Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations Graduate School

2011

Model-Guided Systems Metabolic Engineering of Clostridium thermocellum

Christopher Gowen
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

Part of the Engineering Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/2529

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

المنارة للاستشارات

www.manaraa.com

©Christopher M Gowen 2011

All rights reserved.

www.manaraa.com

MODEL-GUIDED SYSTEMS METABOLIC ENGINEERING OF *CLOSTRIDIUM THERMOCELLUM*

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY AT VIRGINIA COMMONWEALTH UNIVERSITY.

BY

CHRISTOPHER MARK GOWEN
M.S. ENGINEERING, VIRGINIA COMMONWEALTH UNIVERSITY, 2008
B.S. BIOSYSTEMS ENGINEERING, CLEMSON UNIVERSITY, 2006

DIRECTOR: STEPHEN S. FONG, PH.D.
ASSISTANT PROFESSOR, CHEMICAL & LIFE SCIENCE ENGINEERING

VIRGINIA COMMONWEALTH UNIVERSITY
RICHMOND, VIRGINIA
MAY, 2011

المنارة للاستشارات
www.manaraa.com

ACKNOWLEDGEMENT

For Grandy.
"Can't never did do nothin'"

I have many people to whom I am indebted for incredible love, support, and guidance. To begin, I would like to thank my advisor, Dr. Fong, for his guidance, instruction, and flexibility. Few educators can balance so well the sometimes competing drives for impactful research and scientific education, and I am grateful for the patience and flexibility he shows in letting his graduate students find their own independence. I would also like to thank Dr. Sherry Baldwin, Dr. Paul Brooks, Dr. Mark McHugh, Dr. Raj Rao, and Dr. Vamsi Yadavalli, who have also served quite a lot of time advising, challenging, and instructing me as members of my committee.

I have also had the pleasure of working with too many fantastic friends and colleagues over the past five years to list here, and I am indebted to all of them for good advice, commiseration, and fun times. Thanks especially to Jack for his patience and good humor when navigating jungles of paperwork, to Cindy our lab Mom for keeping us all in line and being a good friend, and to my friends and labmates (in order of appearance!), Advait, Oscar, Deng, George, and Niti for all of the fun times that will not be spoken more of here.

Finally, I am hugely blessed to have a loving and supportive family who has always encouraged me to ask questions, and to “just use your brain for a second and figure it out.” You all know who you are. At last, thanks to my beautiful wife Cathy who has patiently (!) endured uncertainty, boring science talk, odd hours, and yet more uncertainty. Thank you for all the times you have picked up my slack.

المنارة للاستشارات

www.manaraa.com

CONTENTS

المنارة للاستشارات

www.manaraa.com

www.manaraa.com

المنارة للاستشارات

www.manaraa.com

المنارة للاستشارات

www.manaraa.com

www.manaraa.com

LIST OF TABLES

www.manaraa.com

LIST OF FIGURES

www.manaraa.com

المنارة للاستشارات

www.manaraa.com

LISTS OF ABBREVIATIONS

A. General abbreviations

BLAST – Basic Local Alignment Search Tool

CAC – *Clostridium acetobutylicum*

CBP – Consolidated bioprocessing

cDNA – complimentary DNA

CoA – Coenzyme A

COBRA – Constraint-based reconstruction and analysis

CTH – *Clostridium thermocellum*

DNA – Deoxyribonucleic acid

EC – Enzyme Commission

FBA – Flux Balance analysis

FVA – Flux variability analysis

GPR – Gene-Protein-Reaction relationship

IMG – Integrated microbial genomics

www.manaraa.com

kDa – kilodaltons

KEGG – Kyoto Encyclopedia of Genes and Genomes

MILP – Mixed-integer linear programming

mmol / g-DW / hr – millimoles per gram of dry cell weight per hour

MOMA – Minimization of metabolic adjustment

NCBI – National Center for Biotechnology Information

ORF – Open reading frame

RNA – Ribonucleic acid

SCE – *Saccharomyces cerevisiae*

TCDB – Transporter Classification Database

TSS – Transcription start site

B. Reactions referenced in the text (a complete list is available in the appendix):

BIFHASE – Bifurcating hydrogenase

FDXHASE – Ferredoxin hydrogenase

LDH – Lactate dehydrogenase

المنارة للاستشارات

www.manaraa.com

NADOX or NADPOX – hydrogen:NAD(P) oxidoreductase

NFO – NAD(P):ferredoxin oxidoreductase

OAADC or OAAdecarb – Oxaloacetate decarboxylase

OOR - 2-oxoglutarate synthase

PEPCK – Phosphoenol pyruvate carboxykinase

PFL – Pyruvate:formate lyase

PPDK – Pyruvate, phosphate dikinase

PTA – Phosphotransacetylase

PYK – Pyruvate kinase

SDH – Succinate dehydrogenase

www.manaraa.com

Abstract

MODEL-GUIDED SYSTEMS METABOLIC ENGINEERING OF *CLOSTRIDIUM THERMOCELLUM*

By Christopher M Gowen, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2011.

Major Director: Stephen S. Fong, Assistant Professor, Chemical and Life Science Engineering

Metabolic engineering of microorganisms for chemical production involves the coordination of regulatory, kinetic, and thermodynamic parameters within the context of the entire network, as well as the careful allocation of energetic and structural resources such as ATP, redox potential, and amino acids. The exponential progression of "omics" technologies over the past few decades has transformed our ability to understand these network interactions by generating enormous amounts of data about cell behavior. The great challenge of the new biological era is in processing, integrating, and rationally interpreting all of this information, leading to testable hypotheses. *In silico* metabolic reconstructions are versatile computational tools for integrating multiple levels of bioinformatics data, facilitating interpretation of that data, and making functional predictions related to the metabolic behavior of the cell.

المنارة للاستشارات

www.manaraa.com

To explore the use of this modeling paradigm as a tool for enabling metabolic engineering in a poorly understood microorganism, an *in silico* constraint-based metabolic reconstruction for the anaerobic, cellulolytic bacterium *Clostridium thermocellum* was constructed based on available genome annotations, published phenotypic information, and specific biochemical assays. This dissertation describes the analysis and experimental validation of this model, the integration of transcriptomic data from an RNAseq experiment, and the use of the resulting model for generating novel strain designs for significantly improved production of ethanol from cellulosic biomass. The genome-scale metabolic reconstruction is shown to be a powerful framework for understanding and predicting various metabolic phenotypes, and contributions described here enhance the utility of these models for interpretation of experimental datasets for successful metabolic engineering.

المنارة للاستشارات

www.manaraa.com

I. INTRODUCTION

A. Systems biology: the informatics challenge

When the complete nucleotide sequence of bacteriophage ΦX174 was published by Fred Sanger's group in 1977 (Sanger et al. 1977), Maria Szekely of Imperial College London echoed the hopes of scientists everywhere when she wrote that "(f)or the first time, it is possible to look at a viral genome as a giant molecule of known structure, all functions of which should be accessible to interpretation in terms of its molecular structure"(Szekely, 1977). Yet, despite enormous advances in our understanding of molecular biology since that publication, each success has shed light on yet deeper complexity, so that a full interpretation of biological function from a nucleotide sequence still confounds us. Although unmistakably invaluable, a complete genome sequence has proven to be merely a prerequisite for understanding an organism's function.

As sequencing technology charged forward and more and more sequenced genomes became available, researchers quickly realized that they needed a way to measure which genes were being transcribed under different conditions and at different times. Maskos' and Southern's introduction of the DNA microarray in 1992 met this challenge by providing a practical, extensible, and cost-effective method to quantify RNA levels within the cell (Maskos & Southern, 1992). Fewer than 20 years later, the Gene Expression Omnibus database hosted by NCBI now holds expression data for more than 500,000 samples. Fueled by exponential increases in computational power, the same technological advance has been seen in our

المنارة للاستشارات
www.manaraa.com

abilities to measure proteins, small molecules, and even metabolic conversions within living cells. Our ability to gather data about cell behavior is no longer the sole limiting factor in the progress of science. One great challenge of the new era is in processing, integrating, and rationally interpreting all of this information, leading to testable hypotheses. Biology now has an informatics problem, much like modern internet technology. As a result, the field of systems biology has grown dramatically in past decade, making numerous contributions in methodology for managing genomic, transcriptomic, proteomic, and metabolomic datasets. Systems biology is focused on observing and interpreting biological function at a cellular level, where emergent behavior can be observed by integrating very large datasets. Typically, however, measurements from a single conceptual layer (e.g. gene, transcript, or protein level) are only horizontally integrated, partly because frameworks for vertical integration of multiple layers do not currently exist. If biologists want to truly capture the many interactions within a cell, large-scale analysis should reach across multiple levels of information.Systems biology seeks to address this challenge by developing modeling frameworks within which all of this information can be integrated systematically, hopefully leading to broader interpretation of complex and dynamic biological systems.

B. *In silico* metabolic reconstructions

One common approach to integrating and managing this complexity to address questions associated with metabolism is the *in silico* metabolic reconstruction. These reconstructions are genome-scale models that provide a framework for integrating omics datasets and predicting metabolic behavior using constraint-based modeling. Models for a range of different organisms

المنارة للاستشارات
www.manaraa.com

with varying metabolic capabilities have been published over the last ten years, including many species of bacteria, yeast, protozoa, and even humans (Chavali, Whittemore, Eddy, Williams, & J. a Papin, 2008; Duarte et al., 2007; J. Lee, Yun, Feist, Palsson, & S. Y. Lee, 2008; Oh, Palsson, S. M. Park, Schilling, & Radhakrishnan Mahadevan, 2007). Each of these models is fundamentally defined by a list of mass-balanced, and, in some cases, charge-balanced reactions. Thermodynamic (Beard, Babson, Curtis, & Qian, 2004; Jol, Kümmel, Hatzimanikatis, Beard, & Heinemann, 2010) and other types of constraints can also be incorporated to these models to provide added detail. The resulting models can simulate regulatory changes (Jensen & J. A. Papin, 2009), genetic engineering effects (Fong et al., 2005; L. Yang, Cluett, & Radhakrishnan Mahadevan, 2011), and dynamic cell behavior (Mahadevan et al. 2002). These are powerful tools for understanding and manipulating biological behavior on a genome scale.

C. Cellulosic ethanol production

1. Motivation

The utilization of cellulose for the production of ethanol or other fuels has recently been highlighted (Demain, 2009; Lynd et al., 2008) as a necessary step in order to sustainably harvest a renewable energy source that can reduce dependence on petroleum derived fuels. Global petroleum fuel consumption has risen exponentially over the past several decades, and this trend is likely to continue as demand surges from emerging economies in China and India. These effects coupled with steady or dwindling supplies of crude oil reserves worldwide foreshadow huge increases in fuel costs, to say nothing of environmental costs incurred through atmospheric release of CO_2.

www.manaraa.com

Although a number of promising and diverse solutions have been proposed to these energy challenges, the immense scale of the problem implies that wide ranging and multifaceted solutions will be necessary. Cellulosic ethanol is one such solution because it can be derived from inexpensive and abundant feedstocks like switchgrass, paper pulp waste, or corn stover. Additionally, the use of cellulosic feedstocks does not interfere with global food markets in the same ways that corn and sugar cane derived fuels do. Finally, most cellulosic crops are relatively low maintenance, meaning they can be grown on marginal lands that do not displace farmland and they require less fertilizer loading, both of which drastically improve the overall atmospheric carbon footprint.

2. Production strategies

Currently, standard practice in industrial production of cellulosic ethanol is to derive it from biomass in two distinct steps: 1) enzymatic saccharification of biomass, and 2) fermentation. The cost of enzyme production and treatment has significantly impeded more widespread use of this technology (Stephanopoulos, 2007). *Clostridium thermocellum* has generated a great deal of interest, because it both hydrolyzes cellulose and produces ethanol as a byproduct and therefore has potential to be a model organism for consolidated bioprocessing (CBP) that eliminates the separate saccharification step (Lynd, van Zyl, McBride, & M. Laser, 2005). If such a process could be industrialized, it would represent an important new and efficient method of bioethanol production.

المنارة للاستشارات
www.manaraa.com

3. Clostridium thermocellum *metabolic engineering*

Industrialization of such a process has been limited in part due to relatively low ethanol yields and productivity by *C. thermocellum*, which, in addition to ethanol, produces significant quantities of acetate, hydrogen, carbon dioxide, and formate as fermentation products. By using FBA to identify an ideal set of gene knockouts eliminating or minimizing the production of these byproducts, as well as streamlining other aspects of metabolism, it should be possible to engineer a strain of *C. thermocellum* capable of significantly improved ethanol yield and productivity.

D. Dissertation overview

The genomic revolution has resulted in a huge amount of biochemical data that is transforming the way in which scientists can approach the problems presented by metabolic engineering. My dissertation focuses on methods for generating, managing, and interpreting all of this data in ways that are meaningful to researchers while maintaining physiological relevance. *In silico* metabolic reconstructions provide the broadest available platform upon which to layer this data where it can be interpreted within the biochemical context of the cell. As a case study for the types of data integrations possible and how they can be used for metabolic engineering, my dissertation research has focused on the construction and analysis of an *in silico* model of the anaerobic cellulolytic bacterium *Clostridium thermocellum*.

The next chapter focuses on the process of construction and validation of the metabolic model as well as the ongoing refinement of the model to incorporate new physiological findings and to

المنارة للاستشارات
www.manaraa.com

adjust incorrect predictions. The resulting metabolic reconstruction is a functionally complete metabolic model containing 577 enzymatic reactions and 432 associated genes. This model provides a phenotypic context for predicting the annotation of 33 previously unknown or partially-understood genes, successfully mimics growth behavior on cellobiose and fructose in batch and continuous culture, predicts gene knockouts that couple ethanol production with growth rate, provides quantitative explanation of the bioenergetic tradeoff between hydrogen and ethanol production in *C. thermocellum*, and explores potential alternative media formulations that would improve the overall ethanol flux.

Chapter III describes the planning, execution, and initial analysis of an RNAseq transcriptomic experiment for basic growth of *C. thermocellum*. . It describes novel bioinformatics algorithms that utilize RNAseq datasets to quantify gene expression, detect operon structures and predict transcription start sites. These algorithms are used in parallel to facilitate accurate interpretation of the existing *C. thermocellum* genome annotation.

Chapter IV, describes important strides towards addressing this disconnect by integrating an RNAseq dataset with a genome-scale metabolic model in a way that is independent of a biomass objective function. The described method allows a global interpretation of quantitative expression levels within the context of the entire metabolic network and represents a novel application of a previously-described optimization algorithm towards modeling non-optimal prokaryotic metabolism and demonstrates significant improvements in prediction accuracy over both FBA alone and FBA with a Boolean integration of RNAseq data. This method is also applied to the genome-scale model combining data from the RNAseq experiment described in

المنارة للاستشارات
www.manaraa.com

Chapter III as well as proteomics data obtained during growth of *C. thermocellum* on a rich cellobiose medium. I will demonstrate how the algorithmic combination of both data types maximizes the respective strengths of the different datasets and improves physiological interpretation.

Finally, Chapter V will explore the use of the metabolic model as a tool for metabolic engineering of *C. thermocellum* for improved ethanol production. As a test case study, it will discuss the use of the method of minimization of metabolic adjustment (MOMA) to predict the phenotypic impact of a pyruvate formate lyase deletion. OptKnock simulations are also performed to search for genetic perturbations that would couple improved ethanol production with optimal growth, and these experiments discovered three main strategies for enhancing ethanol production while proposing a number of candidate genes that carry out these strategies, predicting a significant improvement in ethanol production.

المنارة للاستشارات
www.manaraa.com

II. CONSTRUCTION OF A GENOME-SCALE MODEL OF *CLOSTRIDIUM THERMOCELLUM*

A. Introduction

Although *Clostridium thermocellum* remains one of the best studied cellulolytic microorganisms known, much of the previous engineering and scientific efforts focused specifically on its mechanism and regulation of cellulose hydrolysis. A handful of studies (Desvaux, 2006; Islam, Cicek, Sparling, & D. Levin, 2006; Tolman, Kanodia, & M. F. Roberts, 1987) focused on carbon flux through central metabolism, but these results had yet to be integrated with secondary metabolic pathways and studied in the context of the entire metabolic or regulatory network. Believing that a systematic analysis of metabolism could provide valuable insight towards effective metabolic engineering of *C. thermocellum*, we set out to construct an *in silico* metabolic reconstruction based on existing genome annotations and experimental observation. The inherent limitations of automated gene annotation coupled with the complexity of any biological system dictate that curation of any metabolic reconstruction is an iterative process, where inconsistencies between prediction and observation are constantly being integrated into the model to update and correct previous iterations. This chapter describes the process of building and partially validating the initial metabolic reconstruction, up until its publication in *BioMed Central – Systems Biology* (S. B. Roberts, Gowen, Brooks, & Fong, 2010), and explores some of the physiological predictions of the model. The result is a functionally complete metabolic model containing 577 enzymatic reactions and 432 associated genes. This model provides a phenotypic context for predicting the annotation of 33 previously unknown or partially-understood genes, successfully mimics growth behavior on cellobiose and fructose in batch and continuous culture, predicts gene knockouts that couple ethanol production with

www.manaraa.com

growth rate, provides quantitative explanation of the bioenergetic tradeoff between hydrogen and ethanol production in *C. thermocellum*, and explores potential alternative media formulations that would improve the overall ethanol flux. The final section in this chapter discusses the iterative changes that have been made to the model since the initial publication and examines their physiological implications.

