as inputs for the "uma19.pl" application with the input parameters in Table 2 below. Results in the output file were visualized using Keleidagraph for Windows.

TABLE 2

Parameter	Value
Homology matrix	blosum 30
Gap to gap penalty	0
Gap to aa penalty	−4
Component averaging (k)	5
Final averaging (gamma)	20
Sim score weight	10
Struc score weight	1
Hydro score weight	5

Dehydratase Assays.

Dehydratase activity was measured in a hydration assay by using Crotonyl-CoA and Crotonyl-NAC as substrates. Crotonyl-NAC was synthesized from crotonic acid and N-acetylcysteamine using a DCC coupling strategy as describes by the prior art and purified by flash column chromatography on silica gel using 1:1 ethyl acetate:ethyl ether. For the dehydration assay β-hydroxybutyryl-CoA was used as the substrate. Enzymatic reactions were followed spectrophotometrically by monitoring the absorbance at 260 nm in a 96-well plate format on. The total volume was 200 μL (25 mM Tris, 150 mM NaCl, 10% glycerol, pH 8.0, 3.20 μM DH1-DH2-UMA, and 117 μM of substrate). The values for the absorbance slope (given in mAU/min) were converted to units of μmole of product per minute by using the following equation:

$$\mu\text{mole of product/min}=[\text{Slope}/\varepsilon\times b]\times \text{Vol}_{total} \quad \text{(Eq1)}$$

in which the slope is given by the instrument in units of milliabsorbance (mAU) per minute, b is the path length measured to be 0.89 cm for a Vol_{total}=200 μl in our 96-well plates. The ε is the molar extinction coefficient resulting from the loss of a double bond as defined by the difference in absorbance between crotonyl-CoA and β-hydroxybutyryl-CoA at a particular wavelength. The extinction coefficient was calculated to be ε=969.9 M^{-1} cm^{-1} for the reaction monitored at 260 nm and ε=790.7 M^{-1} cm^{-1} for the reaction monitored at 235 nm.

For the kinetic assays the reaction was monitored at a wavelength of 235 nm and using a range of substrate concentrations between 0 and 600 μM. The data was fit to a simple Michaelis-Menten Equation (Eq 2) using Kaleidagraph v4.03.

$$V_o=V\max[S]/([S]+K_m) \quad \text{(Eq2)}$$

Fatty Acid Profiles.

E. coli BL21-DE3-CodonPlus (RIL) SEQ.ID16 cells expressing DH1-DH2-UMA SEQ.ID14 in the pET200Topo vector were cultured in LB media and the expression was induced as described for protein production. Protein expression was confirmed by SDS-PAGE. Cells were collected by centrifugation at 4,400 rpm, 10 min, 4° C. and freeze-dried. The fatty acid components of the cell culture were obtained as their methyl esters by the reaction of 0.05 g of dried cell pellet with 10.0 mL of methanolic HCl, refluxed for 2 hr followed by workup with hexane twice. The organic layer was dried over $MgSO_4$ and concentrated in vacuo. The fatty acid methyl esters were analyzed by GC-MS (at 70) equipped with a 30 m×0.25 mm special performance capillary column (HP-5MS) of polymethylsiloxane crossed-linked with 5% phenyl methylpolysiloxane. The temperature program was as follows: 130° C. for one minute, increase at a rate of 3° C./min to a 270° C., where the temperature is maintained for 30 min. Methyl heneicosanoate was used as an internal standard for quantification of fatty acid methyl esters.

Results

Design and Expression of Putative DH Domains from the PUFA Synthase.

The pfaC protein of the PUFA synthase complex harbors two homologues of FabA/Z dehydratases as shown in FIG. **1**. Initially, a number of protein constructs were designed on the basis of FabA homology and sequence conservation alone as summarized in FIG. **2**A. Two "short" fragments, DH1S (H1318-51491) and DH2S (I1787-C-term) were designed to include only the sequence homologous to FabA. Two "longer" fragments, DH1L (F1249-51491) and DH2L (S1733-C-term), were designed to include additional conserved sequence N-terminal to the FabA-homology region. Finally, to explore the possibility that the two FabA-homology regions stabilized one another, we also generated protein fragments which contained both Fab-homology regions, DH1-DH2S (H1318-C-Term) and DH1-DH2L (F1249-C-Term). All of these protein fragments were expressed as GST fusion proteins and as His-tagged proteins in *E. coli* and all were found to be insoluble as evidenced by their presence in the lysis pellet (data not shown).

In order to more accurately define the boundaries for the putative DH domains from pfaC so as to increase the likelihood of generating a functional enzyme fragment, we analyzed the sequence using the Udwary-Merski Algorithm (UMA) which assigns a numerical score to each amino acid based on the probability that it is located within a structured domain, as opposed to it being located in an unstructured linker region. UMA analysis of the pfaC sequence revealed six domain regions as defined by their high UMA score as shown in FIG. **2**B. Four of the six domains had been previously identified based on sequence alignments: two KS domains in the N-terminal portion of the protein (KS1 and KS2), and the two FabA homologs (DH1 and DH2). The two other areas of high UMA score were located directly N-terminal to the putative DH domains. The predicted secondary structure for the two new pseudodomains (termed DH1' and DH2') was that of a hot-dog fold, much like the predicted secondary structure for the Fabhomology domains.

Based on the UMA analysis and on the secondary structure prediction, fragment DH1-DH2-UMA (I1096-N-Term) was designed and expressed as a His-tagged protein in soluble form. After nickel resin purification and anion exchange chromatography, a total yield of 1.0 mg of pure protein was obtained per liter of culture. Gel filtration chromatography of this protein revealed an equilibrium between a monomer and a dimer in equal proportions (data not shown).

Preliminary Activity of DH1-DH2-UMA.

Incubation of DH1-DH2-UMA with crotonyl-CoA resulted in a decrease in the absorbance at 260 nm, consistent with the hydration of the double bond as shown in FIG. **3**. The N-acetyl cysteamine (NAC) thioester of crotonic acid was not hydrated suggesting the importance of the pantetheine carrier for substrate recognition. The dehydration of the β-hydroxybutyryl-CoA was also monitored but no activity was detected in the forward reaction, probably due to the fact that the chemical equilibrium favors the reverse reaction. Initial efforts to measure the Michaelis-Menten kinetic parameters were frustrated by the fact that the amount of substrate required to saturate the enzyme was too high for spectrophotometric determination at a wavelength of 260 nm. In order to lower the absorption intensity of the acyl-CoA substrate, the reaction was monitored at a wavelength of 235 nm and the initial velocity was measured at different substrate concentration. From the saturation curve for crotonyl-CoA, the kinetic parameters were extracted (FIG. **4**:

Vmax=0.0001 μmol product/min; Km=156 μM). The measured activity towards NAC-loaded substrates or β-hydroxybutyryl-CoA was too low to yield a reliable saturation curve for the determination of the kinetic parameters.

Effect of DH1-DH2-UMA Overexpression on the Fatty Acid Profile of *E. coli.*

The overexpression of enzymes has been employed as a strategy to enhance fatty acid production or to alter the normal fatty acid profile of *E. coli*. In order to investigate whether DH1-DH2-UMA would interact with the fatty acid biosynthesis machinery of *E. coli* and result in the formation of polyunsaturated fatty acids, we measured the production of fatty acids in a strain overexpressing DH1-DH2-UMA. No polyunsaturated fatty acids were detected in any of the bacterial extracts, indicating that the expression of DH1-DH2-UMA is not sufficient to catalyze the formation of multiple cis double bonds in the fatty acids normally made by *E. coli*. It was observed, however, a 4-fold to 5-fold increase in the total production of free saturated and monounsaturated fatty acids without a change in the percentage composition of fatty acids as shown in FIG. **5**. The fact that the expression of DH1-DH2-UMA affected the production of all fatty acids in equal proportions suggests that the protein is capable of interacting with the *E. coli* machinery for fatty acid biosynthesis in a way that does not discriminate based on fatty acid chain length.