B. Methods

1. Construction of an in silico genome-scale stoichiometric model of C. thermocellum metabolism

a) Reaction list

The core of the stoichiometric metabolic model is a list of metabolic reactions occurring in *C. thermocellum*, compiled based on evidence from genome annotations and experimental observations. An initial list of biochemical reactions was assembled based on predicted enzymatic functions in the genomic annotations available from IMG, UniProt, and KEGG (Kanehisa & Goto, 2000; Kanehisa, Goto, Furumichi, Tanabe, & Hirakawa, 2010; Kanehisa et al., 2006; Magrane & Consortium, 2011; Markowitz et al., 2010). Specifically, Enzyme Commission (EC) numbers of annotated *C. thermocellum* genes were used to select reactions from a set of database reactions, assembled using previously published constraint-based metabolic models (Chavali et al., 2008; Duarte, Herrgård, & Palsson, 2004; Feist, Scholten, Palsson, Brockman, & Ideker, 2006; J. Lee et al., 2008). Transport reactions, representing the movement of metabolites between the extracellular space and cytosol, were initially added based on the annotations, or based on similarity searches between the *C. thermocellum* genome and the

www.manaraa.com

Transport Classification Database (TCDB) (Saier, Yen, Noto, Tamang, & Elkan, 2009). For the latter, we identified reciprocal best hits between C. thermocellum genes and genes in the TCDB using BLASTP with an e-value cutoff of 10-5, and then added reactions according to mechanisms proposed for the genes from TCDB, where available. This initial assembly contained many broken, incomplete, and/or isolated pathways and subsystems. Furthermore, many reactions involving the synthesis of large, complex molecules, e.g., cell wall teichoic acids and extracellular proteins, were missing from this initial reaction list.

To represent the synthesis of large molecules such as phospholipids and cell wall teichoic acid, we created lumped reactions that produced an "average" of that molecular species from a representative fraction of small molecules. For example, to account for the synthesis of 1-acyl-glycerol 3-phosphate, we first created lumped reactions synthesizing necessary fatty acyl-CoA molecules from malonyl-CoA, acetyl-CoA, and other cofactors. Next, we used percent w/w fatty acid content for *C. thermocellum* as determined by (Chan, Himes, & Akagi, 1971) and (Herrero, Gomez, & M. F. Roberts, 1982) to construct a reaction synthesizing a representative molecule of 1-acyl-glycerol 3-phosphate. The coefficients for each of the fatty acids in this reaction are based on the fatty acid content of the cell (see calculations in the supplementary documentation). The resulting average glycerophospholipids were then available for conversion to other components of the phospholipid bilayer and cell wall. This approach has been used previously (Chavali et al., 2008; Oh et al., 2007).

www.manaraa.com

b) Definition of biomass flux reaction

In order to analyze a constraint-based model by FBA, one must specify a metabolic objective. We specified maximization of biomass as *C. thermocellum's* metabolic objective. As in previous work, we represented biomass production by a lumped reaction, with metabolites crucial for biomass formation drained in proportion to their respective requirements (Chavali et al., 2008; Fong, Marciniak, & Palsson, 2003). The identity of these metabolites and the required amounts for each was determined for *C. thermocellum* for metabolites for which experimental biomass composition data was available. In the absence of experimental data specific for *C. thermocellum*, we used similar data from the genome scale metabolic model of the gram positive bacterium *Bacillus subtilis* (Oh et al., 2007). Although the specific biomass requirements of *B. subtilis* no doubt differ from those of *C. thermocellum*, the results of FBA are relatively insensitive to minor variations of secondary metabolites in the biomass requirements (Varma & Palsson, 1993), although predicted phenotype has been shown to be sensitive to major energetic components like ATP (Senger, 2010). The cellular dry weight was divided into the broad categories of protein, RNA, DNA, lipids and cell wall, and ions and metabolites, and each category was divided into subcategories to obtain the molar contribution of individual metabolites to cell mass. Distribution of individual amino acids, deoxyribonucleic acids, and ribonucleic acids in the protein, DNA, and RNA categories was determined by performing counts of each in either the whole genome (for DNA) or in every open reading frame (for RNA and protein). See supplementary information for detailed calculations and results.

www.manaraa.com

c) Gene-protein-reaction (GPR) relationships

GPR relationships specify the putative relationship between genes and enzymatic activities in an organism. Following previous work (Duarte et al., 2007), we represented these relationships as Boolean statements. The simplest such statement was: gene A implies reaction X, i.e., gene A -> reaction X. Enzymatic activities associated with protein complexes required more complicated statements, e.g., (gene A or gene B) and (gene C) -> reaction X.

To develop the GPR relationships for *C. thermocellum*, we used the annotations described above and information on protein complexes from UniProt. For each EC number in the *C. thermocellum* annotations, we searched UniProt (Magrane & Consortium, 2011) to determine whether that EC is associated with protein complexes, and if so, what type of complex exists across different organisms (homodimer, heterotrimer, etc.). Based on this information, and the information in the annotations, we assigned putative GPR relationships for *C. thermocellum*, conforming to known enzyme complex architecture whenever possible. As a specific example of this process, consider the reaction corresponding to carbamoyl-phosphate synthase ('R_CBPS'), EC 6.3.5.5. Several UniProt entries (e.g., CARA_ECOLI) with this EC number are annotated as being members of a complex composed of two chains, a small glutamine-hydrolyzing chain and a large chain that synthesizes carbamoyl phosphate. We found two *C. thermocellum* genes annotated as "carbamoyl-phosphate synthase, small subunit" (Cthe_1867, Cthe_0950) and two genes annotated as "carbamoyl-phosphate synthase, large subunit" (Cthe_1868, Cthe_0949). Thus, we expressed the GPR relationship for this reaction as: (Cthe_1867 or Cthe_0950) and (Cthe_1868 or Cthe_0949) -> R_CBPS.

www.manaraa.com

d) Accounting for cellulosome production

C. thermocellum is one of a number of cellulolytic microorganisms that breaks down cellulose via a large (from 2 to 16,000 kDa), multi-functional polypeptide assembly of extracellular enzymes and scaffold proteins called the cellulosome (Lynd & Y. Zhang, 2002). The amount of cellulosome produced by *C. thermocellum* varies considerably across specific conditions. For example, the amount was found to be nine-fold greater in Avicel-grown batch cultures compared to cellobiose-grown batch cultures (Y. Zhang & Lynd, 2003). Cellulosome-associated polypeptides can be present in quantities as high as 20% percent of the cell mass exclusive of cellulosomes (Y. Zhang & Lynd, 2003), and this likely places a significant additional metabolic burden on the cell. To address this issue, *i*SR432 includes a means of including an additional "cellulosome" requirement in the biomass objective. A lumped reaction was created to represent the formation/export of the cellulosome from the constitutive amino acids and using ATP. The pseudometabolite "cellulosome," representing the outcome of this reaction, can be optionally added to the "cellmass" fraction of biomass discussed above to form total biomass. The molar coefficients of each amino acid to form a single "mole" of cellulosome were determined based on a proteomics study of the *C. thermocellum* cellulosome, which produced the relative abundances of individual peptide constituents of the cellulosome (Gold & Martin, 2007). These data were combined with the protein sequences of each peptide to determine the relative molar abundance of amino acids present in the cellulosome, where one mole of cellulosome equals the average molecular weight of one of its amino acid residues. The ATP cost of assembly as well as the proportion of cellulosome to cellmass needed to create biomass can both be adjusted either directly based on calculated anabolic costs or measured

المنارة للاستشارات
www.manaraa.com

cellulosome production or indirectly to match observed experimental growth patterns. Based on the cost of one mole of ATP hydrolyzed to AMP and pyrophosphate per mole of charged tRNA and one mole of GTP hydrolyzed per mole of amino acid added to a peptide chain during translation, it was assumed that one ATP molecule must be hydrolyzed to ADP and inorganic phosphate to build and export a single average cellulosome building block. The complete external cellulosome, when expressed, was taken to be present at 20% of the total cell mass (Y. Zhang & Lynd, 2003).

e) Model naming convention

We followed the previously established convention (Reed, Vo, Schilling, & Palsson, 2003) for naming genome-scale constraint-based models, i.e., "*i*" to denote "*in silico*," followed by the initials of the first author ("SR") from the model's original publication (S. B. Roberts et al., 2010), followed by the number of genes included in the model ("432").

2. Cell growth

C. thermocellum (ATCC 27405) was cultured anaerobically at 55 °C in serum bottles sealed with butyl rubber stoppers and purged using ultra high purity nitrogen gas. Standard growth medium for *C. thermocellum* contained in final concentrations 5 g/L cellobiose, 3 g/L sodium citrate tribasic dihydrate, 1.3 g/L ammonium sulfate, 1.43 g/L potassium phosphate monobasic, 1.8 g/L potassium phosphate dibasic, 0.13 g/L calcium chloride dihydrate, 6 g/L glycerol-2-phosphate disodium, 4.5 g/L yeast extract, 2.6 g/L magnesium chloride hexahydrate, 1.1 mg/L ferrous

www.manaraa.com

sulfate heptahydrate, 0.5 g/L cystein-HCl, 50 mM MOPS buffer, and 0.1% resazurin. *Escherichia coli* was cultured aerobically at 37 °C in M9 minimal growth medium containing 2 g/L glucose.

3. Assay for succinate dehydrogenase (SDH) and lactate dehydrogenase (LDH) enzymatic activity

The enzymatic activity of succinate dehydrogenase in *C. thermocellum* was tested using a modification of the method used by Kun and Abood (Kun & Abood, 1949) in which a tetrazolium dye is reduced by the dehydrogenase enzyme in the conversion of succinate to fumarate. This assay was also used to detect the activity of lactate dehydrogenase by substituting lactate for succinate in the assay. Cells were harvested in late log phase, centrifuged for 15 min at 8000 g, washed once and resuspended in half the original volume of distilled water. Into a 15 mL centrifuge tube were added 0.5 mL of phosphate buffer at pH 7.8, 0.5 ml of either succinate or lactate at 0.2 M concentration, 1.0 mL of either 100% or 50% cell suspension, and 1.0 mL of 0.1% nitro blue tetrazolium. The tubes were incubated in either a 37 °C (*E. coli*) or 50 °C (*C. thermocellum*) water bath for 20 minutes. Tubes were removed, and the reaction was stopped by adding 7 mL acetone and shaking well. A blue color indicated reduction of the tetrazolium dye, and thus, the presence of the query enzyme activity.

4. Model analysis

a) Flux Balance Analysis

One of the core analytical tools used during model construction and analysis was flux balance analysis (FBA) (Edwards, Covert, & Palsson, 2002). In essence, FBA uses linear programming to identify a single point in the space of possible steady state metabolism that optimizes a given

www.manaraa.com

metabolic objective. As discussed above, we specified maximization of flux on the biomass reaction as the metabolic objective. The optimization problem of FBA can be expressed as

$$\max_{v_i} v_{biomass}$$

$$\text{s.t.: } S \cdot v = 0$$

$$\text{s.t.: } a_i \leq v_i \leq b_i \text{ for all reactions i,}$$

where $v_{biomass}$ is the flux on the biomass reaction, S is the stoichiometric matrix (discussed below), and v is the vector representing the flux values of each reaction – including the biomass reaction – in the model (the flux distribution) (Edwards et al., 2002). The first statement gives the objective. The second statement encapsulates the steady-state assumption, that the concentrations of all metabolites are unchanging. The third statement encapsulates specific flux constraints for each reaction, e.g., whether the reaction is reversible, or not (if the reaction is irreversible, then the lower bound is set to zero). These bounds can also be used to constrain individual reactions to zero flux ($a_i = b_i = 0$) or other experimentally-determined flux values. If no information was available, the upper and lower flux bounds were set to 1000 and -1000, respectively. The stoichiometric matrix S is the mathematical representation of the reaction list. It has one row for each metabolite and one column for each reaction. Each element of S, s_{ij}, represents the stoichiometric coefficient of the *i*th metabolite in the *j*th reaction (the coefficients are positive when the metabolite is a product of the given reaction and negative when the metabolite is a reactant). The vector ***v*** contains the flux values for all reactions in the model, i.e., v_i is the flux on the *i*th reaction.

www.manaraa.com

b) Gene Essentiality Predictions

Constraint-based models can be analyzed by FBA to make comprehensive *in silico* gene essentiality predictions (Oh et al., 2007). For each gene in the model, that gene is assumed to be deleted or nonfunctional. Then, using the GPR relationships, the effect of the *in silico* gene knockout on reaction activities is assessed. If the gene is crucial to the activity of a reaction in the model, then that reaction is constrained to have zero flux to simulate the effect of the gene deletion. Finally, the model with this new constraint is analyzed by FBA. If the maximum flux on the biomass reaction is zero, then the deleted gene is predicted to be essential. If the maximum flux on the biomass reaction is greater than zero, the deleted gene is predicted to be nonessential. All constraints are reset to their default values, and the process is repeated for the next gene in the model.

c) Flux Variability Analysis

FBA is guaranteed to produce an optimal solution to the linear programming problem stated above. However, there generally are many optimal solutions, i.e., different flux distributions that give the same optimal objective value (Bonarius, Schmid, & Tramper, 1997; R Mahadevan, 2003; Reed & Palsson, 2004). In order to examine production capabilities for ethanol, we used flux variability analysis (FVA) (R Mahadevan, 2003)to account for possible variation in ethanol production across optimal solutions. In flux variability analysis, the model is first analyzed by FBA, thus determining the optimal value for the biomass reaction flux. This flux is then imposed as a constraint, i.e., the model is forced to operate at the optimal level. For the present study, we constrained the biomass flux lower and upper bounds to 99% and 100% of the value from

المنارة للاستشارات

www.manaraa.com

FBA, respectively. Then, the flux on a reaction of interest (e.g., ethanol production) is maximized and then minimized by FBA, to find upper and lower bounds on that reaction flux that are consistent with optimal biomass reaction flux.

d) New Techniques for Model-Model Comparison

We compared genome-scale constraint based models for three key organisms in ethanol production, *Clostridium acetobutylicum* (represented by two models: Cac1 (Senger & Papoutsakis, 2008a, 2008b) and Cac2 (J. Lee et al., 2008)) *C. thermocellum* (Cth), and *Saccharomyces cerevisiae* (Sce) (Duarte et al., 2004). Direct comparison of model reactions is hampered by the fact that different models use different conventions for naming metabolites and writing reactions. As an example, consider the reaction for hexokinase. In the published *C. thermocellum* and *S. cerevisiae* models, this reaction is written as,

atp + glc-D --> adp + g6p + h

while in the Cac1 model, this reaction is written as,

ATP + beta-D-Glucose <=> ADP + beta-D-Glucose 6-phosphate

and in the Cac2 model, the reaction is written,

bDG6P + ADP <-> ATP + bDGLC.

www.manaraa.com

These simple variations in notation can be corrected rather easily, but more complex differences in the model conventions exist that confound direct comparison of model contents. Aside from the different metabolite identifiers, reactions in the Cac1 and Cac2 models are written using the "biochemical" convention, as opposed to the charge-balanced "chemical reactions" used in the others. Usually, this difference is manifested in whether hydrogen is represented as a distinct metabolite in a reaction or not (see above). However, it also results in other differences, e.g., the metabolite NH_4 is used instead of NH_3 in the *C. thermocellum* and *S. cerevisiae* models, while the opposite is true in the Cac1 and Cac2 models. In addition, the Cac1 and Cac2 models distinguish between different metabolite anomers (e.g., alpha-D-glucose and beta-D-glucose) while the others do not. For these reasons and others, we decided to use EC numbers to compare the contents of each model. EC numbers have the advantage that they are relatively stable over time, represent international consensus on biochemical reaction categorization, and can overcome ambiguities regarding how to represent a given metabolic reaction in a constraint-based model. It should be noted, however, that comparison of models based on EC numbers has limitations, because similar reactions (e.g. reactions performing the same “backbone” conversion using different cofactors) will not be recognized as related by this analysis.

Conventions in assigning EC numbers to reactions can vary. In the Cac1 and Cth models, a single reaction could be associated with multiple EC numbers while only one EC number was assigned to each reaction in the Cac2 and Sce models. To ensure that the comparison was uniform, the lists of EC numbers used to represent Cac1 and Cth were compared with the other models, and

المنارة للاستشارات
www.manaraa.com

EC numbers unique to these models (Cac1 and Cth) were only kept if no other EC number for that same reaction occurred in the other models.

There are also no universally adhered to guidelines for assigning reactions (and thus their associated EC numbers) to subsystems. To minimize the impact of this on our EC based model comparisons, we attempted to use KEGG as an "external standard." Specifically, for each EC number encountered in any of the models, we assigned all pathways associated to that EC by KEGG.

This EC-based method of comparison necessarily excludes reactions not linked to an EC number. This means that significant numbers of reactions from subsystems such as extracellular transport and cell envelope biosynthesis are not a part of the comparison. However, the EC-based comparisons do cover most reactions in all four models. Excluding transport reactions, the percentages of reactions not associated with an EC number are 8%, 17%, 12%, ~5%, for the Cth, Sce, Cac1, Cac2 models, respectively.

e) Simulation of alternative media formulations

*i*SR432 was used to simulate the addition of an arbitrary amount of various common metabolites in order to determine the effect on ethanol productivity and yield. To simulate an added metabolite, an exchange flux for that metabolite was allowed with the constraints $-20 \leq v \leq 0$ mmol / g-DW / hr. This reflects the thermodynamic prevention of any production of that metabolite and allows the use of it as a source within a reasonable range. For each alternative media formulation, FVA was performed for the ethanol exchange flux and the maximum

المنارة للاستشارات
www.manaraa.com

ethanol flux was used along with the predicted maximum growth rate to calculate the ethanol:biomass yield.

C. Results

We integrated genomic, biochemical, and physiological information to construct the first complete genome-scale *in silico* metabolic reconstruction of *C. thermocellum,* which includes a unique model representation of a cellulosome, a major functional unit in cellulose hydrolysis that accounts for a high percentage of the total protein content in *C. thermocellum*. The accuracy of the constructed model was determined by comparing the results of model simulations to experimental results for growth on cellobiose and fructose from two different independent studies. FVA was used to study the range of ethanol production capabilities of both wild-type and gene deletion strains of *C. thermocellum* in different chemical environments. Metabolic network comparisons were also conducted between *C. thermocellum* and a related butanol-producing species *C. acetobutylicum* and between *C. thermocellum* and the model ethanologen, *Saccharomyces cerevisiae*. All computational and experimental work was based upon the wild-type strain of *C. thermocellum* (ATCC 27405).

1. Model Description

Based upon genomic and available physiologic evidence, a genome-scale constraint-based metabolic model for *C. thermocellum*, hereafter denoted *i*SR432, was developed. *i*SR432 contains 577 reactions representing the function of 432 genes (see Table 1). Of the 577 reactions, 73 represent transport processes from the extracellular space to the cytoplasm, or

المنارة للاستشارات
www.manaraa.com

vice-versa. Four hundred sixty-three of the reactions are associated with genes. Of the reactions that are not associated with genes, 60 are intracellular and 54 are transport reactions. The metabolic subsystems with the highest number of reactions missing gene assignments are transporters and cell envelope biosynthesis. The fact that many transporters are not assigned genes is not surprising in light of the fact that these processes are generally not as well-characterized as metabolic reactions due to the potential multi-functionality of transport mechanisms. The absence of gene assignments for many of the reactions in cell envelope biosynthesis is mainly due to the fact that many of these are lumped reactions, representing several sequential metabolic transformations (see Methods section II.B.1.a)). There are 525 distinct metabolites included, counting extracellular and intracellular forms of the same species as one.

The breakdown of *i*SR432 by functional categories is represented in Figure 1. Amino acid metabolism forms the largest of these, with 126 reactions. Experimentally, *C. thermocellum* appears to have the capacity to synthesize all 20 amino acids (E. A. Johnson, Madia, & Demain, 1981), and *i*SR432 reflects this. Other large groups of reactions include subsystems related to carbon source processing (glycolysis, pentose phosphate pathway, pyruvate metabolism, and the citric acid cycle) and cell envelope biosynthesis (including reactions related to phospholipid, peptidoglycan, and teichoic acid synthesis). The model includes reactions relating to the synthesis of a variety of vitamins and cofactors, including biotin, NAD, pantothenate, and riboflavin.

المنارة للاستشارات
www.manaraa.com

Table 1: Overview of four genome-scale constraint-based models related to ethanol or butanol production (S. B. Roberts et al., 2010)**.**

	C. thermocellum *iSR432*	*C. acetobutylicum*[a]	*C. acetobutylicum* *CacMBEL502*[b]	*S. cerevisiae* *iND750*[c]
Genome size	3.8 Mb	4.1 Mb	4.1 Mb	12.2 Mb
ORFs	3307	4017	4017	6276
Included genes	432	458	432	750
Enzyme complexes	72	n/a[d]	36	86
Isozyme cases	70	n/a	n/a	145
Reactions (excluding exchanges)	577	552	502	1150
Transport	73	80	71	308
Gene associated	463	414	431	810
Non-gene associated intracellular	60	119	n/a	123
Non-gene associated transports	54	19	n/a	216
Distinct metabolites	525	488	479	646

[a] (Senger & Papoutsakis, 2008a)
[b] (J. Lee et al., 2008)
[c] (Duarte et al., 2004)
[d] n/a – data not available

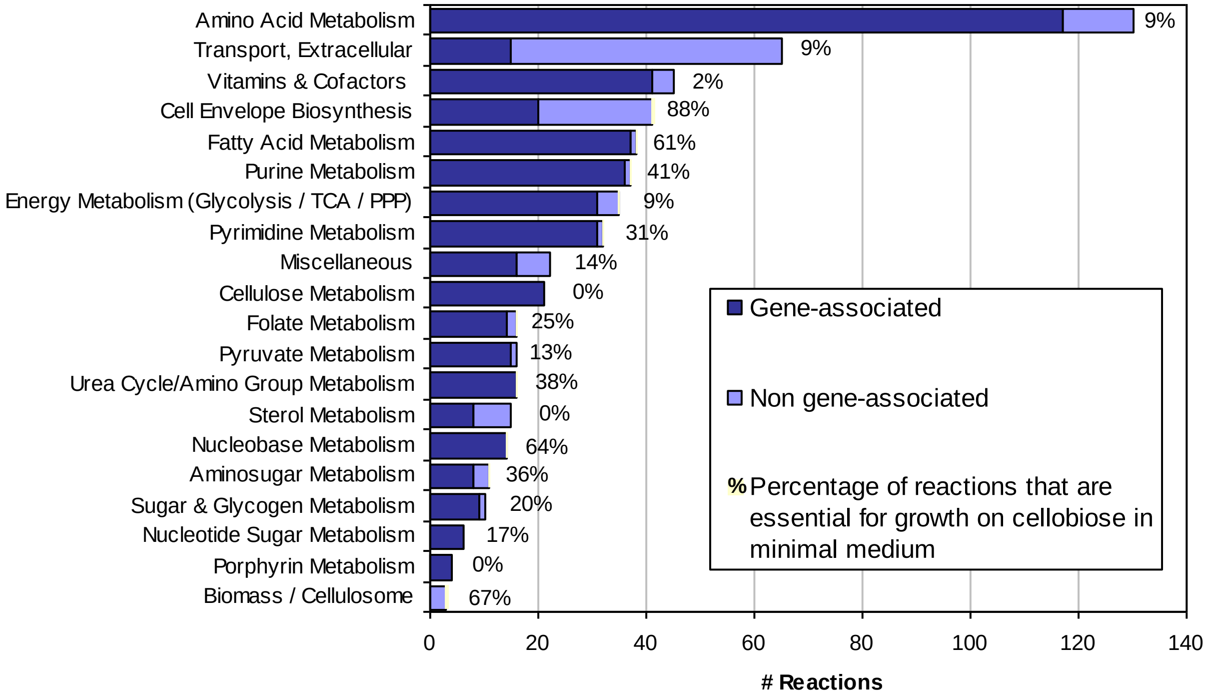


Figure 1: Distribution of reactions in C. thermocellum iSR432 model by functional category (S. B. Roberts et al., 2010)**.** Each reaction in the model is assigned to a single functional category. The length of each bar indicates the total number of reactions falling into each category, and the dark- and light-blue portions indicate the number of reactions that are currently mapped to

www.manaraa.com

C. thermocellum open reading frames (gene-associated) or not (non gene-associated), respectively. The bar labels indicate the percentage of reactions in each category which are predicted to be essential for growth on cellobiose in MJ minimal medium (E. A. Johnson et al., 1981).

a) Initial reconstruction

The initial reconstruction of *C. thermocellum* metabolism was assembled based on the existing annotations (see Methods section II.B.1.a)) and using a database of biochemical reactions compiled from previously published genome-scale constraint-based models, resulting in an initial set of approximately 400 reactions. This initial version was incomplete in several respects, including gaps in critical metabolic subsystems and the inability to synthesize metabolites known to be crucial (e.g., phospholipids). This initial version was then supplemented with reactions representing processes known to be crucial, but missing in this first version. Some of these reactions were missing because they represent either processes specific to *C. thermocellum*, or processes that must be represented by manually constructed, lumped reactions. Examples include cellulosome production, cellulose and chitin breakdown, fatty acid synthesis, phospholipid synthesis, teichoic acid and peptidoglycan synthesis, steroid metabolism, and various transport processes known to occur in *C. thermocellum*. We also added missing transport reactions for important inputs (e.g., water) and known metabolic fermentation products (e.g., formate). This second model version resulted in the addition of about 150 additional reactions.

المنارة للاستشارات
www.manaraa.com

b) Gap filling

After the aforementioned changes, the resulting second version was still incomplete due primarily to gaps in critical metabolic subsystems. These were addressed by manual curation of the individual subsystems, using temporary exchange reactions (temporary artificial transport mechanisms used to isolate a subsystem of pathways) to identify missing components. For example, in glycolysis, there was no pyruvate kinase included in the early versions of the model, because none of the existing genome annotations included a gene with this annotated function. However, upon inspection of the glycolysis pathway, this gap became self-evident. Because essentially all of the surrounding reactions (i.e., reactions leading up to and away from the missing pyruvate kinase reaction) were present and because pyruvate kinase activity has been experimentally observed in *C. thermocellum* (Patni & Alexander, 1971), we decided to include this reaction even though there was initially no genetic evidence[1]. In an attempt to discover genes of *C. thermocellum* that might code for a missing enzyme, we used BLASTP similarity searches between the translated set of *C. thermocellum* genes and enzymes from public databases with the annotation of interest. For example, for pyruvate kinase (EC 2.7.1.40) all sequences from UNIPROT annotated with this EC number (2.7.1.40) were downloaded. Using BLASTP, we identified reciprocal best hits (RBH) between the set of genes in *C. thermocellum* and the set of genes from UNIPROT annotated to EC 2.7.1.40, using an e-value cutoff of 10^{-5}. There were generally several RBH, and from these we selected a candidate *C. thermocellum* gene that could plausibly perform the function in question. Similar procedures were repeated

---

[1] The reaction for pyruvate kinase (PYK) has since been removed in later versions of the model and replaced with phosphoenolpyruvate carboxykinase (R_PEPCK) based on inability to detect PYK activity in *C. thermocellum*. See section II.D.1 for details.

www.manaraa.com

for all other model subsystems, until we were able to generate a positive flux on the biomass reaction when analyzing the model by FBA. Gaps that were filled using reactions that were not associated with direct genetic evidence were noted. The results of this analysis (shown in Table 2) are 27 genes that likely should have annotations added/modified for *C. thermocellum*.

www.manaraa.com

Table 2: Possible new annotations for C. thermocellum ORFs based on identified metabolic gaps (S. B. Roberts et al., 2010)**.**

Missing EC number	Enzyme name	Possible Cth ORF	Current annotation	Reciprocal best hit[a]	E value [b]
2.7.1.107	diacylglycerol kinase	Cthe_3168	hypothetical protein	A0R923	7.00E-35
3.2.1.52	hexosaminidase	Cthe_0787	isoleucyl-tRNA synthetase	A4N847	0
2.7.7.39	CDP-glycerol pyrophosphorylase	Cthe_1276	pantetheine-phosphate adenylyltransferase	A1SHB9	3.00E-07
2.7.8.8	phosphatidylserine synthase	Cthe_0158	Ribonuclease	A9JBA9	1.00E-68
4.1.1.65	phosphatidylserine decarboxylase	Cthe_0505	formate acetyltransferase	A4NBN7	0
1.3.99.1	succinate dehydrogenase	Cthe_2355	L-aspartate oxidase	Q97W79	1.00E-94
1.3.1.6	NADH-fumarate reductase	Cthe_2355	L-aspartate oxidase	B0VG44	5.00E-98
1.3.5.1	succinate dehydrogenase	Cthe_2355	L-aspartate oxidase	A4YEK0	2.00E-85
6.2.1.5	succinate--CoA ligase	Cthe_1907	amino acid adenylation domain	A3P3B3	3.00E-40
6.4.1.2	acetyl-CoA carboxylase	Cthe_0699	carboxyl transferase	A0RY61	8.00E-169
6.3.4.14	biotin carboxylase	Cthe_0949	carbamoyl-phosphate synthase, large subunit	B2J980	0
4.1.3.38	aminodeoxychorismate lyase	Cthe_0026	queuosine biosynthesis protein	Q03L66	3.00E-54
3.1.3.1	alkaline phosphatase	Cthe_2965	binding-protein-dependent transport systems inner membrane component	B0USD4	1.00E-59
2.6.1.2	alanine transaminase	Cthe_0755	aminotransferase, class I and II	Q7LYW0	3.00E-66
2.6.1.51	serine--pyruvate transaminase	Cthe_0265	aminotransferase, class V	B4BE13	0
2.7.1.39	homoserine kinase	Cthe_0397	ABC transporter related protein	A5M0U7	3.00E-140
3.1.3.3	phosphoserine phosphatase	Cthe_0256	histidine kinase	A9G173	3.00E-54
2.7.1.40	pyruvate kinase	Cthe_1955	RNA binding S1	A5LC67	0
1.2.2.1	formate dehydrogenase	Cthe_0199	4Fe-4S ferredoxin, iron-sulfur binding	Q2LVY6	9.00E-11
1.7.99.4	nitrate reductase	Cthe_0200	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	Q11VH4	4.00E-24
2.2.1.2	transaldolase	Cthe_0217	Glucose-6-phosphate isomerase	Q2S6E8	1.00E-22
6.3.4.1	GMP synthase	Cthe_0375	GMP synthase, large subunit	A2C5P2	4.00E-176
1.2.1.2	formate dehydrogenase	Cthe_0341	NADH dehydrogenase (quinone)	B5IPC7	6.00E-126
5.3.3.2	isopentenyl-diphosphate Delta-isomerase	Cthe_1022	Glycerol-3-phosphate dehydrogenase	A8VXT5	2.00E-94
2.5.1.29	farnesyltranstransferase	Cthe_0831	Polyprenyl synthetase	B2J443	4.00E-69
2.5.1.33	trans-pentaprenyltranstransferase	Cthe_0564	Trans-hexaprenyltranstransferase	Q6KZR8	3.00E-25
3.2.1.108	lactase	Cthe_0212	Beta-glucosidase	P09848	3.00E-89
3.5.1.19	nicotinamidase	Cthe_1178	isochorismatase hydrolase	Q6F6U3	6.00E-08
1.2.4.4	branched chain keto acid dehydrogenase	Cthe_0547	periplasmic solute binding protein	A8VXE7	2.00E-27

[a] UniProt accession numbers

[b] E value based on reciprocal best hit against *C. thermocellum* gene

www.manaraa.com

c) Citric acid cycle and the fate of succinate

We noted during model curation that the citric acid cycle was complete except for two reactions, namely succinate dehydrogenase and succinate:CoA ligase. There was no clear genetic evidence for these genes, although similarity searches produced some probable reciprocal best hits (see Table 2). Due to the importance of these genes and their potential effect on metabolism, we conducted experimental assays for succinate dehydrogenase (SDH) activity in *C. thermocellum*. Using a colorimetric assay that can detect the activities of various dehydrogenase enzymes, we measured succinate dehydrogenase and lactate dehydrogenase (LDH) activity in both *E. coli* (control) and *C. thermocellum* (LDH as a control). We were able to detect LDH activity in both *C. thermocellum* and E. coli as well as SDH activity in *E. coli* (data not shown). We were unable to detect SDH activity in *C. thermocellum*. Based on the absence of genetic and experimental biochemical evidence for this reaction, we elected to exclude the succinate dehydrogenase reaction from *i*SR432. We did, however, add a reaction for succinate:CoA ligase, based on the fact that succinate has been reported as a metabolic fermentation product of *C. thermocellum* (Lalaurette, Thammannagowda, Mohagheghi, Maness, & Logan, 2009; Mcbee, 1954) and genetic evidence from our similarity searches.

2. Comparison of model results to experimental results

To test the predictions of *i*SR432, we simulated growth of *C. thermocellum* by applying FBA, assuming minimal media conditions with one of two possible carbon sources (cellobiose or fructose). We then compared model predictions to experimentally observed growth rates and fermentation product secretion profiles of *C. thermocellum* grown in either continuous

www.manaraa.com

(Hogsett, 1995) or batch (Rydzak, D. B. Levin, Cicek, & Sparling, 2009) culture. In addition to the carbon source, the *in silico* minimal medium used for simulations contained water (h2o), ammonia (nh4), sulfate (so4), phosphate (pi), calcium (ca2), ferrous iron (fe3), hydrogen sulfide (h2s), potassium (k), magnesium (mg2), pantothenate (pnto-r), and nicotinate D-ribonucleotide (nmn). In each of the three simulation conditions, we applied progressively more experimentally determined constraints associated with by-product secretion rates to determine how closely the computational results could match the experimental results given the possibility of alternate optimal solutions (Reed & Palsson, 2004). Figure 2 shows the simulation results (given as the possible range of reaction fluxes) for each growth condition when exchange rates for the carbon sources, acetate, and formate were constrained to match experimental observations.