Three-Dimensional Models of DH Domains and Pseudodomains.

In order to verify the presence of amino acid residues normally associated with dehydratase activity, we built three-dimensional models of all domains and pseudo-domains using the Phyre Server from Imperial College London as shown in FIG. **6**. The 3D models generated for the actual Fab-homology domains (DH1 and DH2) feature an active site His70 and the conserved Glu84 typical of dehydratases (amino acid sequence numbers are based on the FabA numbering). Interestingly, even though the newly identified N-terminal pseudodomains (DH1' and DH2') do not have a high enough sequence similarity with any known protein, their secondary structure prediction in the Phyre server was found to be consistent with the formation of a hotdog fold, possibly the first half of a double hotdog. Instead of the expected His70 conserved in dehydratases, DH1' featured an Asp72 and DH2' featured a Glu70 in the corresponding region as shown in FIG. **6**. These acidic residues in the active site are not typically observed in the hotdog dehydratases but they are a defining feature of the hotdog hydrolases, suggesting a possible involvement of DH1' and DH2' in hydrolysis.

Discussion

The biosynthesis of PUFAs in deep-sea bacteria is carried by a family of enzymes that contain a unique and conserved arrangement of enzyme domains. PUFA synthases have been found in metagenomic DNA from marine samples collected throughout the world, indicating that anaerobic PUFA biosynthesis is a widely selected mechanism for microbial adaptation to high-pressure and low temperature environments. Despite much interest in elucidating how the PUFA synthase carries out its function, published work on the enzymatic activities of PUFA synthases has been sparse. Bumpus et al., 2008 showed for the first time the in vitro activity of the enoyl reductase (pfaD) enzyme from *Shewanella oneidensis* PUFA synthase and Jiang et al., 2008 interrogated the role of the tandem ACP arrangement, which is a hallmark of PUFA synthases. The present invention addressed another conserved feature of PUFA synthases, a pair of conserved DH domains arranged in tandem near the C-terminus of the multidomain protein, pfaC.

Analysis of the sequence of pfaC protein using the Udwary-Merski Algorithm revealed the presence of two new pseudodomains located directly N-terminal to the regions of FabA homology. These pseudodomains were found to be essential for the proper expression of protein fragments, since only the protein fragments that included both pseudodomains were soluble, stable and active. This result alone would suggest that DH' pseudodomains are important components of the three-dimensional structure of dehydratase domains. This finding also confirms the general applicability of the Udwary-Merski Algorithm for the identification of functional units within multidomain proteins with unknown functions or from unexplored lineages.

The predicted secondary structure for both DH' pseudodomains was that of a hotdog fold, which is also the expected three-dimensional topology of the FabA-homology DH domains. This predicted arrangement of contiguous hotdog folds points towards an overall double hotdog structure, which has become the widely accepted model for embedded dehydratases based on structural and biochemical evidence. However, several differences exist between the PUFA arrangement and its FAS and PKS evolutionary cousins. While in FAS/PKS DH, the pseudodomains are located C-terminal to the Fabhomology domain, in the PUFA DH, the pseudodomains are located N-terminal to the Fab-homology domain. This alternative gene structure of the PUFA DH suggests a tandem gene duplication event that took place independently in terrestrial FAS/PKS and marine PUFA synthase for the generation of functional DH dimers, resulting in two alternative convergent topological solutions. Another difference between FAS/PKS DH and PUFA DH is that, while FAS/PKS DH domains consist of didomains (one FabA homology domain plus one pseudodomain), the PUFA DH complex invariably consists of a tetradomain (two FabA homology plus two pseudodomains). This invention does not address the question of how the four protein domains are paired in the functional assembly. Additional structural characterization of DH1-DH2-UMA will have to be carried out in order to elucidate how the different domains are arranged in a functional complex.

Substantial work has been dedicated to determining the specific role of pseudodomains in the activity of FAS DH domains beyond stabilizing the dimeric structure by partnering with the FabA-homology domain. Amino acids in the DH pseudodomain have been implicated in the partial activity of the FAS ketoreductase domain. Additionally, an Asp residue in the FAS pseudodomain has been found to be essential and a Gln residue in the pseudodomain has been found to be important for dehydratase activity. In the PUFA DH in this report, multiple sequence alignment of the pseudodomains reveal levels of sequence conservation (67% and 71% for DH1' and DH2', respectively) that were comparable to the sequence conservation of the FabA homology domains (61% and 75% for DH1 and DH2, respectively). This high level of sequence similarity among the pseudodomains is suggestive of a role in DH function beyond that of a structural scaffold for dimerization.

The soluble DH1-DH2-UMA fragment was competent to catalyze the hydration of crotonyl-CoA with a specific activity of 0.009 μmol product/(min*mg enzyme). When this number is converted to the units of specific activity employed in Pasta et al., 2007, it becomes 0.83 mol product/(min*mol enzyme), at least two orders of magnitude lower than the specific activity reported for the FAS 1-1168 construct (204 mol product/(min*mol enzyme)). It has been

shown that dehydratase activity decreases dramatically with decreasing length of the acyl chain. Although that report does not include the activity toward crotonyl-ACP (3:1), the difference between the specific activity against octenoyl-ACP(8:1) and butenoyl-ACP (4:1) was about one order of magnitude. In addition a similar dramatic effect was observed when comparing ACP-linked substrate to pantetheine-linked substrates. The PUFA DH in this report was assayed for activity against crotonyl-CoA (3:1). Thus, it is not surprising that the specific activity is low considering that the acyl chain is even shorter that the shortest one in Pasta et al., 2007 and that the substrate in this report is not loaded on an ACP. Further work will need to be carried out to determine the substrate preference for PUFA DH domains in a more physiological context.

Additional confirmation of the activity of DH1-DH2-UMA came from measuring the effect of its overexpression on the production of fatty acids in *E. coli*. According to the invention, a significant increase in the production of fatty acids was observed in the BL21 *E. coli* strain expressing the DH1-DH2-UMA protein. Previous work by others has shown that overexpression of the *E. coli* FabA dehydratase does not increase the production of fatty acids in *E. coli*. Thus, it is hard to argue that the observed increase in fatty acid production in this report is due to the dehydratase activity of DH1-DH2-UMA although it cannot be entirely ruled out. It has been well established that the overexpression of thioesterases and other hydrolases results in the enhancement of the production of fatty acids and other high-energy biofuel precursors. Therefore, it is possible that an adventitious or unphysiological hydrolase activity, possibly an artifact arising from high enzyme concentration inside overexpressing bacterial cells, could be responsible for the observed enhancement of fatty acid production in *E. coli*.

Inspection of the homology model made for the DH' pseudodomains reveals a hotdog fold similar to that expected for the FabA-homology regions, although with a different amino acid occupying the active site position as shown in FIG. **6**. While the model for the FabA-homology region contains a His residue in position 70 and a Glu in position 84, consistent with dehydratase function, the homology model for the DH' pseudodomains reveals a Glu70 and a Glu84, which are more commonly found in hotdog hydrolases than in dehydratases. There have been reports of bona fide DH domains with the His70 and Glu84, that have hydrolase activity. Moriguchi et al., 2010 reported a hidden thioesterase function in what appeared by sequence homology to be an embedded dehydratase domain in the 6-MSA Synthase fungal multienzyme. The thioesterase activity in that domain was abolished when the conserved active site His70 residue was replaced by Ala, thus showing that an apparent DH domain could catalyze either dehydration or hydrolysis. Therefore, based on our results and on the three-dimensional models for the DH domains according to the invention, it cannot be ruled out that the DH tetradomain of PUFA synthases houses a hydrolase or esterase activity in addition to the reported dehydratase activity.