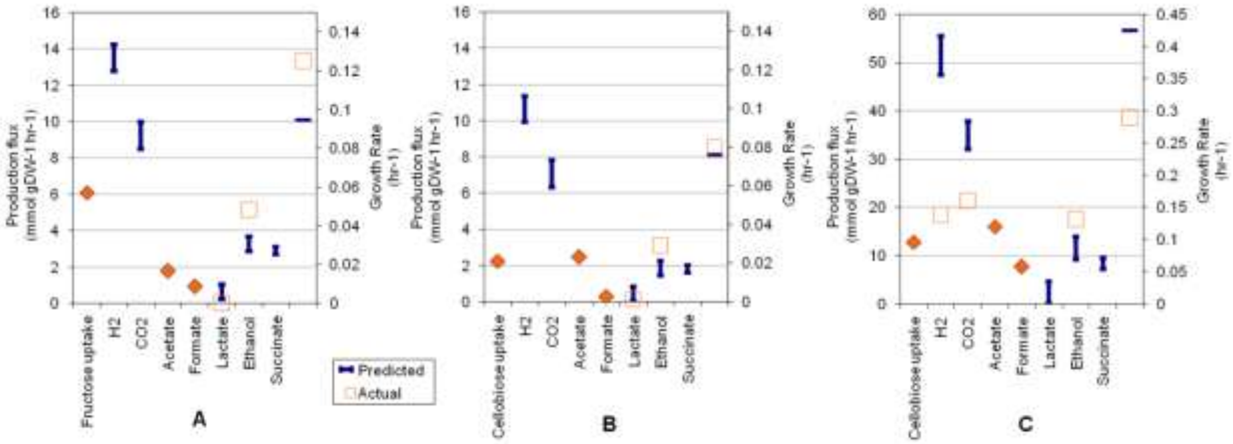


Figure 2: Comparison of model predictions to experimental observations (S. B. Roberts et al., 2010)**.** *C. thermocellum i*SR432 was used to simulate growth in multiple conditions. Actual (□) and predicted (▬) reaction flux rates are shown, and predicted fermentation product production rates are shown as ranges as determined by flux variability analysis (see II.B.4.c)). For each simulation, the boundary fluxes for cellobiose, acetate, and formate were constrained to match the measured fluxes during (A) chemostat growth on cellobiose and (B) fructose(Hogsett, 1995), and (C) batch growth on cellobiose(Rydzak et al., 2009).

www.manaraa.com

a) Continuous growth on cellobiose

Fermentation experiments in which *C. thermocellum* grew in continuous culture and consumed cellobiose at 2.2472 mmol / g-DW / hr (Hogsett, 1995) were the basis for testing model predictions of continuous growth on cellobiose. The experimental observations for growth rate, cellobiose uptake and major metabolic byproduct production rates can be found along with selected model predictions in Figure 2(A). In this case, the experimentally measured growth rate was 0.08 hr^{-1} (Hogsett, 1995). Based on experimental evidence that growth on cellobiose is associated with less cellulase activity (Demain, Newcomb, & Wu, 2005), we set the cellulosome production requirement for these simulations to zero. Applying one constraint, for cellobiose uptake flux, gave a predicted growth rate of ~0.094. Applying an additional constraint for acetate production resulted in a predicted growth rate of ~0.076 hr^{-1} and greatly improved the prediction for ethanol production. Fermentation products predicted by the model, but for which we had no experimental measurements, included CO_2, and H_2. To consider the effect of our computational representation of the cellulosome on growth predictions, we restored the cellulosome production requirement back into the objective function. This resulted in a decrease of ~4-17% in predicted growth rates, depending on which constraints were applied. This decrease in the predicted growth rate represents the metabolic impact (burden) of producing cellulosomes in *C. thermocellum*.

b) Continuous Growth on Fructose

In the case of continuous growth on fructose, the experimentally measured growth rate was 0.125 hr^{-1} (Hogsett, 1995). Experimental observations and major metabolic fermentation

www.manaraa.com

product productions rates for these conditions are in Figure 2(B). Based on experimental evidence that growth on fructose is associated with more cellulase activity (Demain et al., 2005), cellulosome production was required as part of these simulations. Applying a single constraint for fructose uptake flux resulted in a predicted growth rate of 0.123 hr^{-1}. Applying one further constraint for acetate production resulted in a predicted growth rate of 0.095 hr^{-1}. Applying constraints for acetate, lactate, ethanol, and formate production resulted in a predicted growth rate of 0.090 hr^{-1}. Removing the cellulosome production requirements resulted in an increase of ~4-8% in growth rates, depending on which constraints were applied.

c) Batch Growth on Cellobiose

Fermentation experiments in which *C. thermocellum* grew in batch culture and consumed cellobiose at 12.8 mmol / g-DW / hr (Rydzak et al., 2009) were the basis for testing model predictions of batch growth on cellobiose. The experimental observations for growth rate, cellobiose uptake and major metabolic product secretion rates can be found along with selected model predictions in Figure 2(C). When only the cellobiose uptake rate is specified, *i*SR432 predicts a maximum growth rate of 0.518 hr^{-1}, which correlates with high production of H_2, CO_2, and acetate and very low production of all other major metabolites. As other constraints are added to match observed fluxes, the maximum growth rate, as in other cases, tends to drop, and carbon flux is redistributed to other metabolites. When constraints for acetate, formate, ethanol, H_2, and CO_2 are added, flux balance analysis predicts a maximum growth rate of 0.387 hr^{-1} and very low or absent production of lactate and succinate. Furthermore, when succinate and lactate production are also constrained, excess carbon is

المنارة للاستشارات
www.manaraa.com

relieved by an increase in the production of aspartate to 8.48 mmol / gDW / hr. With these constraints, the maximum predicted growth rate is 0.380 hr-1, and as additional constraints are added to reduce all amino acid export fluxes to zero (not shown), the maximum growth rate gradually drops to below the observed value until in silico growth is impossible with no additional amino acid production.

3. *Gene Deletions: essentiality and effects on ethanol secretion*

We conducted comprehensive *in silico* single gene deletions with *i*SR432, using cellobiose as a carbon source and the other minimal media components (as described above), and constraining cellobiose uptake to its experimentally observed value for batch growth (12.8 mmol/gDW-hr, shown in Figure 2). Gene essentiality results are shown in Figure 1. In the case of growth on cellobiose, we found that 208 (36%) of *C. thermocellum* genes included in *i*SR432 were predicted to be essential. We also examined which subsystems of *i*SR432 contained the highest percentage of essential reactions ('vulnerable subsystems'). Among the most vulnerable subsystems are the cell envelope biosynthesis, nucleobase metabolism, and fatty acid metabolism.

المنارة للاستشارات

www.manaraa.com

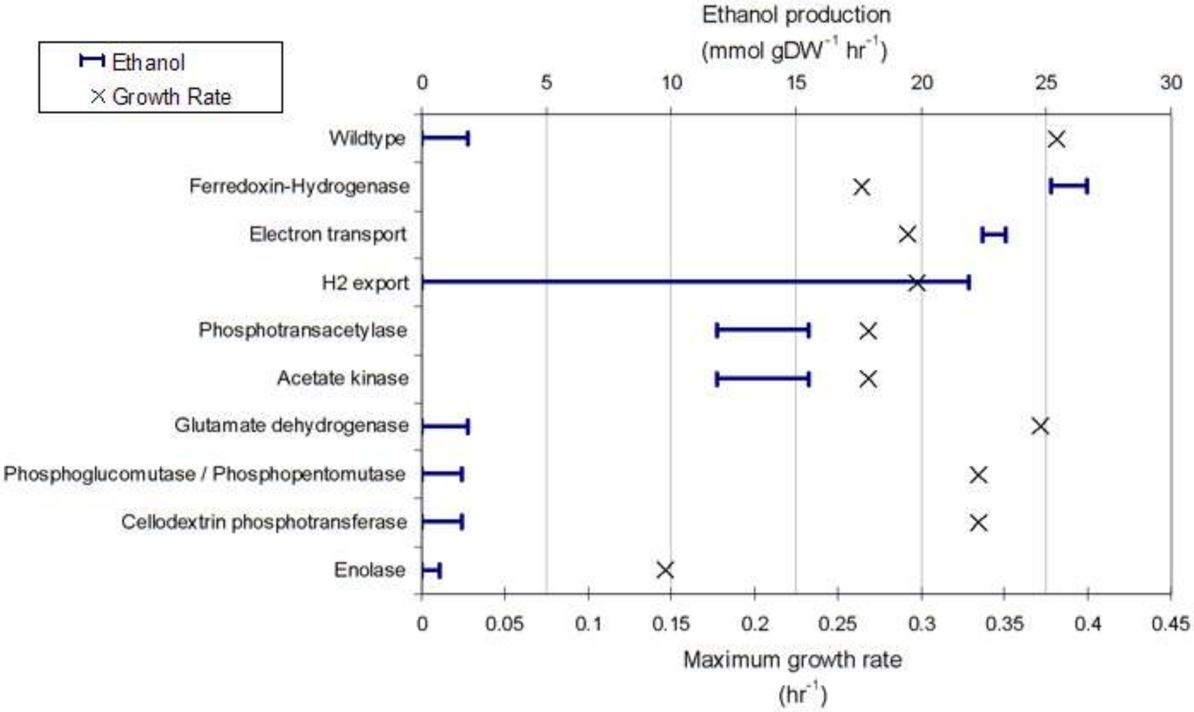


Figure 3: Single gene deletions associated with increased ethanol production (S. B. Roberts et al., 2010)**.** The predicted maximum growth rate (X) is shown for each simulated deletion scenario during growth on a cellobiose minimal medium with a measured cellobiose uptake rate of 12.8 mmol / g-DW / hr. Also shown is the range of ethanol production that can be achieved at the maximum growth rate. See the Appendix for details about the genes deleted and the affected reactions.

For each gene deletion predicted to be non-lethal, we conducted flux variability analysis to determine the effect of the deletion on the lower and upper bounds for ethanol secretion. Results are shown in Figure 3. There were nine single gene deletions (Cthe_1028, Cthe_1029, Cthe_2430, Cthe_2431, Cthe_2432, Cthe_2433, Cthe_2434, Cthe_2435, and Cthe_3003), associated with seven distinct model reactions, that were predicted to increase the upper bound on ethanol secretion relative to the wild-type ethanol secretion capabilities. Each of these was also predicted to increase the lower bound on ethanol secretion indicating that a deletion of one of these genes should force an increase in ethanol production. Deletion of gene Cthe_3003 (Coding for R_FDXHASE or Ferredoxin hydrogenase 1.12.7.2) was predicted to result

www.manaraa.com

in the greatest change that resulted in a ~15 fold increase in maximum ethanol secretion and a concomitant 30% decrease in biomass production, when compared to ethanol production at optimal growth rate of the wild-type strain. Genes predicted to increase ethanol secretion upon deletion were either involved in reactions affecting reduction-oxidation balance or in reactions producing acetate.

4. *Tradeoff of H_2 and ethanol production*

In examining the trends shown in Figure 2, we noted a general tradeoff between the flux on H_2 production and ethanol production. A number of studies (Brener & B. F. Johnson, 1984; Chinn, Nokes, & Strobel, 2007; Islam et al., 2006; Lamed, Lobos, & Su, 1988) have suggested that by thermodynamically restricting hydrogen escape, it is possible to increase the production of ethanol by *C. thermocellum*. To investigate this phenomenon, we tested the relationship between H_2 escape and the production of other fermentation products during batch growth on cellobiose. Over the entire ranges of hydrogen and ethanol production that would allow *in silico* growth, we constrained H_2 and ethanol escape and used FBA to determine the maximum growth rate. The results are displayed in Figure 4 and indicate that at the highest possible growth rate, ethanol production drops to zero while H_2 production yield is

$$\frac{61.85 \cdot mmol\_H_2\_produced/g-DW/hr}{12.80 \cdot mmol\_cellobiose\_consumed/g-DW/hr} = 4.832 \cdot (mol \cdot H_2/mol\_cellobiose).$$

www.manaraa.com

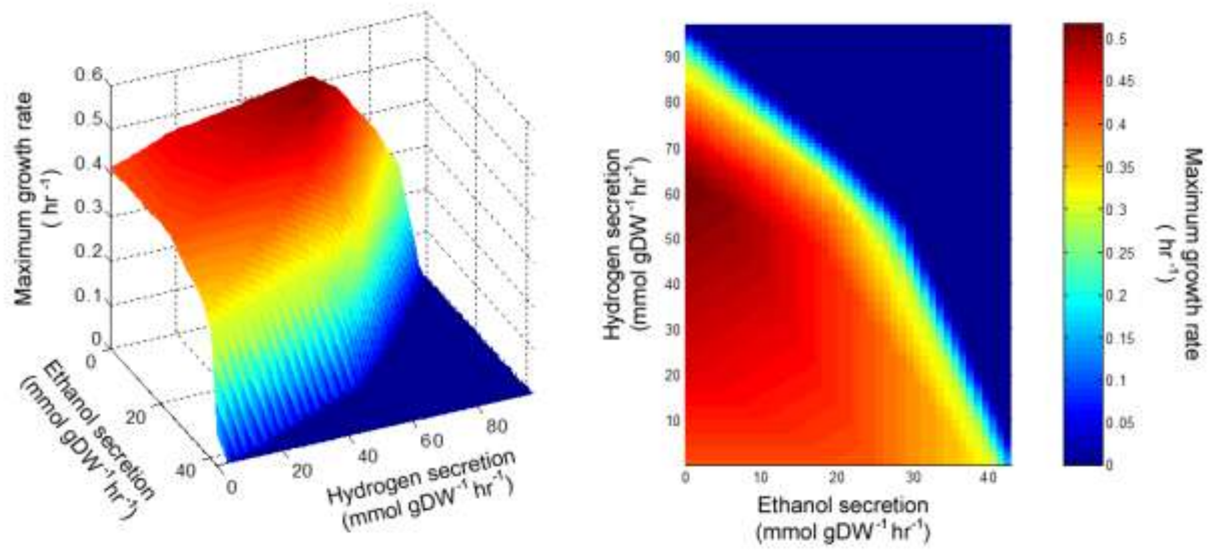


Figure 4: The metabolic solution space for growth of *C. thermocellum i*SR432 on cellobiose (S. B. Roberts et al., 2010)**.** The maximum growth rate was simulated across a two-dimensional surface of ethanol and hydrogen production rates, given a cellobiose uptake rate of 2.2472 mmol / g-DW / hour. The maximum growth rate achievable occurs when hydrogen production is at 61.85 mmol / g-DW / hr and ethanol production is at zero, reflecting the improved stoichiometric efficiency from using H_2 gas as a final electron acceptor rather than ethanol.

If hydrogen or ethanol secretion are constrained away from this global maximum, the *in silico* organism is forced to utilize less efficient pathways to varying degrees. For example, when ethanol and hydrogen escape fluxes are both forced to zero, the carbon flux to ethanol is rerouted at acetyl-CoA to acetate production, and since H_2 can no longer act as an electron acceptor for reduced ferredoxin produced by pyruvate:ferredoxin oxidoreductase (R_POR2_i; 1.2.7.1), acetyl-CoA is instead produced by pyruvate formate lyase (R_PFL; 2.3.1.54), resulting in a drastic increase in formate export (see the Appendix for the complete simulation results).