Although the present invention has been described herein with reference to the foregoing exemplary embodiment, this embodiment does not serve to limit the scope of the present invention. Accordingly, those skilled in the art to which the present invention pertains will appreciate that various modifications are possible, without departing from the technical spirit of the present invention.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 17

<210> SEQ ID NO 1
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 38 bases

<400> SEQUENCE: 1

catgcatggg atccaacttg ctagacgcaa atatcgca 38

<210> SEQ ID NO 2
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 39 bases

<400> SEQUENCE: 2

catgcatgcc cgggtcatga ttcttctttg atcatcacg 39

<210> SEQ ID NO 3
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 28 bases

<400> SEQUENCE: 3

caccttctct tacgaatgtt tcgttggc 28

-continued

<210> SEQ ID NO 4
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 39 bases

<400> SEQUENCE: 4

catgcatgcc cgggtcatga ttcttctttg atcatcacg 39

<210> SEQ ID NO 5
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 39 bases

<400> SEQUENCE: 5

catgcatggg atccaactta ctggataaag aaagccgtt 39

<210> SEQ ID NO 6
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 24 bases

<400> SEQUENCE: 6

tcaggcttct tcaatacaga ttgc 24

<210> SEQ ID NO 7
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 37 bases

<400> SEQUENCE: 7

catgcatggg atccttcagc ttcgaactca gtaccga 37

<210> SEQ ID NO 8
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 24 bases

<400> SEQUENCE: 8

tcaggcttct tcaatacaga ttgc 24

<210> SEQ ID NO 9
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 28 bases

<400> SEQUENCE: 9

caccaacttg ctagacgcaa atatcgca 28

<210> SEQ ID NO 10
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Sequence source: PCR primer, 24 bases

<400> SEQUENCE: 10

tcaggcttct tcaatacaga ttgc 24

<210> SEQ ID NO 11
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 28 bases

<400> SEQUENCE: 11

caccttctct tacgaatgtt tcgttggc 28

<210> SEQ ID NO 12
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 24 bases

<400> SEQUENCE: 12

tcaggcttct tcaatacaga ttgc 24

<210> SEQ ID NO 13
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 27 bases

<400> SEQUENCE: 13

cacccgcaaa ccttgtatct gggatta 27

<210> SEQ ID NO 14
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence source: PCR primer, 24 bases

<400> SEQUENCE: 14

tcaggcttct tcaatacaga ttgc 24

<210> SEQ ID NO 15
<211> LENGTH: 5741
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Plasmid vector; pET200TOPO