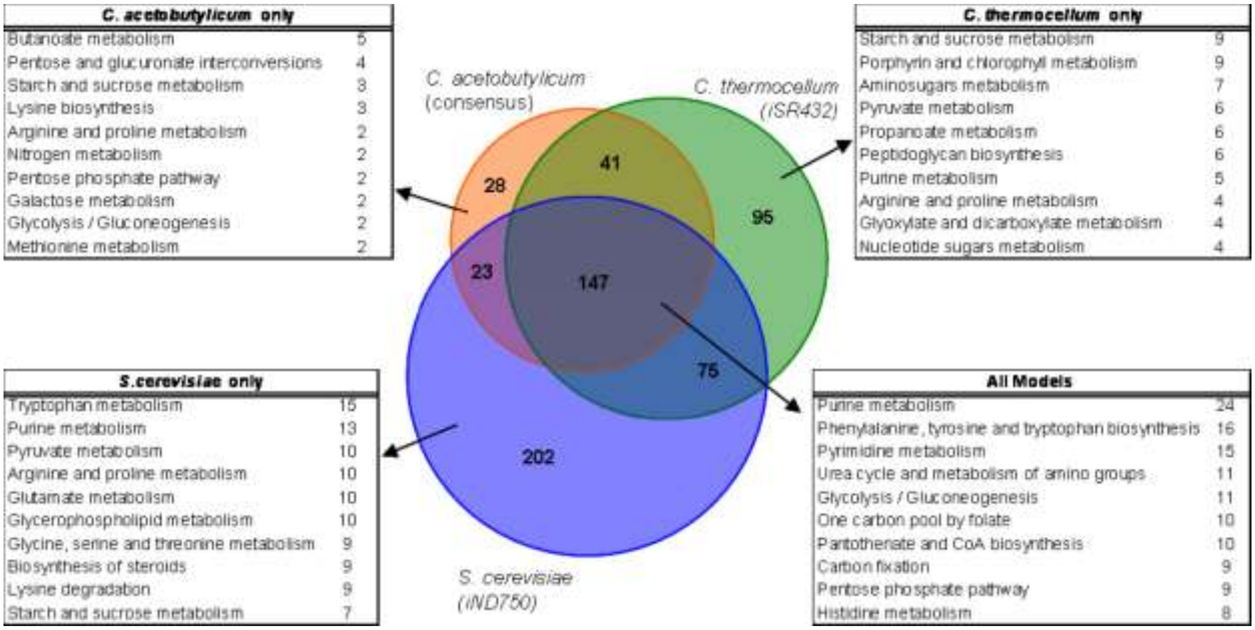
www.manaraa.com

5. *Comparative Analysis of metabolisms – compare to C. acetobutylicum and S. cerevisiae*

One consequence of constructing genome-scale models is the ability to obtain a comprehensive overview of an organism's metabolic network. Using our reconstructed network for *C. thermocellum*, we sought to computationally compare the metabolic network of *C. thermocellum* to a related, but butanol-producing Clostridia species (*C. acetobutylicum*) and a different ethanologenic microorganism (*S. cerevisiae*). We conducted our analysis by comparing the content of the metabolic models for these three organisms, shown in Figure 5. As described in Methods, we represented reaction content of the models by EC numbers. The EC numbers unique to *C. thermocellum* are commonly associated to pathways such as Starch and Sucrose Metabolism (especially due to enzymes involved in cellulose metabolism) and Porphyrin Metabolism (several enzymes related to Vitamin B12 metabolism and production of various porphyrinogens). EC numbers unique to *C. acetobutylicum* are commonly associated to pathways such as Butanoate Metabolism (CAC is the only one of the three organisms that natively produces butanol) and Pentose and Glucuronate Interconversions. EC numbers unique to *S. cerevisiae* are commonly associated to pathways such as Tryptophan Metabolism and Purine Metabolism. Finally, the EC numbers shared by all three models are commonly associated to pathways such as Purine and Pyrimidine Metabolism and aromatic amino acid biosynthesis. One reason that these particular pathways are well represented is likely because they are associated with many EC numbers, i.e., they are "big" subsystems. As expected, core pathways of metabolism, such as glycolysis and the pentose phosphate pathway, are also

المنارة للاستشارات
www.manaraa.com

among the most commonly represented in the set of EC numbers common to all three organisms.



C. acetobutylicum only	
Butanoate metabolism	5
Pentose and glucuronate interconversions	4
Starch and sucrose metabolism	3
Lysine biosynthesis	3
Arginine and proline metabolism	2
Nitrogen metabolism	2
Pentose phosphate pathway	2
Galactose metabolism	2
Glycolysis / Gluconeogenesis	2
Methionine metabolism	2

C. thermocellum only	
Starch and sucrose metabolism	9
Porphyrin and chlorophyll metabolism	9
Aminosugars metabolism	7
Pyruvate metabolism	6
Propanoate metabolism	6
Peptidoglycan biosynthesis	6
Purine metabolism	5
Arginine and proline metabolism	4
Glyoxylate and dicarboxylate metabolism	4
Nucleotide sugars metabolism	4

S.cerevisiae only	
Tryptophan metabolism	15
Purine metabolism	13
Pyruvate metabolism	10
Arginine and proline metabolism	10
Glutamate metabolism	10
Glycerophospholipid metabolism	10
Glycine, serine and threonine metabolism	9
Biosynthesis of steroids	9
Lysine degradation	9
Starch and sucrose metabolism	7

All Models	
Purine metabolism	24
Phenylalanine, tyrosine and tryptophan biosynthesis	16
Pyrimidine metabolism	15
Urea cycle and metabolism of amino groups	11
Glycolysis / Gluconeogenesis	11
One carbon pool by folate	10
Pantothenate and CoA biosynthesis	10
Carbon fixation	9
Pentose phosphate pathway	9
Histidine metabolism	8

Figure 5: Comparison of model content for three ethanologenic organisms (S. B. Roberts et al., 2010)**.** *C. thermocellum i*SR432, *C. acetobutylicum* consensus [*CacMBEL502* (J. Lee et al., 2008) and (Senger & Papoutsakis, 2008a)], and *Saccharomyces cerevisiae i*ND750 (Duarte et al., 2004) were found to exclusively represent 95, 28, and 202 EC numbers, respectively, and 147 EC numbers were shared among all three. The EC numbers were mapped to pathway names using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and the top ten most frequently occurring pathways for each of the exclusive lists and the combined list are shown along with the count of occurrences for each pathway.

6. Simulation of Alternative Media Formulations

Media formulations and growth conditions can significantly affect the fermentation characteristics of any microorganism. In light of this, we used *i*SR432 to make context-dependent predictions of growth and fermentation product secretion profiles. Specifically, the additions of 35 metabolites were systematically simulated, both individually and in

المنارة للاستشارات
www.manaraa.com

combinations of two, and flux variability analysis was performed to determine the resulting maximum growth rate and concomitant ethanol production range. This analysis was repeated for each of the reaction deletion strains that were shown to result in an increase in ethanol production (Figure 3). Figure 6 explores the predicted maximum ethanol yields when supplementing lactate, malate, or both lactate and malate to either the wild-type or gene deletion strains. Comprehensive results for all 35 tested metabolites are available in supplementary documentation. When these metabolites are added to the wildtype strain, there is no increase in the maximum ethanol yield, but in the deletion strains, these additions significantly improve ethanol production. When both lactate and malate are added to the deletion strains, maximum ethanol yield is increased by ~50% for the deletion strains affecting hydrogen production (ΔFDXHASE) or electron transport (ΔNFO) and ~140% for Δpta and Δack, which knock out acetate production. Other media additions increased ethanol production by thermodynamically blocking production of another competing metabolic byproduct, such as acetate, as is seen experimentally in a metabolic shift assay. For example, if acetate is supplied to the model, acetate production is blocked thermodynamically, and the maximum ethanol yield is increased 11 fold. For the purposes of the model, this effect is exactly the same as deleting the genes coding for acetate kinase (R_ACK) or phosphoacetyltransferase (R_PTA2). By combining the implementation of a single gene deletion and alternative media formulations, the maximum ethanol yield at optimal growth can be increased by as much as 35 fold.

www.manaraa.com

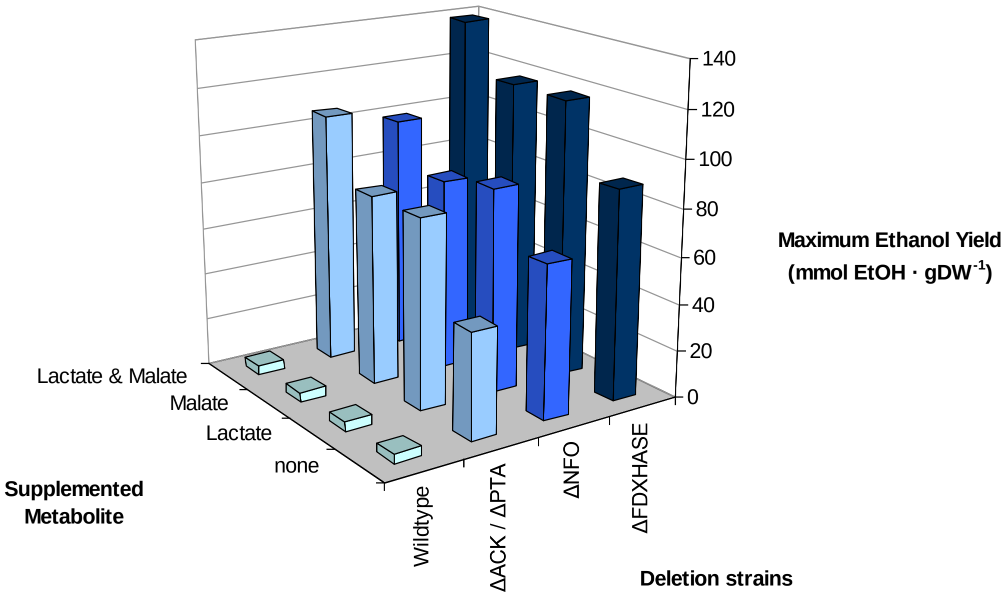


Figure 6: Alternative media formulations for single-reaction deletion strains (S. B. Roberts et al., 2010)**.** *C. thermocellum i*SR432 was used to simulate the addition of various potential media components individually and in pairs (see II.B.4.e) for details). The maximum possible ethanol yield per biomass is shown for four deletion strains during simulated growth on cellobiose and supplemented alternative carbon sources.

D. Modifications to the original *i*SR432 model

The development of any mathematical model is inherently an ongoing and iterative process, where changes to the model are made to accommodate new mechanistic knowledge or discrepancies between model predictions and experimental observations. This process gradually improves both the reliability of the model and our understanding of the studied system. Although the previous sections describe in detail this iterative process up to the publication of the original *i*SR432 model, a number of changes have been made to the *C. thermocellum* model since that time. This section details those changes and some of the physiological implications that they carry. The complete summary of changes to the model is in the appendix.

www.manaraa.com

1. *Pyruvate kinase*

During the initial construction of the model, we noted that there was no gene present in *C. thermocellum* coding for the enzyme pyruvate kinase, which is the typical means of converting phosphoenolpyruvate (PEP) to pyruvate (PYR) in glycolysis:

R_PYK: Pyruvate kinase | adp + h + pep --> atp + pyr

The conversion of PEP to PYR is a crucial step in the central carbon metabolism of any carbohydrate-utilizing organisms, and we had no available evidence of other present mechanisms for this conversion during the initial construction process. Therefore, we added this R_PYK to the model and used BLAST to search for potential genetic candidates coding for this enzyme, as detailed in section 0 above. Since the original publication, enzymology experiments by our collaborator Richard Sparling at the University of Manitoba have been unable to detect any pyruvate kinase activity. Instead, Sparling et al. have detected relatively high protein levels for the enzymes pyruvate kinase (PPDK), phosphoenolpyruvate carboxykinase (PEPCK), oxaloacetate decarboxylase (OAADC), malate dehydrogenase (MDH), and malic enzyme (ME) (personal communication, May 2011). The proposed reactions are represented in the model as

R_PEPCK: Cthe_2874 | [c] : atp + oaa <==> co2 + adp + pep

R_OAAdecarb: Cthe_0699 and Cthe_0700 and Cthe_0701 | [c] : oaa <==> pyr + co2

R_MDH: Cthe_1053 or Cthe_0345 | [c] : mal-l + nad <==> h + nadh + oaa

www.manaraa.com

R_ME1: Cthe_0344 or (Cthe_0699 and Cthe_0700 and Cthe_0701) | [c] : mal-l + nad --> co2 + nadh + pyr

All of these reactions except for PPDK were in the *i*SR432 model, but R_PEPCK was irreversible in the direction of PEP production, based on thermodynamic constraints set in other models. In the *i*SR432 model, therefore, deletion of the PYK reaction resulted in the only means to connect the upper part of glycolysis to pyruvate being through a circuitous and unlikely series of reactions beginning at 3-phosphoglycerate and branching through the serine and cysteine pathways. The replacement of R_PYK with a reversible R_PEPCK does not affect the net balance of the model around this conversion step, but it does affect the results of any potential genetic manipulations along these routes by shifting metabolic flux to PEPCK and either OAADC or MDH and ME.

2. *Bifurcating hydrogenase and electron transport*

An increasing number of research groups have recently noted the wide diversity present in different hydrogenase enzymes across several species. Specifically, Herrmann et al. noted that the hydrogenase present in *Thermotoga maritime* is dependent on both ferredoxin and NADH concurrently, and they propose that this represents a new type of redox energy conservation in anaerobic bacteria. This hydrogenase utilizes the thermodynamic favorability of ferredoxin (E_0' = –420 mV) oxidation to drive the unfavorable oxidation of NADH (E_0' = –320 mV) while coupling electron flow to hydrogen production (Herrmann, Jayamani, Mai, & Buckel, 2008). Recent comparative genomic work has further explored this diversity in the genus Clostridium, and has suggested that the hydrogenase gene cluster at Cthe_0426 operates in this manner

www.manaraa.com

(Calusinska, Happe, Joris, & Wilmotte, 2010). As a result of this new genetic evidence, the reaction R_BIFHASE has been added:

R_BIFHASE: Cthe_0426 | [c] : 3 h + 2 fdred + nadh <==> fdox + nad + 2 h2

Additionally, conversations with collaborators in the research groups of Richard Sparling and David Levin have led us to correct some of the NADH:Ferredoxin oxidoreductase reactions in the model based on new interpretations of the genome annotation. The following reactions were added:

R_NADPOX: Cthe_3003 | [c] : h2 + nadp <==> h + nadph

R_NADOX: (Cthe_0341 and Cthe_0342) or (Cthe_0429 and Cthe_0430) | [c] : h2 + nad <==> h + nadh

These reactions replaced the following incorrectly annotated reactions:

R_NFO_Na: (Cthe_2430 and Cthe_2431 and Cthe_2432 and Cthe_2433 and Cthe_2434 and Cthe_2435) | nadh[c] + fdox[c] + na1[e] <==> nad[c] + fdred[c] + na1[c]

R_NFO_h: (Cthe_2430 and Cthe_2431 and Cthe_2432 and Cthe_2433 and Cthe_2434 and Cthe_2435) | nadh[c] + fdox[c] + h[e] <==> nad[c] + fdred[c] + h[c]

www.manaraa.com

Additionally, the GPR rules for ferredoxin hydrogenase (R_FDXHASE) were updated from (Cthe_3003) to (Cthe_3019 and Cthe_3020 and Cthe_3021 and Cthe_3022 and Cthe_3023 and Cthe_3024).

E. Discussion

This chapter discussed the construction and validation of the initial version of the genome-scale *in silico* metabolic for *Clostridium thermocellum*. The resulting model was shown to be consistent with available experimental observations for substrate utilization and byproduct production. As described in the previous section, model curation is an ongoing process throughout the model's lifetime. I have described here some of the most fundamental changes that have been made to the model since its initial development.

المنارة للاستشارات
www.manaraa.com

III. WHOLE TRANSCRIPTOME SEQUENCING OF *C. THERMOCELLUM* ON CELLOBIOSE

A. Introduction

One of the most promising advantages of genome-scale metabolic reconstructions is that they can be used as a platform for the integration of multi-level “omics” datasets. This would allow analysis and interpretation of these huge information sources within the context of the metabolic network, which is necessary to fully understand physiological implications of metabolic phenomena. With this in mind, we designed and executed a simple whole-transcriptome sequencing (RNAseq) experiment. The goals of this experiment were specifically to investigate global gene expression in *C. thermocellum* and, more broadly, to test potential methodologies for integrating this type of dataset with a metabolic model. This chapter will focus on the first goal by examining more traditional bioinformatics analysis of the RNAseq dataset, and the following chapter will further explore multiple means to combine this and other omics datasets with the metabolic reconstruction in order to improve the quality of model predictions and the reliability of computational strain designs. In this chapter, I present custom bioinformatics analysis techniques for utilizing RNAseq datasets to quantitatively and globally compare mRNA transcript levels detect transcription start sites and cotranscription of neighboring genes. These analyses are shown to be useful for planning genetic manipulations for metabolic engineering.

www.manaraa.com

B. Background

1. Types of mRNA-level "omics" datasets

Transcription-level datasets can provide useful information about the biochemical mechanisms behind an observed cellular phenotype and about some of the regulatory controls used by the cell. The first genome-scale RNA measurement procedures were DNA microarrays, first described by Uwe Maskos and Edwin Southern (Maskos & Southern, 1992). This technology has since advanced in coverage, accuracy, and affordability, and microarrays have become a core toolset for nearly every area of systems biology. Although technical details between microarray platforms and application areas vary considerably, the basic premise remains constant. DNA oligomers of known sequence are spotted on microarray plates or chips in known locations. Mixed RNA samples are purified of DNA and protein contaminants, and the RNA is converted to cDNA using a reverse transcription reaction. The resulting cDNA mixture is more chemically stable than the RNA template, and fluorescent probes are added to all cDNA molecules, independent of sequence. This mixture is then allowed to anneal to the spotted DNA oligomers, and wash steps are employed to promote good binding specificity. The fluorescent signal at each DNA spot then correlates to a relative abundance of cDNA that is complimentary to the spotted oligomer. A combination of signal processing and statistical normalization can be used to generate average relative signal strengths for each measured gene. The relative nature of the resulting signal requires a careful use of controls as a basis for comparison, and most studies use two separate fluorescent probes that can be read simultaneously to avoid the significant variation that can occur between experimental replicates. Furthermore, the experimental sample space is limited to the DNA oligomers that are spotted on the array, so

www.manaraa.com

researchers must rely on *a priori* predictions about which RNA fragments are expected. Still, microarrays remain a popular choice for transcriptomics because many platforms are relatively cost-effective, especially when examining transcription in many different conditions using the same set of DNA oligomers.