<400> SEQUENCE: 15

caaggagatg gcgcccaaca gtcccccggc cacggggcct gccaccatac ccacgccgaa 60

acaagcgctc atgagcccga agtggcgagc ccgatcttcc ccatcggtga tgtcggcgat 120

ataggcgcca gcaaccgcac ctgtggcgcc ggtgatgccg gccacgatgc gtccggcgta 180

gaggatcgag atctcgatcc cgcgaaatta atacgactca ctatagggga attgtgagcg 240

gataacaatt cccctctaga aataattttg tttaacttta agaaggagat atacatatgc 300

ggggttctca tcatcatcat catcatggta tggctagcat gactggtgga cagcaaatgg 360

gtcgggatct gtacgacgat gacgataagg atcatccctt caccaagggc gagctcaacg 420

atccggctgc taacaaagcc cgaaaggaag ctgagttggc tgctgccacc gctgagcaat 480

-continued

aactagcata acccccttggg gcctctaaac gggtcttgag gagtttttg ctgaaaggag 540

gaactatatc cggatatccc gcaagaggcc cggcagtacc ggcataacca agcctatgcc 600

tacagcatcc agggtgacgg tgccgaggat gacgatgagc gcattgttag atttcataca 660

cggtgcctga ctgcgttagc aatttaactg tgataaacta ccgcattaaa gcttatcgat 720

gataagctgt caaacatgag aattaattct tgaagacgaa agggcctcgt gatacgccta 780

tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg cacttttcgg 840

ggaaatgtgc gcggaacccc tatttgttta tttttctaaa tacattcaaa tatgtatccg 900

ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa gagtatgatt 960

gaacaagatg gattgcacgc aggttctccg gccgcttggg tggagaggct attcggctat 1020

gactgggcac aactgacaat cggctgctct gatgccgccg tgttccggct gtcagcgcag 1080

gggcgcccgg ttctttttgt caagaccgac ctgtccggtg ccctgaatga actgcaggac 1140

gaggcagcgc ggctatcgtg gctggccacg acgggcgttc cttgcgcagc tgtgctcgac 1200

gttgtcactg aagcgggaag ggactggctg ctattgggcg aagtgccggg gcaggatctc 1260

ctgtcatctc accttgctcc tgccgagaaa gtatccatca tggctgatgc aatgcggcgg 1320

ctgcatacgc ttgatccggc tacctgccca ttcgaccacc aagcgaaaca tcgcatcgag 1380

cgggcacgta ctcggatgga agccggtctt gtcgatcagg atgatctgga cgaagagcat 1440

caggggctcg cgccagccga actgttcgcc aggctcaagg cgcgcatgcc cgacggcgag 1500

gatctcgtcg tgacacatgg cgatgcctgc ttgccgaata tcatggtgga aaatggccgc 1560

ttttctggat tcatcgactg tggccggctg ggtgtggcgg accgctatca ggacatagcg 1620

ttggctaccc gtgatattgc tgaagagctt ggcggcgaat gggctgaccg cttcctcgtg 1680

ctttacggta tcgccgctcc cgattcgcag cgcatcgcct tctatcgcct tcttgacgag 1740

ttcttctgag cgggactctg gggttcgaaa tgaccgacca agcgacgcct aactgtcaga 1800

ccaagtttac tcatatatac tttagattga tttaaaactt cattttaat ttaaaaggat 1860

ctaggtgaag atccttttg ataatctcat gaccaaaatc ccttaacgtg agttttcgtt 1920

ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 1980

gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 2040

ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 2100

aaatactgtc cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 2160

gcctacatac ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc 2220

gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 2280

aacgggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 2340

cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 2400

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 2460

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 2520

atgctcgtca gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 2580

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 2640

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 2700

gcgcagcgag tcagtgagcg aggaagcgga agagcgcctg atgcggtatt ttctccttac 2760

gcatctgtgc ggtatttcac accgcaatgg tgcactctca gtacaatctg ctctgatgcc 2820

-continued

gcatagttaa gccagtatac actccgctat cgctacgtga ctgggtcatg gctgcgcccc 2880

gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt 2940

acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac 3000

cgaaacgcgc gaggcagctg cggtaaagct catcagcgtg gtcgtgaagc gattcacaga 3060

tgtctgcctg ttcatccgcg tccagctcgt tgagtttctc cagaagcgtt aatgtctggc 3120

ttctgataaa gcgggccatg ttaagggcgg ttttttcctg tttggtcact gatgcctccg 3180

tgtaaggggg atttctgttc atgggggtaa tgataccgat gaaacgagag aggatgctca 3240

cgatacgggt tactgatgat gaacatgccc ggttactgga acgttgtgag ggtaaacaac 3300

tggcggtatg gatgcggcgg gaccagagaa aaatcactca gggtcaatgc cagcgcttcg 3360

ttaatacaga tgtaggtgtt ccacagggta gccagcagca tcctgcgatg cagatccgga 3420

acataatggt gcagggcgct gacttccgcg tttccagact ttacgaaaca cggaaaccga 3480

agaccattca tgttgttgct caggtcgcag acgttttgca gcagcagtcg cttcacgttc 3540

gctcgcgtat cggtgattca ttctgctaac cagtaaggca accccgccag cctagccggg 3600

tcctcaacga caggagcacg atcatgcgca cccgtggcca ggacccaacg ctgcccgaga 3660

tgcgccgcgt gcggctgctg gagatggcgg acgcgatgga tatgttctgc caagggttgg 3720

tttgcgcatt cacagttctc cgcaagaatt gattggctcc aattcttgga gtggtgaatc 3780

cgttagcgag gtgccgccgg cttccattca ggtcgaggtg gcccggctcc atgcaccgcg 3840

acgcaacgcg gggaggcaga caaggtatag ggcggcgcct acaatccatg ccaacccgtt 3900

ccatgtgctc gccgaggcgg cataaatcgc cgtgacgatc agcggtccaa tgatcgaagt 3960

taggctggta agagccgcga gcgatccttg aagctgtccc tgatggtcgt catctacctg 4020

cctggacagc atggcctgca acgcgggcat cccgatgccg ccggaagcga gaagaatcat 4080

aatggggaag gccatccagc ctcgcgtcgc gaacgccagc aagacgtagc ccagcgcgtc 4140

ggccgccatg ccggcgataa tggcctgctt ctcgccgaaa cgtttggtgg cgggaccagt 4200

gacgaaggct tgagcgaggg cgtgcaagat tccgaatacc gcaagcgaca ggccgatcat 4260

cgtcgcgctc cagcgaaagc ggtcctcgcc gaaaatgacc cagagcgctg ccggcacctg 4320

tcctacgagt tgcatgataa agaagacagt cataagtgcg gcgacgatag tcatgccccg 4380

cgcccaccgg aaggagctga ctgggttgaa ggctctcaag ggcatcggtc gagatcccgg 4440

tgcctaatga gtgagctaac ttacattaat tgcgttgcgc tcactgcccg ctttccagtc 4500

gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt 4560

gcgtattggg cgccagggtg gtttttcttt tcaccagtga gacgggcaac agctgattgc 4620

ccttcaccgc ctggccctga gagagttgca gcaagcggtc cacgctggtt tgccccagca 4680

ggcgaaaatc ctgtttgatg gtggttaacg gcgggatata acatgagctg tcttcggtat 4740

cgtcgtatcc cactaccgag atatccgcac caacgcgcag cccggactcg gtaatggcgc 4800

gcattgcgcc cagcgccatc tgatcgttgg caaccagcat cgcagtggga acgatgccct 4860

cattcagcat ttgcatggtt tgttgaaaac cggacatggc actccagtcg ccttcccgtt 4920

ccgctatcgg ctgaatttga ttgcgagtga gatatttatg ccagccagcc agacgcagac 4980

gcgccgagac agaacttaat gggcccgcta acagcgcgat ttgctggtga cccaatgcga 5040

ccagatgctc cacgcccagt cgcgtaccgt cttcatggga gaaaataata ctgttgatgg 5100

gtgtctggtc agagacatca agaaataacg ccggaacatt agtgcaggca gcttccacag 5160

caatggcatc ctggtcatcc agcggatagt taatgatcag cccactgacg cgttgcgcga 5220

-continued

gaagattgtg caccgccgct ttacaggctt cgacgccgct tcgttctacc atcgacacca 5280

ccacgctggc acccagttga tcggcgcgag atttaatcgc cgcgacaatt tgcgacggcg 5340

cgtgcagggc cagactggag gtggcaacgc caatcagcaa cgactgtttg cccgccagtt 5400

gttgtgccac gcggttggga atgtaattca gctccgccat cgccgcttcc actttttccc 5460

gcgttttcgc agaaacgtgg ctggcctggt tcaccacgcg ggaaacggtc tgataagaga 5520

caccggcata ctctgcgaca tcgtataacg ttactggttt cacattcacc accctgaatt 5580

gactctcttc cgggcgctat catgccatac cgcgaaaggt tttgcgccat tcgatggtgt 5640

ccgggatctc gacgctctcc cttatgcgac tcctgcatta ggaagcagcc cagtagtagg 5700

ttgaggccgt tgagcaccgc cgccgcaagg aatggtgcat g 5741

<210> SEQ ID NO 16
<211> LENGTH: 14994
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli

<400> SEQUENCE: 16

aaattgaaga gtttgatcat ggctcagatt gaacgctggc ggcaggccta acacatgcaa 60

gtcgaacggt aacaggaaac agcttgctgt ttcgctgacg agtggcggac gggtgagtaa 120

tgtctgggaa actgcctgat ggagggggat aactactgga aacggtagct aataccgcat 180

aacgtcgcaa gaccaaagag ggggaccttc gggcctcttg ccatcggatg tgcccagatg 240

ggattagcta gtaggtgggg taacggctca cctaggcgac gatccctagc tggtctgaga 300

ggatgaccag ccacactgga actgagacac ggtccagact cctacgggag gcagcagtgg 360

ggaatattgc acaatgggcg caagcctgat gcagccatgc cgcgtgtatg aagaaggcct 420

tcgggttgta aagtactttc agcggggatg aagggagtaa agttaatacc tttgctcatt 480

gacgttaccc gcagaagaag caccggctaa ctccgtgcca gcagccgcgg taatacggag 540

ggtgcaagcg ttaatcggaa ttactgggcg taaagcgcac gcaggcggtt tgttaagtca 600

gatgtgaaat ccccgggctc aacctgggaa ctgcatctga tactggcaag cttgagtctc 660

gtagaggggg gtagaattcc aggtgtagcg gtgaaatgcg tagagatctg gaggaatacc 720

ggtggcgaag gcggcccccc ggacgaagac tgacgctcag gtgcgaaagc gtggggagca 780

aacaggatta gataccctgg tagtccacgc cgtaaacgat gtcgacttgg aggttgtgcc 840

cttgaggcgt ggcttccgga gctaacgcgt taagtcgacc gcctggggag tacggccgca 900

aggttaaaac tcaaatgaat tgacgggggc ccgcacaagc ggtggagcat gtggtttaat 960

tcgatgcaac gcgaagaacc ttacctggtc ttgacatcca cggaagtttt cagagatgag 1020

aatgtgcctt cgggaaccgt gagacaggtg ctgcatggct gtcgtcagct cgtgttgtga 1080

aatgttgggt taagtcccgc aacgagcgca acccttatcc tttgttgcca gcggtccggc 1140

cgggaactca aaggagactg ccagtgataa actggaggaa ggtggggatg acgtcaagtc 1200

atcatggccc ttacgaccag ggctacacac gtgctacaat ggcgcataca aagagaagcg 1260

acctcgcgag agcaagcgga cctcataaag tgcgtcgtag tccggattgg agtctgcaac 1320

tcgactccat gaagtcggaa tcgctagtaa tcgtggatca gaatgccacg gtgaatacgt 1380

tcccgggcct tgtacacacc gcccgtcaca ccatgggagt gggttgcaaa agaagtaggt 1440

agcttaacct tcgggagggc gcttaccact ttgtgattca tgactggggt gaagtcgtaa 1500

caaggtaacc gtaggggaac ctgcggttgg atcacctcct taccttaaag aagcgtactt 1560

-continued

tgcagtgctc acacagattg tctgatgaaa atgagcagta aaacctctac aggcttgtag 1620

ctcaggtggt tagagcgcac ccctgataag ggtgaggtcg gtggttcaag tccactcagg 1680

cctaccaaat ttgcacggca aatttgaaga ggttttaact acatgttatg gggctatagc 1740

tcagctggga gagcgcctgc tttgcacgca ggaggtctgc ggttcgatcc cgcatagctc 1800

caccatctct gtagtgatta aataaaaaat acttcagagt gtacctgcaa aggttcactg 1860

cgaagttttg ctctttaaaa atctggatca agctgaaaat tgaaacactg aacaacgaaa 1920

gttgctcgtg agtctctcaa attttcgcaa ctctgaagtg aaacatcttc gggttgtgag 1980

gttaagcgac taagcgtaca cggtggatgc cctggcagtc agaggcgatg aaggacgtgc 2040

taatctgcga taagcgtcgg taaggtgata tgaaccgtta taaccggcga tttccgaatg 2100

gggaaaccca gtgtgtttcg acacactatc attaactgaa tcaaattgaa gagtttgatc 2160

atggctcaga ttgaacgctg gcggcaggcc taacacatgc aagtcgaacg gtaacaggaa 2220

gcagcttgct gcttcgctga cgagtggcgg acgggtgagt aatgtctggg aaactgcctg 2280

atggaggggg ataactactg gaaacggtag ctaataccgc ataacgtcgc aagaccaaag 2340

agggggacct tagggcctct tgccatcgga tgtgcccaga tgggattagc tagtaggtgg 2400

ggtaacggct cacctaggcg acgatcccta gctggtctga gaggatgacc agccacactg 2460

gaactgagac acggtccaga ctcctacggg aggcagcagt ggggaatatt gcacaatggg 2520

cgcaagcctg atgcagccat gccgcgtgta tgaagaaggc cttcgggttg taaagtactt 2580

tcagcgggga ggaagggagt aaagttaata cctttgctca ttgacgttac ccgcagaaga 2640

agcaccggct aactccgtgc cagcagccgc ggtaatacgg agggtgcaag cgttaatcgg 2700

aattactggg cgtaaagcgc acgcaggcgg tttgttaagt cagatgtgaa atccccgggc 2760

tcaacctggg aactgcatct gatactggca agcttgagtc tcgtagaggg gggtagaatt 2820

ccaggtgtag cggtgaaatg cgtagagatc tggaggaata ccggtggcga aggcggcccc 2880

ctggacgaag actgacgctc aggtgcgaaa gcgtggggag caaacaggat tagataccct 2940

ggtagtccac gccgtaaacg atgtcgactt ggaggttgtg cccttgaggc gtggcttccg 3000

gagctaacgc gttaagtcga ccgcctgggg agtacggccg caaggttaaa actcaaatga 3060

attgacgggg gcccgcacaa gcggtggagc atgtggttta attcgatgca acgcgaagaa 3120

ccttacctgg tcttgacatc cacggaagtt ttcagagatg agaatgtgcc ttcgggaacc 3180

gtgagacagg tgctgcatgg ctgtcgtcag ctcgtgttgt gaaatgttgg gttaagtccc 3240

gcaacgagcg caacccttat cctttgttgc cagcggtccg gccgggaact caaaggagac 3300

tgccagtgat aaactggagg aaggtgggga tgacgtcaag tcatcatggc ccttacgacc 3360

agggctacac acgtgctaca atggcgcata caaagagaag cgacctcgcg agagcaagcg 3420

gacctcataa agtgcgtcgt agtccggatt ggagtctgca actcgactcc atgaagtcgg 3480

aatcgctagt aatcgtggat cagaatgcca cggtgaatac gttcccgggc cttgtacaca 3540

ccgcccgtca caccatggga gtgggttgca aaagaagtag gtagcttaac cttcgggagg 3600

gcgcttacca ctttgtgatt catgactggg gtgaagtcgt aacaaggtaa ccgtagggga 3660

acctgcggtt ggatcacctc cttaccttaa agaagcgtac tttgcagtgc tcacacagat 3720

tgtctgatag aaagtgaaaa gcaaggcgtc ttgcgaagca gactgatacg tccccttcgt 3780

ctagaggccc aggacaccgc cctttcacgg cggtaacagg ggttcgaatc ccctagggga 3840

cgccacttgc tggtttgtga gtgaaagtcg ccgaccttaa tatctcaaaa ctcatcttcg 3900

ggtgatgttt gagatatttg ctctttaaaa atctggatca agctgaaaat tgaaacactg 3960

-continued

aacaatgaaa gttgttcgtg agtctctcaa attttcgcaa ctctgaagtg aaacatcttc 4020

gggttgtgag gttaagcgac taagcgtaca cggtggatgc cctggcagtc agaggcgatg 4080

aaggacgtgc taatctgcga taagcgtcgg taaggtgata tgaaccgtta taaccggcga 4140

tttccgaatg gggaaaccca gtgtgtttcg acacactatc attaactgaa tccataggtt 4200

aatgaggcga accgggggaa ctgaaacatc taagtacccc gaggaaaaga aatcaaccga 4260

gattccccca gtagcggcga gcgaaaattg aagagtttga tcatggctca gattgaacgc 4320

tggcggcagg cctaacacat gcaagtcgaa cggtaacagg aagcagcttg ctgcttcgct 4380

gacgagtggc ggacgggtga gtaatgtctg ggaaactgcc tgatggaggg ggataactac 4440

tggaaacggt agctaatacc gcataacgtc gcaagaccaa agaggggac cttcgggcct 4500