2. Whole-transcriptome sequencing analysis (RNAseq)

In the early and mid-1990's, concurrently with microarray development, a number of sequence-based methods were developed to measure intracellular mRNA levels. The most basic approach was to perform Sanger sequencing of cDNA (Gerhard et al., 2004; Provot, Persuy, & Mercier, 1989), and the low throughput of this approach was later improved using tag-based sequencing techniques like Expressed Sequence Tags (EST), Serial Analysis of Gene Expression (SAGE) (Harbers & Carninci, 2005; Sun et al., 2004) and its derivatives. More recently, with the exponential advance in high-throughput DNA sequencing technology, deep sequencing of all RNA molecules in a cell has become feasible in many circumstances (Marioni, Mason, Mane, Stephens, & Gilad, 2008; Z. Wang, Gerstein, & Snyder, 2009; Wilhelm & Landry, 2009). RNAseq employs "next generation" sequencing platforms like the Illumina IG and Roche 454 Life Science systems to generate sequence data for thousands of cDNA fragments. In most cases, coverage is high enough that this analysis is qualitative, in contrast to transcriptomics obtained from Sanger sequencing. Because all cDNA molecules can be sequenced uniformly, this analysis requires no preliminary knowledge of the genome sequence or of how and where exons are spliced together, making RNAseq an extremely powerful tool for studying microbial communities and eukaryotes with complex transcriptomes. These technological advantages can

www.manaraa.com

often outweigh the greatly increased per-sample cost that is incumbent with next-generation sequencing.

C. Methods

1. Sample Preparation

Clostridium thermocellum ATCC 27405 cultures were maintained in batch culture at 55° C in anaerobic serum bottle culture as described in section II.B.2 above. Cultures were grown in 50mL of MJ minimal medium (E. A. Johnson et al., 1981) with 5 g/L cellobiose as the carbon source. Anaerobic conditions were created and maintained by culturing in 100mL serum bottles with the headspace purged and replaced by ultra-high purity nitrogen gas. Cultures were passed as a 0.5% inoculum during mid-log phase (OD_{600} between 0.6 and 0.9) for at least two days prior to mRNA isolation.

RNA was isolated from mid-log cultures of *C. thermocellum* using QIAGEN RNA Protect reagent in combination with the QIAGEN RNeasy Mini kit with on-column DNAse treatment. The resulting total RNA samples were then pooled and enriched for mRNA using the MICROBExpress kit from Ambion, Inc. Denaturing RNA electrophoresis gels were used at all steps to monitor the quality of mRNA. Complimentary DNA was synthesized from the enriched mRNA using the Bioline Tetro cDNA synthesis kit with random hexamer primers. The resulting cDNA was sequenced using the Roche 454 sequencer.

www.manaraa.com

2. *Analysis of raw data*

The raw sequences of cDNA fragments in FASTA format were aligned to all RefSeq genomic BLAST databases using megaBLAST. These queries were subsequently filtered using a Python script as follows. For each query id (i.e. each read sequence from the 454), the best alignment as judged by the BLAST bit score was kept, and all other suggested alignments were discarded. Additionally, for any queries for which there were multiple alignments that scored equally well, the entire query was discarded. These queries are available in the Appendix. The queries were then filtered according to the rules

min length = 30

max e-value = 0.001

min % identity = 95.00

Subject id = NC_009012.1 (*C. thermocellum* RefSeq code)

Queries that did not meet these requirements were discarded from further analyses. In all, 16% of total reads aligned optimally in multiple locations, 1% did not meet the minimum agreement, 27% did not match to within a single gene locus, 0.1% did not match to *C. thermocellum*, and 0.007% did not meet the minimal read length. 55% of total reads were accepted for further analysis. The remaining reads were compared to the IMG annotation to assign each read to a gene locus. Reads not located entirely within any one gene locus were not counted. The results from the remaining queries are on the sheet 'Main Out' listed by gene locus and sorted by the

المنارة للاستشارات
www.manaraa.com

number of reads mapping to that locus. This analysis resulted in a discrete count of 'hits' for each gene locus, and the relative expression of each gene was calculated by normalizing this count to the gene length.

D. Results

1. Distribution of gene expression levels

A whole-transcriptome sequencing (RNAseq) experiment was carried out to investigate global gene expression in *C. thermocellum* during exponential growth on cellobiose minimal media. RNA sequencing using the Roche 454 technology resulted in over 230,000 unique reads with an average length of 367bp. These reads were blasted against the RefSeq database (Pruitt, Tatusova, & Maglott, 2007) using MegaBLAST (Z. Zhang, Schwartz, L Wagner, & W. Miller, n d) with default parameters. After processing and filtering the best alignments as described in Materials and Methods, a total of 128,114 unique reads were successfully matched to a single gene locus on the *C. thermocellum* genome. This analysis produced a normalized expression score for each gene locus representing the number of mRNA fragments that were detected and aligned within that locus, normalized to gene length. We found that gene expression frequency had a log-log distribution across the genome, as shown in Figure 7.

www.manaraa.com

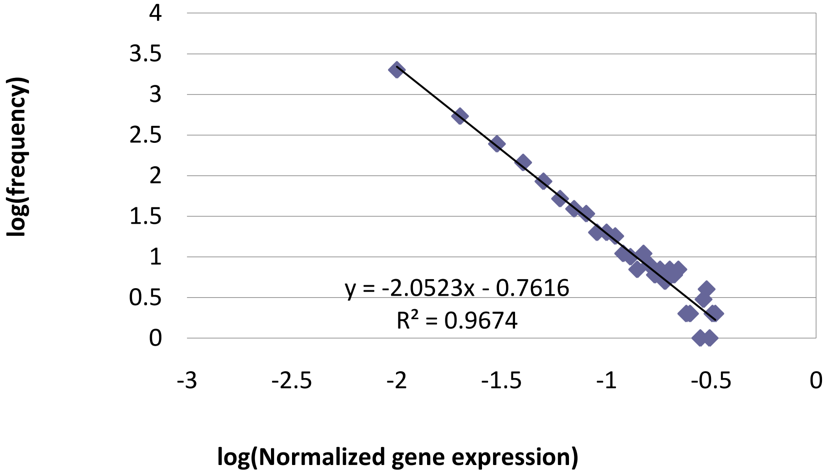


Figure 7: Log-log frequency plot for normalized RNAseq analysis of cellobiose-grown *Clostridium thermocellum*

www.manaraa.com

Table 3 lists a selection of the most highly-expressed genes for this growth environment that are directly relevant to cellulose utilization and fermentation. It is notable that all twelve of the listed reactions are within the top 32 most transcribed genes, and many of the other top genes are related to cellular maintenance processes (RNA polymerase, ribosomal proteins, elongation factors, etc.). Even during growth on cellobiose where cellulosome expression has been shown to be down-regulated, the structural cellulosome anchoring proteins are still very highly expressed. The PAS/PAC sensor protein coded by Cthe_0426 is annotated in the SEED database as a Fe-containing hydrogenase. The PAS domain has been shown to be an internal sensor of redox potential in the cell (Taylor & Zhulin, 1999), and this enzyme likely plays a vital role in regulating electron flow to hydrogen.

المنارة للاستشارات
www.manaraa.com

Table 3: Most highly-expressed genes with direct relevance to fermentation of cellulose to ethanol

Gene locus	Annotation	Gene Length	Unique Reads	Normalized Hit Count
Cthe_3078	cellulosome anchoring protein, cohesin region	6941	10580	1.524
Cthe_0432	iron-containing alcohol dehydrogenase (EC:1.2.1.10)	2621	3768	1.438
Cthe_3077	cellulosome anchoring protein, cohesin region	5561	7082	1.274
Cthe_0426	putative PAS/PAC sensor protein	1670	1930	1.156
Cthe_0429	NADH dehydrogenase (quinone) (EC:1.6.99.5)	1874	1558	0.831
Cthe_3079	cellulosome anchoring protein, cohesin region	2066	1098	0.531
Cthe_0505	formate acetyltransferase (EC:2.3.1.54)	2228	1115	0.500
Cthe_0412	glycoside hydrolase, family 9	2687	1294	0.482
Cthe_0389	PfkB	971	428	0.441
Cthe_0430	hydrogenase, Fe-only	1700	668	0.393
Cthe_3091	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, beta subunit	557	196	0.352
Cthe_2392	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	1184	412	0.348

2. *Operon structure*

One distinct advantage of RNAseq experiments over microarray experiments is that detection is not limited to the regions of the genome included in the DNA microarray probes. This means that there is no extra cost to look for transcription in non-coding or non-annotated regions, including the regions between coding sequences within a single operon. By looking for mRNA connections between coding sequences, it is possible to experimentally determine operon structure using a very high-throughput approach. Although many of the operons in *C.*

www.manaraa.com

thermocellum could be guessed based on functional similarity of neighboring genes (see Figure 8 A), only a few of its operons have been experimentally validated (Newcomb, C.-Y. Chen, & Wu, 2007; Newcomb, Millen, C.-Y. Chen, & Wu, 2011).

Figure 8 provides an example of how an RNAseq experiment can greatly improve the quality of functional annotation and prediction of operon structure. The genes in the positional cluster from Cthe_0332 – Cthe_0342, which are described in Table 4, are close neighbors on the same strand on the *C. thermocellum* genome, and it is possible that some or all of them are co-transcribed. As shown in part A, the Integrated Microbial Genomes (IMG) Genome Viewer suggests that the entire region may be co-transcribed based on the functional evidence that the first gene – Cthe_0332, a uridine kinase – is annotated in the same KEGG Pathway as the last – Cthe_0342, an NAD(P)-dependent hydrogenase. In contrast, the RNAseq operon analysis shown in part B only finds linkages beginning at Cthe_0333, but those mRNA links continue through Cthe_0342. The TSS score and RNAseq coverage clearly show a significant change in the transcription pattern starting near the 5' end of Cthe_0338. This case study emphasizes the value of integrating diverse molecular and computational analyses to develop a clear picture of molecular behavior. One possible explanation is that all of the genes from Cthe_0333 through Cthe_0342 are transcribed as a single operon, but that there exists an alternate TSS at Cthe_0338 that is separately regulated. Alternatively, there may simply be a leaky terminator at the end of Cthe_0337 that stochastically permits a fraction of RNA polymerase molecules to continue to the following genes.

المنارة للاستشارات

www.manaraa.com

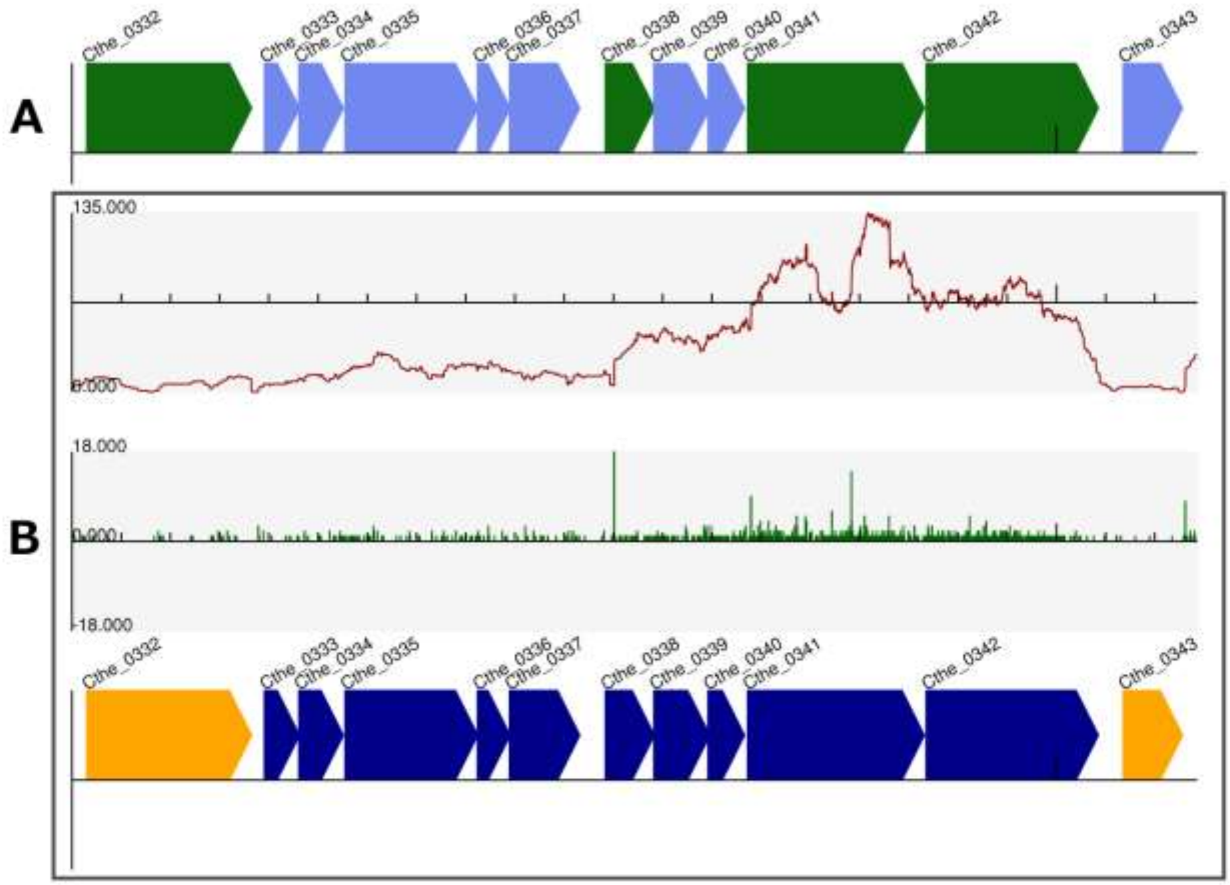


Figure 8: Prediction of co-transcription in a hydrogenase gene cluster. The automated functional gene cluster prediction by the Integrated Microbial Genomes site is shown in (A). Genes Cthe_0332, Cthe_0338, Cthe_0341, and Cthe_0342 are highlighted green because they are members of the same positional cluster annotated with functions within the same KEGG Pathway. RNAseq evidence (B) offers experimental evidence for clustering Cthe_0338-Cthe_0342 as an operon, where the red trace is overall mRNA coverage, the green bars are the positional TSS score, and where the linked genes are drawn in blue and unlinked genes are shown in orange.

Table 4: Genes in a potential hydrogenase operon

Gene locus	Annotation	Gene Length	Unique Reads	Normalized Hit Count
Cthe_0332	phosphoribulokinase/uridine kinase	1673	22	0.013
Cthe_0333	hypothetical protein	353	2	0.006
Cthe_0334	putative anti-sigma regulatory factor, serine/threonine protein kinase	449	5	0.011
Cthe_0335	hydrogenase large subunit-like protein	1346	46	0.034

المنارة للاستشارات
www.manaraa.com

Cthe_0336	DRTGG domain containing protein	329	2	0.006
Cthe_0337	PHP-like protein	710	15	0.021
Cthe_0338	NADH-quinone oxidoreductase, E subunit	497	19	0.038
Cthe_0339	histidine kinase	566	18	0.032
Cthe_0340	ferredoxin	368	5	0.014
Cthe_0341	NADH dehydrogenase (quinone) (EC:1.6.99.5)	1793	371	0.207
Cthe_0342	hydrogenase, Fe-only	1748	257	0.147

3. *Transcriptional start sites*

Since the reads detected using the 454 represent intact mRNA molecules, it is conceivable that we could use a bioinformatics approach to detect the locations of gene and operon transcription start sites (TSS). Previously, Cho et al demonstrated this approach in examining the transcription unit architecture of *E. coli*, but that study utilized a wide variety of omics-scale datasets, including a modified RNAseq analysis in which RNA adapters were added to the mRNA molecules before the reverse-transcription step. Additionally, results were filtered based on agreement with RNA polymerase binding locations that were determined using costly ChIP-Chip analysis (Cho et al., 2009). These methods allowed a clean and reliable detection of the 5' end of the RNA molecule and resulted in remarkable accuracy in TSS detection, but the additional steps required are cost-prohibitive for many research groups. To test the use of a more general 454 deep sequencing approach for TSS determination we attempted to quantify 5' and 3' RNA ends in the aligned RNAseq data from *C. thermocellum*. This assigned an "end count" to each location in the genome. To provide a background context for each location's count, we defined

www.manaraa.com

a location's "neighborhood" based on its placement within a gene or operon. For all locations outside of any gene, the neighborhood was defined as the surrounding 100bp region. The end count for each was location was then normalized to the average RNAseq coverage within the location's neighborhood to yield a TSS score. The histogram in Figure 9 shows the distribution of scores from all locations in the genome. The relatively large peak in frequency around a score of 2 was found to be a numerical artifact, occurring wherever there was a single alignment in an otherwise empty region. As a result, we found that filtering each location by TSS score ≥ 2.1 and a background RNAseq coverage ≥ 1 resulted in a list of relatively feasible putative TSS's.