cttgccatcg gatgtgccca gatgggatta gctagtaggt ggggtaacgg ctcacctagg 4560

cgacgatccc tagctggtct gagaggatga ccagccacac tggaactgag acacggtcca 4620

gactcctacg ggaggcagca gtggggaata ttgcacaatg ggcgcaagcc tgatgcagcc 4680

atgccgcgtg tatgaagaag gccttcgggt tgtaaagtac tttcagcggg gaggaaggga 4740

gtaaagttaa tacctttgct cattgacgtt acccgcagaa gaagcaccgg ctaactccgt 4800

gccagcagcc gcggtaatac ggagggtgca agcgttaatc ggaattactg ggcgtaaagc 4860

gcacgcaggc ggtttgttaa gtcagatgtg aaatccccgg gctcaacctg ggaactgcat 4920

ctgatactgg caagcttgag tctcgtagag gggggtagaa ttccaggtgt agcggtgaaa 4980

tgcgtagaga tctggaggaa taccggtggc gaaggcggcc ccctggacga agactgacgc 5040

tcaggtgcga aagcgtgggg agcaaacagg attagatacc ctggtagtcc acgccgtaaa 5100

cgatgtcgac ttggaggttg tgcccttgag gcgtggcttc cggagctaac gcgttaagtc 5160

gaccgcctgg ggagtacggc cgcaaggtta aaactcaaat gaattgacgg gggcccgcac 5220

aagcggtgga gcatgtggtt taattcgatg caacgcgaag aaccttacct ggtcttgaca 5280

tccacagaac tttccagaga tggattggtg ccttcgggaa ctgtgagaca ggtgctgcat 5340

ggctgtcgtc agctcgtgtt gtgaaatgtt gggttaagtc ccgcaacgag cgcaaccctt 5400

atcctttgtt gccagcggtc cggccgggaa ctcaaaggag actgccagtg ataaactgga 5460

ggaaggtggg gatgacgtca agtcatcatg gcccttacga ccagggctac acacgtgcta 5520

caatggcgca tacaaagaga agcgacctcg cgagagcaag cggacctcat aaagtgcgtc 5580

gtagtccgga ttggagtctg caactcgact ccatgaagtc ggaatcgcta gtaatcgtgg 5640

atcagaatgc cacggtgaat acgttcccgg gccttgtaca caccgcccgt cacaccatgg 5700

gagtgggttg caaaagaagt aggtagctta accttcggga gggcgcttac cactttgtga 5760

ttcatgactg gggtgaagtc gtaacaaggt aaccgtaggg gaacctgcgg ttggatcacc 5820

tccttacctt aaagaagcgt actttgcagt gctcacacag attgtctgat gaaaatgagc 5880

agtaaaacct ctacaggctt gtagctcagg tggttagagc gcacccctga taagggtgag 5940

gtcggtggtt caagtccact caggcctacc aaatttgcac ggcaaatttg aagaggtttt 6000

aactacatgt tatggggcta tagctcagct gggagagcgc ctgctttgca cgcaggaggt 6060

ctgcggttcg atcccgcata gctccaccat ctctgtagtg attaaataaa aaatacttca 6120

gagtgtacct gcaaaggttc actgcgaagt tttgctcttt aaaaatctgg atcaagctga 6180

aaattgaaac actgaacaac gaaagttgtt cgtgagtctc tcaaattttc gcaacacgat 6240

gatgaatcgc aagaaacatc ttcgggttgt gaggttaagc gactaagcgt acacggtgga 6300

-continued

tgccctggca gtcagaggcg atgaaggacg tgctaatctg cgataagcgt cggtgaggtg 6360

atatgaaccg ttataaccgg cgatttccga atggggaaac ccagtgtgat tcgtcacact 6420

atcattaaat tgaagagttt gatcatggct cagattgaac gctggcggca ggcctaacac 6480

atgcaagtcg aacggtaaca ggaagcagct tgctgcttcg ctgacgagtg gcggacgggt 6540

gagtaatgtc tgggaaactg cctgatggag ggggataact actggaaacg gtagctaata 6600

ccgcataacg tcgcaagacc aaagaggggg accttagggc ctcttgccat cggatgtgcc 6660

cagatgggat tagctagtag gtggggtaac ggctcaccta ggcgacgatc cctagctggt 6720

ctgagaggat gaccagccac actggaactg agacacggtc cagactccta cgggaggcag 6780

cagtggggaa tattgcacaa tgggcgcaag cctgatgcag ccatgccgcg tgtatgaaga 6840

aggccttcgg gttgtaaagt actttcagcg gggaggaagg gagtaaagtt aatacctttg 6900

ctcattgacg ttacccgcag aagaagcacc ggctaactcc gtgccagcag ccgcggtaat 6960

acggagggtg caagcgttaa tcggaattac tgggcgtaaa gcgcacgcag gcggtttgtt 7020

aagtcagatg tgaaatcccc gggctcaacc tgggaactgc atctgatact ggcaagcttg 7080

agtctcgtag agggggggtag aattccaggt gtagcggtga aatgcgtaga gatctggagg 7140

aataccggtg gcgaaggcgg cccccctggac gaagactgac gctcaggtgc gaaagcgtgg 7200

ggagcaaaca ggattagata ccctggtagt ccacgccgta aacgatgtcg acttggaggt 7260

tgtgcccttg aggcgtggct tccggagcta acgcgttaag tcgaccgcct ggggagtacg 7320

gccgcaaggt taaaactcaa atgaattgac gggggcccgc acaagcggtg gagcatgtgg 7380

tttaattcga tgcaacgcga agaaccttac ctggtcttga catccacgga agttttcaga 7440

gatgagaatg tgccttcggg aaccgtgaga caggtgctgc atggctgtcg tcagctcgtg 7500

ttgtgaaatg ttgggttaag tcccgcaacg agcgcaaccc ttatcctttg ttgccagcgg 7560

tccggccggg aactcaaagg agactgccag tgataaactg gaggaaggtg gggatgacgt 7620

caagtcatca tggcccttac gaccagggct acacacgtgc tacaatggcg catacaaaga 7680

gaagcgacct cgcgagagca agcggacctc ataaagtgcg tcgtagtccg gattggagtc 7740

tgcaactcga ctccatgaag tcggaatcgc tagtaatcgt ggatcagaat gccacggtga 7800

atacgttccc gggccttgta cacaccgccc gtcacaccat gggagtgggt tgcaaaagaa 7860

gtaggtagct taaccttcgg gagggcgctt accactttgt gattcatgac tggggtgaag 7920

tcgtaacaag gtaaccgtag gggaacctgc ggttggatca cctccttacc ttaaagaagc 7980

gtactttgca gtgctcacac agattgtctg atagaaagtg aaaagcaagg cgtcttgcga 8040

agcagactga tacgtcccct tcgtctagag gcccaggaca ccgcccttc acggcggtaa 8100

caggggttcg aatcccctag gggacgccac ttgctggttt gtgagtgaaa gtcgccgacc 8160

ttaatatctc aaaactcatc ttcgggtgat gtttgagata tttgctcttt aaaaatctgg 8220

atcaagctga aaattgaaac actgaacaat gaaagttgtt cgtgagtctc tcaaattttc 8280

gcaactctga agtgaaacat cttcgggttg tgaggttaag cgactaagcg tacacggtgg 8340

atgccctggc agtcagaggc gatgaaggac gtgctaatct gcgataagcg tcggtaaggt 8400

gatatgaacc gttataaccg gcgatttccg aatggggaaa cccagtgtgt ttcgacacac 8460

tatcattaac tgaatccata ggttaatgag gcgaaccggg ggaactgaaa catctaagta 8520

ccccgaggaa aagaaatcaa ccgagattcc cccagtagcg gcgagcgaaa attgaagagt 8580

ttgatcatgg ctcagattga acgctggcgg caggcctaac acatgcaagt cgaacggtaa 8640

caggaaacag cttgctgttt cgctgacgag tggcggacgg gtgagtaatg tctgggaaac 8700

-continued

tgcctgatgg agggggataa ctactggaaa cggtagctaa taccgcataa cgtcgcaaga 8760

ccaaagaggg ggaccttcgg gcctcttgcc atcggatgtg cccagatggg attagctagt 8820

aggtggggta acggctcacc taggcgacga tccctagctg gtctgagagg atgaccagcc 8880

acactggaac tgagacacgg tccagactcc tacgggaggc agcagtgggg aatattgcac 8940

aatgggcgca agcctgatgc agccatgccg cgtgtatgaa gaaggccttc gggttgtaaa 9000

gtactttcag cggggaggaa gggagtaaag ttaatacctt tgctcattga cgttacccgc 9060

agaagaagca ccggctaact ccgtgccagc agccgcggta atacggaggg tgcaagcgtt 9120

aatcggaatt actgggcgta aagcgcacgc aggcggtttg ttaagtcaga tgtgaaatcc 9180

ccgggctcaa cctgggaact gcatctgata ctggcaagct tgagtctcgt agaggggggt 9240

agaattccag gtgtagcggt gaaatgcgta gagatctgga ggaataccgg tggcgaaggc 9300

ggcccccctgg acgaagactg acgctcaggt gcgaaagcgt ggggagcaaa caggattaga 9360

taccctggta gtccacgccg taaacgatgt cgacttggag gttgtgccct tgaggcgtgg 9420

cttccggagc taacgcgtta agtcgaccgc ctggggagta cggccgcaag gttaaaactc 9480

aaatgaattg acgggggccc gcacaagcgg tggagcatgt ggtttaattc gatgcaacgc 9540

gaagaacctt acctggtctt gacatccaca gaactttcca gagatggatt ggtgccttcg 9600

ggaactgtga gacaggtgct gcatggctgt cgtcagctcg tgttgtgaaa tgttgggtta 9660

agtcccgcaa cgagcgcaac ccttatcctt tgttgccagc ggtccggccg ggaactcaaa 9720

ggagactgcc agtgataaac tggaggaagg tggggatgac gtcaagtcat catggccctt 9780

acgaccaggg ctacacacgt gctacaatgg cgcatacaaa gagaagcgac ctcgcgagag 9840

caagcggacc tcataaagtg cgtcgtagtc cggattggag tctgcaactc gactccatga 9900

agtcggaatc gctagtaatc gtggatcaga atgccacggt gaatacgttc ccgggccttg 9960

tacacaccgc ccgtcacacc atgggagtgg gttgcaaaag aagtaggtag cttaaccttc 10020

gggagggcgc ttaccacttt gtgattcatg actggggtga agtcgtaaca aggtaaccgt 10080

aggggaacct gcggttggat cacctcctta ccttaaagaa gcgtactttg cagtgctcac 10140

acagattgtc tgatgaaaat gagcagtaaa acctctacag gcttgtagct caggtggtta 10200

gagcgcaccc ctgataaggg tgaggtcggt ggttcaagtc cactcaggcc taccaaattt 10260

gcaccgcaaa tttgaagagg ttttaactac atgttatggg gctatagctc agctgggaga 10320

gcgcctgctt tgcacgcagg aggtctgcgg ttcgatcccg catagctcca ccatctctgt 10380

agtgattaaa taaaaaatac ttcagagtgt acctgcaaag gttcactgcg aagttttgct 10440

ctttaaaaat ctggatcaag ctgaaaattg aaacactgaa caacgaaagt tgttcgtgag 10500

tctctcaaat tttcgcaaca cgatgatgaa tcgcaagaaa catcttcggg ttgtgaggtt 10560

aagcgactaa gcgtacacgg tggatgccct ggcagtcaga ggcgatgaag gacgtgctaa 10620

tctgcgataa gcgtcggtga ggtgatatga accgttataa ccggcgattt ccgaatgggg 10680

aaacccagtg tgattcgtca cactatcatt aaattgaaga gtttgatcat ggctcagatt 10740

gaacgctggc ggcaggccta acacatgcaa gtcgaacggt aacaggaaac agcttgctgt 10800

ttcgctgacg agtggcggac gggtgagtaa tgtctgggaa actgcctgat ggagggggat 10860

aactactgga aacggtagct aataccgcat aacgtcgcaa gaccaaagag ggggaccttc 10920

gggcctcttg ccatcggatg tgcccagatg ggattagcta gtaggtgggg taacggctca 10980

cctaggcgac gatccctagc tggtctgaga ggatgaccag ccacactgga actgagacac 11040

-continued

ggtccagact cctacgggag gcagcagtgg ggaatattgc acaatgggcg caagcctgat 11100

gcagccatgc cgcgtgtatg aagaaggcct tcgggttgta aagtactttc agcggggagg 11160

aagggagtaa agttaatacc tttgctcatt gacgttaccc gcagaagaag caccggctaa 11220

ctccgtgcca gcagccgcgg taatacggag ggtgcaagcg ttaatcggaa ttactgggcg 11280

taaagcgcac gcaggcggtt tgttaagtca gatgtgaaat ccccgggctc aacctgggaa 11340

ctgcatctga tactggcaag cttgagtctc gtagaggggg gtagaattcc aggtgtagcg 11400

gtgaaatgcg tagagatctg gaggaatacc ggtggcgaag gcggcccccc ggacgaagac 11460

tgacgctcag gtgcgaaagc gtggggagca aacaggatta gataccctgg tagtccacgc 11520

cgtaaacgat gtcgacttgg aggttgtgcc cttgaggcgt ggcttccgga gctaacgcgt 11580

taagtcgacc gcctggggag tacggccgca aggttaaaac tcaaatgaat tgacgggggc 11640

ccgcacaagc ggtggagcat gtggtttaat tcgatgcaac gcgaagaacc ttacctggtc 11700

ttgacatcca cagaactttc cagagatgga ttggtgcctt cgggaactgt gagacaggtg 11760

ctgcatggct gtcgtcagct cgtgttgtga aatgttgggt taagtcccgc aacgagcgca 11820

acccttatcc tttgttgcca gcggtccggc cgggaactca aaggagactg ccagtgataa 11880

actggaggaa ggtggggatg acgtcaagtc atcatggccc ttacgaccag ggctacacac 11940

gtgctacaat ggcgcataca aagagaagcg acctcgcgag agcaagcgga cctcataaag 12000

tgcgtcgtag tccggattgg agtctgcaac tcgactccat gaagtcggaa tcgctagtaa 12060

tcgtggatca gaatgccacg gtgaatacgt tcccgggcct tgtacacacc gcccgtcaca 12120

ccatgggagt gggttgcaaa agaagtaggt agcttaacct tcgggagggc gcttaccact 12180

ttgtgattca tgactggggt gaagtcgtaa caaggtaacc gtaggggaac ctgcggttgg 12240

atcacctcct taccttaaag aagcgtactt tgcagtgctc acacagattg tctgataaaa 12300

agtgaaaagc aaggcgtctt gcgaagcaga ctgatacgtc cccttcgtct agaggcccag 12360

gacaccgccc tttcacggcg gtaacagggg ttcgaatccc ctaggggacg ccacttgctg 12420

gtttgtgagt gaaagtcacc tgccttaata tctcaaaact catcttcggg tgatgtttga 12480

gatatttgct ctttaaaaat ctggatcaag ctgaaaattg aaacactgaa caacgagagt 12540

tgttcgtgag tctctcaaat tttcgcaaca cgatgatgaa tcgaaagaaa catcttcggg 12600

ttgtgaggtt aagcgactaa gcgtacacgg tggatgccct ggcagtcaga ggcgatgaag 12660

gacgtgctaa tctgcgataa gcgtcggtaa ggtgatatga accgttataa ccggcgattt 12720

ccgaatgggg aaacccagtg tgtttcgaca cactatcatt aactgaatcc ataggttaat 12780

gaggcgaacc gggggaactg aaacatctaa gtaccccgag gaaaagaaat caaccgagat 12840

tcccccagta gcaaattgaa gagtttgatc atggctcaga ttgaacgctg gcggcaggcc 12900

taacacatgc aagtcgaacg gtaacaggaa gcagcttgct gcttcgctga cgagtggcgg 12960

acgggtgagt aatgtctggg aaactgcctg atggaggggg ataactactg gaaacggtag 13020

ctaataccgc ataacgtcgc aagaccaaag aggggggacct tagggcctct tgccatcgga 13080

tgtgcccaga tgggattagc tagtaggtgg ggtaacggct cacctaggcg acgatcccta 13140

gctggtctga gaggatgacc agccacactg gaactgagac acggtccaga ctcctacggg 13200

aggcagcagt ggggaatatt gcacaatggg cgcaagcctg atgcagccat gccgcgtgta 13260

tgaagaaggc cttcgggttg taaagtactt tcagcgggga ggaagggagt aaagttaata 13320

cctttgctca ttgacgttac ccgcagaaga agcaccggct aactccgtgc cagcagccgc 13380