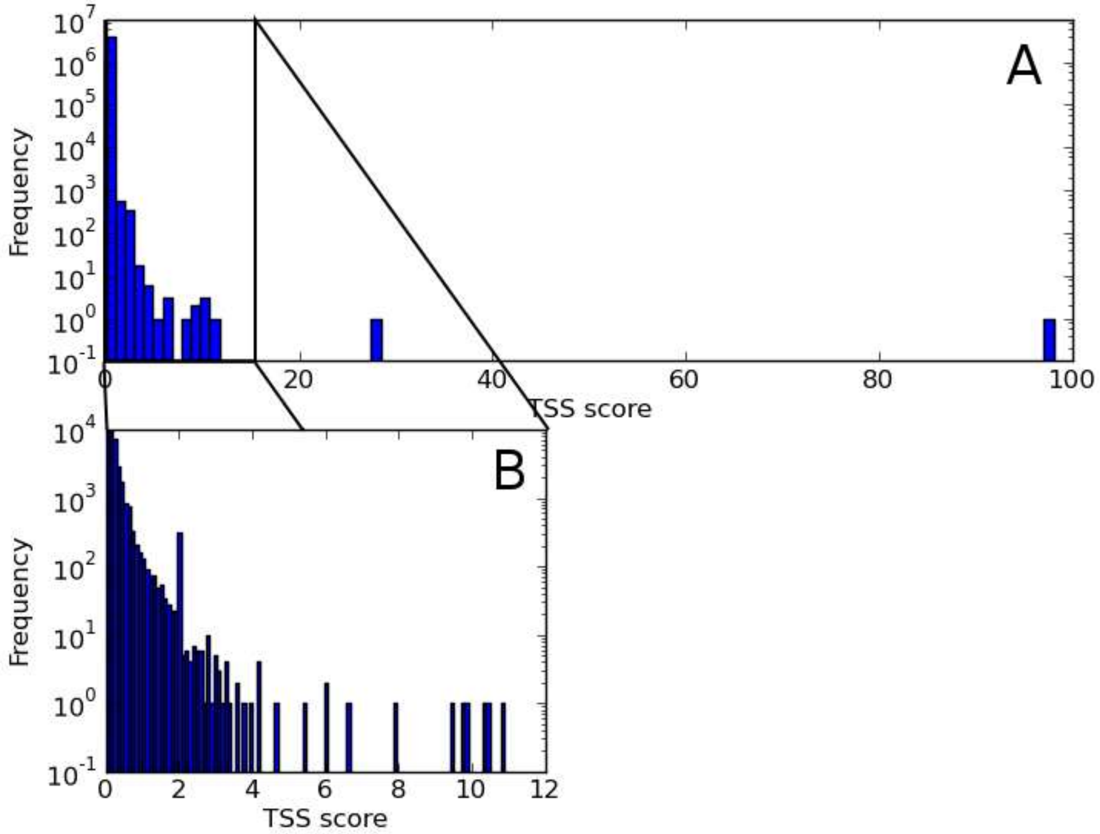


Figure 9: Transcription Start Site (TSS) Score frequency chart. TSS score was calculated by dividing the mRNA "end count" at each location by the average mRNA coverage at the

www.manaraa.com

location's neighborhood. Graph A covers the entire range of scores, while graph B covers a smaller range at a higher resolution.

The resulting list of 22 locations is listed in the appendix, and includes the TSS of the RNA polymerase operon, shown in Figure 10. This operon has many of the genes used in gene expression, with two beta-subunits of the DNA-directed RNA polymerase, three ribosomal proteins, and two translation elongation factors. These genes were found to be co-transcribed using the methods described in 0 above.

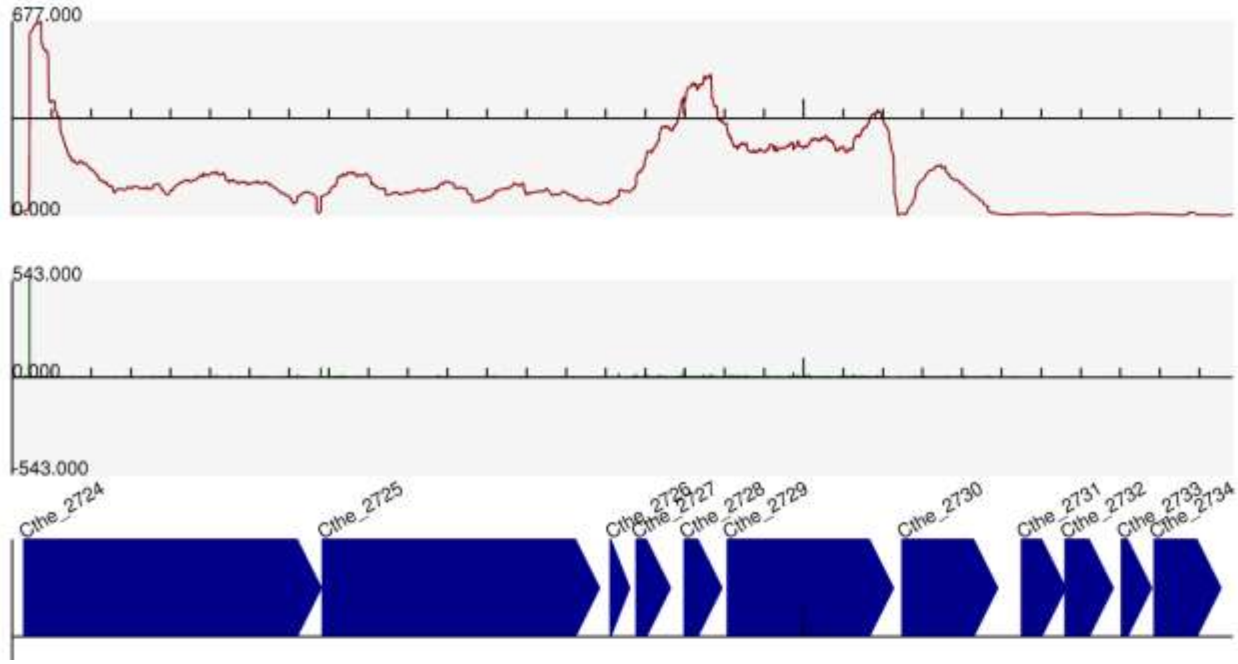


Figure 10: RNA polymerase operon expression in *Clostridium thermocellum*. The TSS count (dark green bars), mRNA coverage (red line), and annotated gene features (blue arrows) are shown for the gene loci Cthe_2724 through Cthe_2734. These genes were determined to be co-expressed as an operon using the methods described in III.D.2 above.

The list also includes start sites from two different operons containing various ribosomal RNA and tRNA subunits. It is also worth noting that not all sites scoring highly as a TSS match existing annotations as expected. For example, as shown in Figure 11, the occurring at position 3722380

المنارة للاستشارات
www.manaraa.com

in the first gene of operon Cthe_3149 does not appear near the 5’ end of the annotated gene aminoacyl-histidine dipeptidase. Instead, the putative TSS occurs near the 3’ end of the gene, and mRNA coverage is high downstream of that point across the untranslated region connecting Cthe_3149 and Cthe_3150 (adenosylcobinamide-phosphate synthase). If the alignments to this region are correct, it is possible that RNA polymerase is mispriming at this location, resulting in an artificially high coverage at this location. The enzymes coded by genes 3149 and 3150 appear to be functionally unrelated, and it may be that the operon linkage predicted here by the method described above is artificially affected by this anomalous transcribed region. This demonstrates one of the limitations of using a single RNAseq dataset for detection of operon and TSS locations in an automated fashion and highlights the importance of maintaining data provenance in any bioinformatics analysis. For example, ChIP-Chip analysis could quickly determine whether RNA polymerase indeed binds near the suggested start site, which would suggest that this region is indeed transcribed independently. The resulting mRNA may have a regulatory role by binding competitively to promoter regions elsewhere in the genome. If RNA polymerase were not found to bind here, another explanation might be that the suggested TSS is actually at a ribonuclease degradation site. The distinction between these different cases would be valuable input for reducing trial and error in attempts at genetic engineering in this region.

المنارة للاستشارات
www.manaraa.com

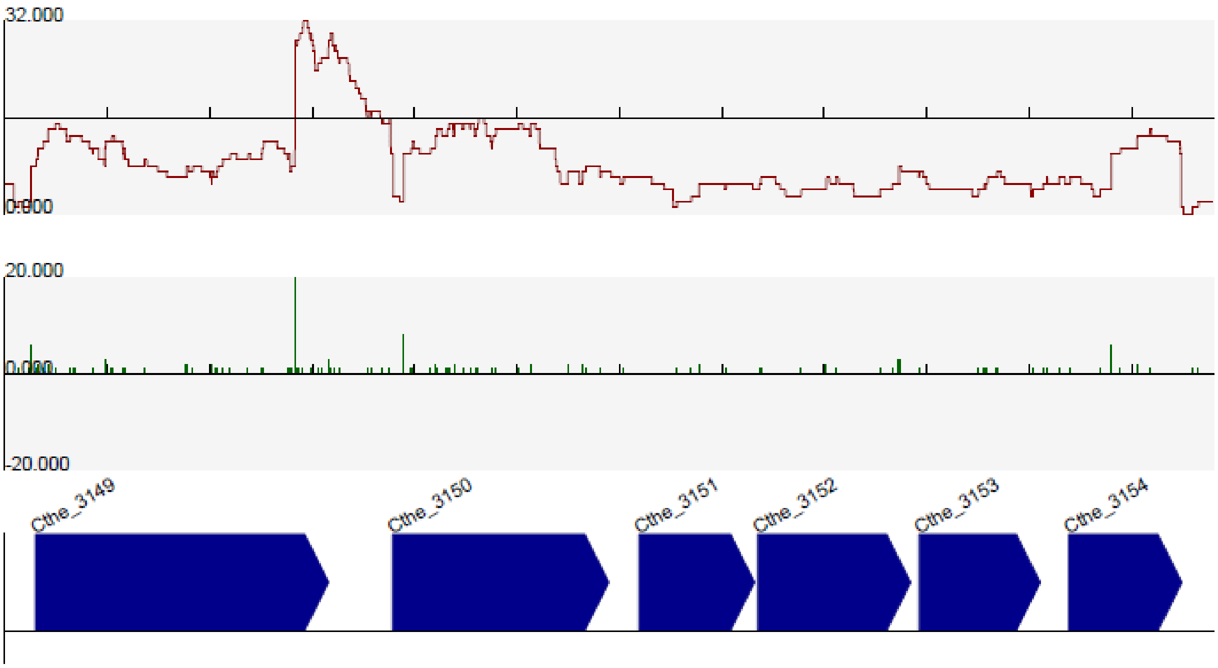


Figure 11: Possible RNA polymerase mispriming in Cthe_3149. The TSS count (dark green bars), mRNA coverage (red line), and annotated gene features (blue arrows) are shown for the gene loci Cthe_3149 through Cthe_3154. These genes were determined to be co-expressed as an operon using the methods described in III.D.2 above, but it is possible that the connection between 3149 and 3150 is only due to the aberrant transcription of the non-coding region between the two.

E. Discussion

Bioinformatics analysis of gene expression levels, co-transcription of neighboring genes, and putative transcription start sites (TSS) in this chapter illustrate the utility of RNAseq analysis for generating a huge amount of usable information from even a single experimental dataset. This study has resulted in the prediction of 686 operons and 22 TSS's with high confidence. Moreover, absolute presence/absence calls for all of the metabolic genes in the genome will prove to be very valuable for our understanding of how metabolism is operating on a genome scale, as detailed in the next chapter.

المنارة للاستشارات
www.manaraa.com

However, it is clear that further quantitative interpretation of many of the observed behaviors would depend on the integration of further omics datasets, including both RNAseq studies at different conditions and omics studies at the protein and metabolite levels. For example, by combining RNAseq studies at other growth conditions or time points with metabolomics data at the same conditions, it should be possible to greatly improve the extent of predicted operon coverage and to observe transcription-level regulatory changes that occur in different environments or at different time-points. This knowledge could be used to detect global regulatory networks in *C. thermocellum*, which would be incredibly valuable for engineering future strains and fermentation processes. Likewise, TSS prediction could benefit greatly through the application of 5’ end tags to mRNA prior to sequencing, as was used in (B.-K. Cho et al., 2009). These tags would distinguish the 5’ ends and protect them from degradation, thus greatly increasing the reliability of TSS predictions.

المنارة للاستشارات
www.manaraa.com

IV. COMPUTATIONAL INTEGRATION OF TRANSCRIPTOME AND PROTEOME WITH THE GENOME-SCALE MODEL OF *C. THERMOCELLUM*

A. Introduction

Physiological interpretation of bioinformatics datasets like the RNAseq and proteomics remains as an unsolved problem in modern biology. Particularly for metabolic phenomena, making a connection between mRNA or protein measurements and phenotype is difficult or impossible without knowledge of the metabolic network. *In silico* metabolic reconstructions can provide a platform for integration of this data for better physiological interpretation. The most direct method for incorporation of this data is the application of Boolean constraints to the reaction list where relevant enzyme mRNA or protein is not detected (Akesson, Förster, & Nielsen, 2004; Colijn et al., 2009). This method trims the solution space, thereby improving predictions using methods like FBA and FVA. In this section, I will describe the application of a recently described optimization procedure (Shlomi, Cabili, Herrgard, Palsson, & Ruppin, 2008) that generates flux predictions based on a transcriptomic dataset. In contrast to previous methods, this protocol can operate independently of a biomass objective and is therefore especially relevant for cases where non-optimal growth is likely. Application of this protocol to the *C. thermocellum* model was shown to be more effective than the previously described methods at accurately predicting fermentation behavior (Gowen & Fong, 2010). Further to that publication, I discuss here an additional step of incorporating proteomics data, both alone and in combination with the RNAseq data.

www.manaraa.com

B. Methods

1. Boolean restriction of solution space

Other methodologies that have incorporated gene expression data with metabolic models (Akesson et al., 2004; Colijn et al., 2009) have utilized gene expression data to specify constraints on reaction fluxes on a gene-by-gene basis. Following the method developed by Akesson et al. (Akesson et al., 2004), any absent gene (normalized expression of zero) will force the flux through the corresponding reaction to also be zero. Flux balance analysis or other analyses can then be performed as normal on the newly restricted metabolic network.

2. Ruppin algorithm

The genome-scale *in silico* metabolic reconstruction, *i*SR432, of *Clostridium thermocellum* (S. B. Roberts et al., 2010) accounts for 432 genes, 577 reactions, and 525 intracellular metabolites. This model was analyzed using the methods described by (Shlomi et al., 2008), in which the Boolean mapping from genes to reactions in *i*SR432 was used to translate gene expression data into predicted reaction activity states. Based on the relative gene expression levels calculated from RNAseq data, gene expression states were determined according to the following rules:

$$gene\ state = \begin{cases} -1 & g = 0 \\ 0 & 0 < g \le \gamma \\ 1 & \gamma < g \end{cases}$$

where g is the normalized expression level as determined by RNAseq and γ is the threshold above which a gene is called 'on.' The resulting gene states were then mapped using the gene-

www.manaraa.com

protein-reaction (GPR) relationships in *i*SR432 to generate lists of reactions predicted to be high or low, as determined by a flux with an absolute value higher than a given reaction threshold epsilon. Mixed integer linear programming (MILP) was then used as in (Shlomi et al., 2008) to find a network flux state that maximizes agreement with the expression-based reaction activities, where a reaction was confirmed to be on when its flux was greater than or equal to a given reaction threshold, epsilon (ε). This method is distinct from traditional flux balance analysis (FBA) because it is able to predict a flux state for the network independent of a biomass objective function, which may be valuable in cases where cells may not have been evolved towards a maximum growth rate and therefore are not operating optimally.

3. *Evaluation of flux prediction quality*

Flux balance analysis (FBA) was performed as previously described in II.B.4.a) above by maximizing flux through the biomass reaction while using experimentally-determined measurements as input constraints. All data processing and programming was implemented in Python (see Appendix), with the exception of optimization problems, where FBA was implemented using the GNU Linear Programming Kit (GLPK) or Gurobi optimization software (http://www.gurobi.com), and MILP optimization problems used Gurobi. In order to approximate the accuracy of model predictions, simulation prediction vectors were compared to a vector of all available exchange measurements (uptake or secretion of acetate, cellobiose, cellulose, xylose, ribose, fructose, carbon dioxide, ethanol, formate, hydrogen, succinate, and lactate) for batch growth on cellobiose in complex media (Rydzak et al., 2009). This vector also included zeros for sources that were not provided in the growth medium. The Euclidean

المنارة للاستشارات
www.manaraa.com

distance between prediction vectors and the measurement vector was calculated using the NumPy python modules (http://numpy.scipy.org/).

C. Results

In this study, the metabolic state of *Clostridium thermocellum* for growth on cellobiose was studied by algorithmic integration of gene expression data with a genome-scale metabolic model. Gene expression data were obtained by direct RNA sequencing using a Roche 454 system. RNAseq data was integrated with a genome-scale metabolic model using a mixed-integer linear programming (MILP) algorithm to predict network-wide metabolic fluxes. The accuracy of generated flux mappings was evaluated by comparison against experimental phenotypic measurements of growth and metabolite exchanges (consumption and secretion).