ggtaatacgg agggtgcaag cgttaatcgg aattactggg cgtaaagcgc acgcaggcgg 13440

-continued

tttgttaagt cagatgtgaa atccccgggc tcaacctggg aactgcatct gatactggca 13500

agcttgagtc tcgtagaggg gggtagaatt ccaggtgtag cggtgaaatg cgtagagatc 13560

tggaggaata ccggtggcga aggcggcccc ctggacgaag actgacgctc aggtgcgaaa 13620

gcgtggggag caaacaggat tagataccct ggtagtccac gccgtaaacg atgtcgactt 13680

ggaggttgtg cccttgaggc gtggcttccg gagctaacgc gttaagtcga ccgcctgggg 13740

agtacggccg caaggttaaa actcaaatga attgacgggg gcccgcacaa gcggtggagc 13800

atgtggttta attcgatgca acgcgaagaa ccttacctgg tcttgacatc cacggaagtt 13860

ttcagagatg agaatgtgcc ttcgggaacc gtgagacagg tgctgcatgg ctgtcgtcag 13920

ctcgtgttgt gaaatgttgg gttaagtccc gcaacgagcg caacccttat cctttgttgc 13980

cagcggtccg gccgggaact caaaggagac tgccagtgat aaactggagg aaggtgggga 14040

tgacgtcaag tcatcatggc ccttacgacc agggctacac acgtgctaca atggcgcata 14100

caaagagaag cgacctcgcg agagcaagcg gacctcataa agtgcgtcgt agtccggatt 14160

ggagtctgca actcgactcc atgaagtcgg aatcgctagt aatcgtggat cagaatgcca 14220

cggtgaatac gttcccgggc cttgtacaca ccgcccgtca caccatggga gtgggttgca 14280

aaagaagtag gtagcttaac cttcgggagg gcgcttacca ctttgtgatt catgactggg 14340

gtgaagtcgt aacaaggtaa ccgtagggga acctgcggtt ggatcacctc cttaccttaa 14400

agaagcgtac tttgcagtgc tcacacagat tgtctgatag aaagtgaaaa gcaaggcgtc 14460

ttgcgaagca gactgatacg tcccccttcgt ctagaggccc aggacaccgc cctttcacgg 14520

cggtaacagg ggttcgaatc ccctagggga cgccacttgc tggtttgtga gtgaaagtcg 14580

ccgaccttaa tatctcaaaa ctcatcttcg ggtgatgttt gagatatttg ctctttaaaa 14640

atctggatca agctgaaaat tgaaacactg aacaatgaaa gttgttcgtg agtctctcaa 14700

attttcgcaa ctctgaagtg aaacatcttc gggttgtgag gttaagcgac taagcgtaca 14760

cggtggatgc cctggcagtc agaggcgatg aaggacgtgc taatctgcga taagcgtcgg 14820

taaggtgata tgaaccgtta taaccggcga tttccgaatg gggaaaccca gtgtgtttcg 14880

acacactatc attaactgaa tccataggtt aatgaggcga accgggggaa ctgaaacatc 14940

taagtacccc gaggaaaaga aatcaaccga gattccccca gtagcggcga gcga 14994

<210> SEQ ID NO 17
<211> LENGTH: 1755
<212> TYPE: DNA
<213> ORGANISM: Photobacterium profundum

<400> SEQUENCE: 17

tcatgatgcc tttccctctt tttgcgccca tacagccagt ttttctctaa aaatcttggc 60

gttatggcgt atatcaacgg gaaaattaga gtgaattaga aaatcgacaa tacccacagt 120

ttgttgatgt ttagcgccaa tcgttttaag ctcggcaaaa aacgtattac gggcagtatc 180

atccgtcagg cttgtgcctt tattaagctc aacacacagc agtggcactg tttcaccatg 240

acgttcaatg ccgaccaatg cggtacgctt gaccgaagag tgagtattaa atatacgttc 300

acaaggaata gaaaaatacg gcagcttgtc tgatgcaacc gcaacgcgat tggcttcaac 360

acggtgagct ttacgaccac acatccacag ttgcccgctg tcgtccaagt agcctaaatc 420

ccccatgcgg tgacgtacag tattcgcccc atcggtaatt ttagcctgaa cggttgcatg 480

atcacggtga taatacgcac ggctcaccat gggacctttt acaacgattt cgccaatttg 540

-continued

gtttacaggt aaacgtaacg tttcgtccca cgtagtaatt ggatcgtcgg tgatcgcaat 600

gatcgcaata tccacaccat caacagcttg accgacacaa ataccaccgc cattatcagt 660

gacatccgtg gttttcatta actgatcact accaatcatg gtaagaggca atgattctgt 720

cgcaccataa gaattcagta cttcaacccc atcctttagc attttactaa agcgtgaaat 780

agacgaaata gtagcaggtg cgccagcgga aataacacgt ttgatactgg gtaatgtatg 840

aggtttatga tgctgtgtac ccgctttccc taaacgctct atcaacgcag gattcacaaa 900

catattggta cattggtatt gttcaatcgc tgcaaacaaa ctgtcgggat tagccgttat 960

tggtttactg gcatccatat caggcacaat cgatgccatg cctaaagctg gaccaaaaag 1020

agagaataac gggaaggtcg caagatcacg ctcgccgtgg gcgataccat aatcattttt 1080

caatacactg atttgtgctt caaacattgc atgggtatac acaacacctt tcggcgtacc 1140

cgtactgcca ctggtaaata gaatggctgc catttcatca tttttaagct tggcaatatc 1200

ataatcaatc gcttgatgca ttttactgcg cggcttaata acagtattac gtttaagtaa 1260

tgcttttagt gttgtaccac caaaaatatt ggcaaataca tcatttccac caacggtgag 1320

taaacgtcga acactcggtt taccccaacc aaataaacaa cgggcaatat gcgctttagg 1380

gataccaata aacgcatcgg gcttagcttc atcaaagcac tgtttgaggt ttttaacccc 1440

catgccagga tcaaccaaaa taggaacaat acccgcttta aataatgcaa atgttagagc 1500

aaaaaaatca agactaggcg ttaccatcag taccgctttc atgcctcgtt taataccgtg 1560

atcattcaac gcttgtgcaa ttgcgttact gtcagtgttt aactgcccaa acgttaattc 1620

ttcataccgt agtttgccga ataacgatcg tttttgaaca gcaacagcaa gcgaagcagg 1680

tgtttctttt gccgcgcgtg ttaagtggcg acagatattg gctttagagg catcaatctc 1740

atgcccttta tccat 1755

We claim:

1. A gene vector for modifying *Escherichia coli* (*E. coli*) comprising:

the dehydratase tetradomain gene fragment DH1-DH2-UMA from *Photobacterium profundum* of SEQ ID:17 cloned into the plasmid vector of SEQ ID:15.

2. The gene vector of claim **1**, wherein said dehydratase tetradomain gene fragment encodes a DH1-DH2 protein.

3. The gene vector of claim **1**, wherein said gene vector modifies *E. coli* to produce about 4-5 times more free saturated and monounsaturated fatty acids than wild-type *E. coli*.

4. The gene vector of claim **1**, wherein said dehydratase tetradomain gene fragment is over-expressed at about room temperature.

5. A method for increasing the production of free saturated and monounsaturated fatty acids in *E. coli* comprising:

cloning the dehydratase tetradomain gene fragment DH1-DH2-UMA from *Photobacterium profundum* of SEQ ID: 17 into the plasmid vector of SEQ ID:15; and

inserting said cloned vector into *E. coli*.

6. The method of claim **5**, wherein said dehydratase tetradomain gene fragment is over-expressed at about room temperature.

7. The method of claim **5**, wherein said dehydratase tetradomain gene fragment encodes a DH1-DH2 protein.

* * * * *