1. *Algorithmic integration of RNAseq with metabolic model*

After associating RNAseq reads to a gene locus, the discrete read count for each gene was normalized to gene length in order to derive a relative expression level. Algorithmic integration of gene expression data with a metabolic model seeks to identify the network state that best matches the gene expression data. For this purpose, expression data for each gene is used to predict whether a gene is present (if normalized gene expression is greater than the parameter γ), absent (if gene expression was found to equal zero), or undetermined (if normalized gene expression is greater than zero and less than γ). These gene states were mapped to predicted reaction states using Boolean gene-reaction associations present in the model. The MILP algorithm developed by Shlomi et al. (Shlomi et al., 2008) searches for a flux distribution

المنارة للاستشارات
www.manaraa.com

satisfying network constraints that maximizes agreement with the predicted reaction states. The activity state of a reaction is determined to be on if the absolute value of its flux value is greater than or equal to the parameter ε. Here, "agreement" refers to the number of reactions in the model for which the gene expression and reaction activity states both call a reaction on (flux $>= \varepsilon$) or off (flux $< \varepsilon$). In order to quantitatively evaluate the quality of model predictions against experimental observations, a vector representing the experimentally measured phenotype was generated where each component was a discrete quantitative measurement. Overall the vector contained 13 components, a value for growth rate and 12 exchange reactions (transport in or out of: acetate, cellobiose, cellulose, xylose, ribose, fructose, carbon dioxide, ethanol, formate, hydrogen, succinate, and lactate). Overall similarity between any solution and experimental observations can therefore be quantified by constructing a similar vector from the solution fluxes and calculating the Euclidean distance between the solution and experimental vectors.

Constraint-based modeling is by nature very dependent on the boundary conditions applied to the metabolic network. In general, adding more boundary constraints (i.e., forcing individual metabolite uptake or production rates to match experimental observations) has the tendency to trim the possible solution space and thereby improve the quality of the predictions, but the ability to generate reliable predictions with a minimum of inputs is clearly advantageous. In Figure 12, we explore the quality of predictions given a number of sets of boundary conditions with the goal of determining a minimum set of constraints necessary to achieve reasonable predictive quality. The parameter γ was varied and the distance from the experimental flux

www.manaraa.com

vector was calculated for each state. As expected, additional constraints tended to improve the quality of predictions, although it was not sufficient to force the uptake of only one carbon source (cellobiose) to match the experimental value because the model was still able to utilize other carbon sources in an unbounded fashion. Adding a constraint on growth rate further improves predictions, reducing the distance measure by up to 40%. Carbon source uptake and growth rates are relatively easy to measure experimentally and for growth in minimal media these two measurements combined with algorithmic integration of gene expression data are able to provide predictions of reasonable quality.

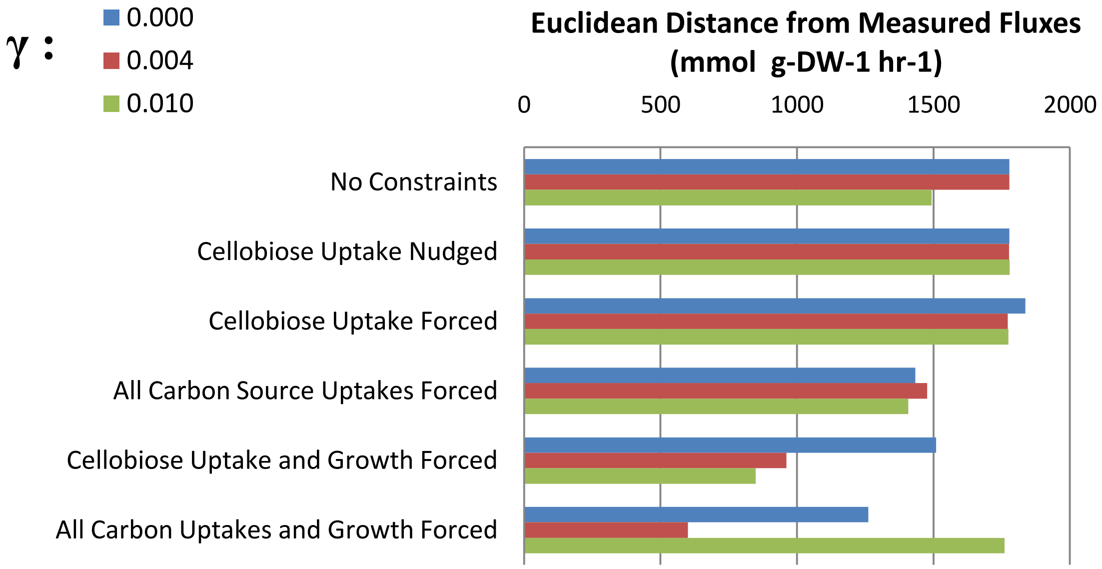
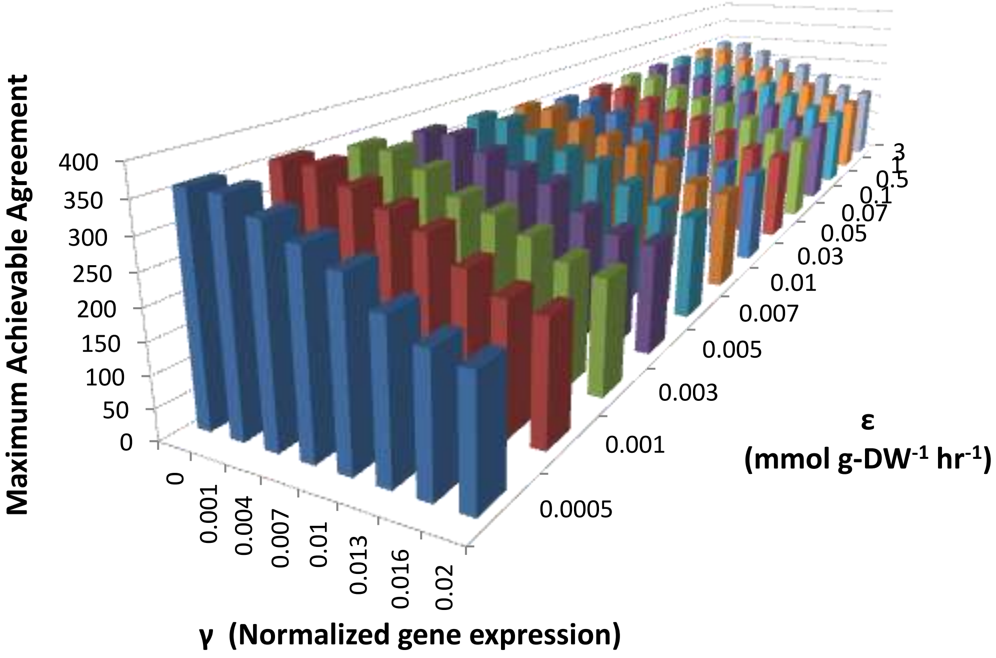


Figure 12: Impact of boundary constraint selection on the accuracy of model predictions.

Due to the high degree of confidence that can be associated with the presence or absence of an mRNA transcript using RNAseq data, the first analysis was conducted where all genes were

www.manaraa.com

called either present or absent (γ = 0). To account for potential errors in the gene expression data, 7 additional scenarios were analyzed with increasing values for the parameter γ, thereby providing more flexibility because an increasing subset of genes remained undetermined. For all of these scenarios, the agreement between the expression data and the computational predictions were scored for each represented reaction in the metabolic model (Figure 13). Figure 13 shows the agreement between experimental and computational predictions for combinations of the different values for the parameters γ and ε. For each parameter, maximum agreement increases as the parameter decreases. For example, one combination is where genes were called absent when normalized expression was zero, present when normalized expression > 0.004, and undetermined when normalized expression was between 0 – 0.004 and the associated reaction was called active if the flux was > 0.0005 and inactive if flux was < 0.0005 resulted in agreement of 335 calls (gene-reaction pairings).



www.manaraa.com

Figure 13: Maximum achievable agreement between gene expression data and the resulting optimal flux distribution. This quantity is identical to the objective value determined by the MILP optimization problem.

2. Comparison to other data integration methodologies

In the MILP algorithm implemented by Shlomi et al., there was no noted qualitative difference in their solutions by varying ε (Shlomi et al., 2008); however, in our case, Figure 13 and Figure 14 demonstrate an effect on both the maximum achievable agreement and the quality of the predictions when compared to experimental observations when varying the parameter ε, particularly as the value is reduced to allow a larger number of reactions to count as "on" even if the nominal flux value is relatively low.

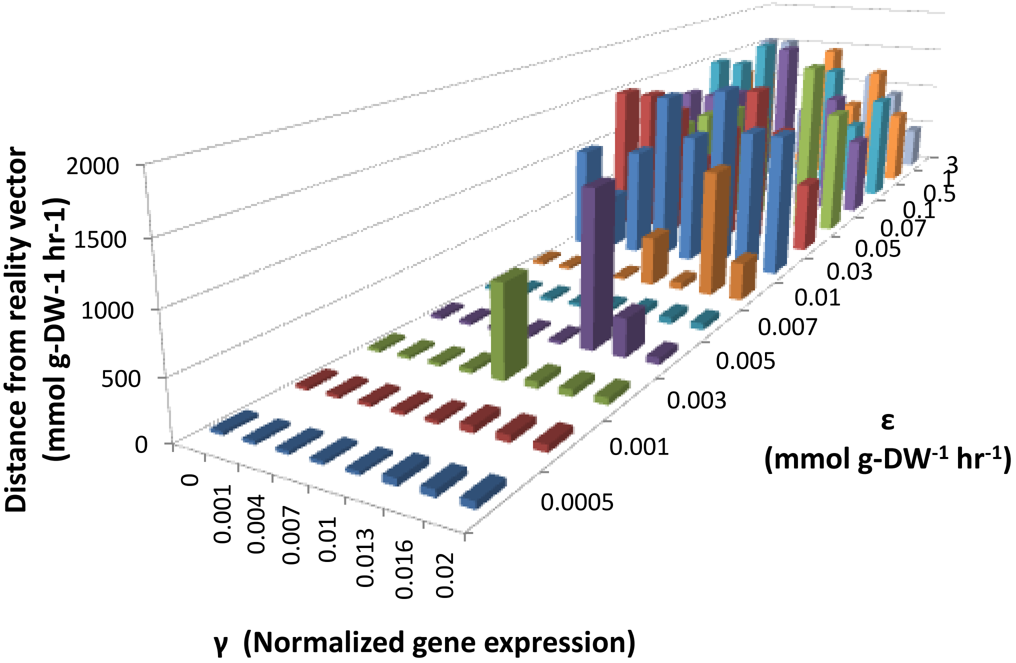


Figure 14: Euclidean distance between the solution vector found by the MILP optimization approach described and a vector of measured flux values during exponential batch growth on complex media, given ranges of the solution parameters γ and ε.

www.manaraa.com

Other methodologies that have incorporated gene expression data with metabolic models (Akesson et al., 2004; Colijn et al., 2009) have utilized gene expression data to specify constraints on reaction fluxes on a gene-by-gene basis. Following the method developed by Akesson et al. (Akesson et al., 2004), any absent gene (normalized expression of zero) will force the flux through the corresponding reaction to also be zero. From our data set, there are a total of 669 genes with a normalized expression of zero. Only 13 of these 669 genes are included in the *i*SR432 model of *C. thermocellum* (the remaining 656 genes are mostly hypothetical genes or genes involved in transcription/translation processes). The 13 absent genes correspond to 8 different biochemical reactions in our model. Of these eight reactions, seven are already inactive in a growth-optimized FBA simulation. The last remaining reaction is for a dihydrofolate reductase (gene *Cth_1226*) that is predicted to be essential in *C. thermocellum* and thus cannot have a zero flux. Thus, incorporation of the RNAseq data using the Akesson method in this case results in no phenotypic difference from an FBA growth-optimized simulation.

3. *Comparison to cellular phenotype*

After generating possible metabolic flux states using different cutoff thresholds for presence/absence (gene expression) and activity (model reaction predictions), we evaluated which flux states corresponded most closely to experimentally measured phenotypes (Rydzak et al., 2009). Figure 14 shows the calculated distance from experimental observation for each of the scenarios discussed above. Even using the same boundary conditions throughout, the algorithmic integration of gene expression data can result in a very wide range of prediction quality, depending at least partially on the selected parameter values. The large variability can

www.manaraa.com

be partially explained by the existence of alternative optimal solutions (R Mahadevan, 2003). As ε decreases, however, the algorithm tends to be much more consistent at selecting flux distributions that most closely mimic observation, and the best distance achieved is 25.8 mmol g-DW-1 hr-1 with parameters γ = 0.004 and ε = 0.01.

4. Resulting flux distribution

To examine the impact of the integration of gene expression data with the model, flux balance analysis was performed while constraining only the carbon source, cellobiose, to simulate batch culture growth on a single substrate. Because the simulation described here is independent of a biomass objective, it can more reliably describe a non-optimal growth phenotype. A comparison of the RNAseq-dependent flux distribution with the optimal growth flux distribution reveals several notable differences in phenotype (see Table 6).

Table 5 shows a comparison of relevant exchange fluxes as determined by each approach, along with the experimentally observed values for each of these variables. FBA alone uses only the most efficient pathway available in order to maximize growth, so it only predicts acetate, carbon dioxide, and hydrogen as fermentation products, leaving out many other important byproducts that are experimentally observed. In contrast, by incorporating gene expression data independent of a biomass objective function, this algorithm is able to predict that these pathways are active and much more closely match experimental observation. Solving this model using FBA yielded 210 reactions with a flux greater than 0.01 in either direction, while algorithmic integration yielded 258 such reactions by turning on 68 and turning off 20.

www.manaraa.com

Because the simulation described here is independent of a biomass objective, it can more reliably describe a non-optimal growth phenotype. A comparison of the RNAseq-dependent flux distribution with the optimal growth flux distribution reveals several notable differences in phenotype (see Table 6).

Table 5: Comparison of predicted and observed measurement vectors. This table compares the flux distributions predicted by flux balance analysis (**FBA**) and by algorithmic integration of gene expression data (**With RNAseq** where $\gamma = 0.004$ and $\varepsilon = 0.01$) against the vector of experimentally observed fluxes (**Reality**) (Rydzak et al., 2009). All units are in mmol g-DW^{-1} hr^{-1}; for biomass flux this converts to hr^{-1}.

	FBA	**With RNAseq**	**Reality**
Biomass	0.54	0.29	0.29
Acetate	43.6	32.1	16.0
Cellobiose	-12.8	-12.8	-12.8
Fructose	0.0	0.0	0.0
CO_2	45.0	17.9	21.4
Ethanol	0.0	0.0	17.6
Formate	0.0	17.0	7.7
H_2	63.5	18.6	18.4
Succinate	0.0	1.0	0.0
Lactate	0.0	0.0	0.0

Table 6 contains the number of reactions in each pathway that are determined to be used by the respective simulation methods, and growth rate optimization with FBA pares down the pathway usage to only the most efficient reactions while integration with RNAseq data reflects a more diverse pathway utilization strategy. This pattern is also clear in the flux diagrams obtained from FBA and from RNAseq integration (not shown). As a specific example of diverse pathway utilization, RNAseq data indicates that the gene for methylglyoxalsynthase

www.manaraa.com

(Cthe_0955) is on, which leads to the production and moderate intracellular accumulation of methylglyoxal (allowed by an artificial cytosolic escape reaction), which is cytotoxic to bacteria and has been suggested as a possible growth regulator (Booth et al., 2003). This non-optimal behavior cannot be captured by FBA alone.

Table 6: Counts and (%) of reactions found to have a flux value greater than 0.01 mmol gDW^{-1} hr^{-1} in either direction when maximizing growth rate using FBA and when maximizing agreement to RNAseq data.

Pathway	FBA only (%)		With RNAseq (%)		Total Reactions
Amino Acid Metabolism	70	(56)	110	(87)	126
Aminosugar Metabolism	4	(31)	12	(92)	13
Cell Envelope Biosynthesis	22	(54)	21	(51)	41
Cellulose Metabolism	1	(5)	3	(14)	21
Citric Acid Cycle	6	(86)	7	(100)	7
Fatty Acid Metabolism	9	(24)	5	(13)	38
Folate Metabolism	4	(25)	15	(94)	16
Glycogen Metabolism	0	(0)	3	(100)	3
Glycolysis/Gluconeogenesis	12	(63)	15	(79)	19
Miscellaneous	5	(36)	9	(64)	14
Nucleobase Metabolism	7	(50)	12	(86)	14
Nucleotide Sugar Metabolism	1	(17)	4	(67)	6
Nucleotides	0	(0)	1	(25)	4
Pentose Phosphate Pathway	6	(67)	9	(100)	9
Porphyrin Metabolism	0	(0)	4	(100)	4
Purine Metabolism	14	(38)	25	(68)	37
Pyrimidine Metabolism	8	(24)	23	(70)	33
Pyruvate Metabolism	6	(38)	13	(81)	16
Sterol Metabolism	0	(0)	7	(47)	15
Sugar Metabolism	0	(0)	2	(33)	6
Transport, Extracellular	13	(18)	19	(26)	72
Urea Cycle/Amino Group Metabolism	7	(44)	15	(94)	16
Vitamins & Cofactors	2	(4)	5	(11)	45
Total	**197**		**339**		**575**

www.manaraa.com

5. *Integration of multiple omics data types*

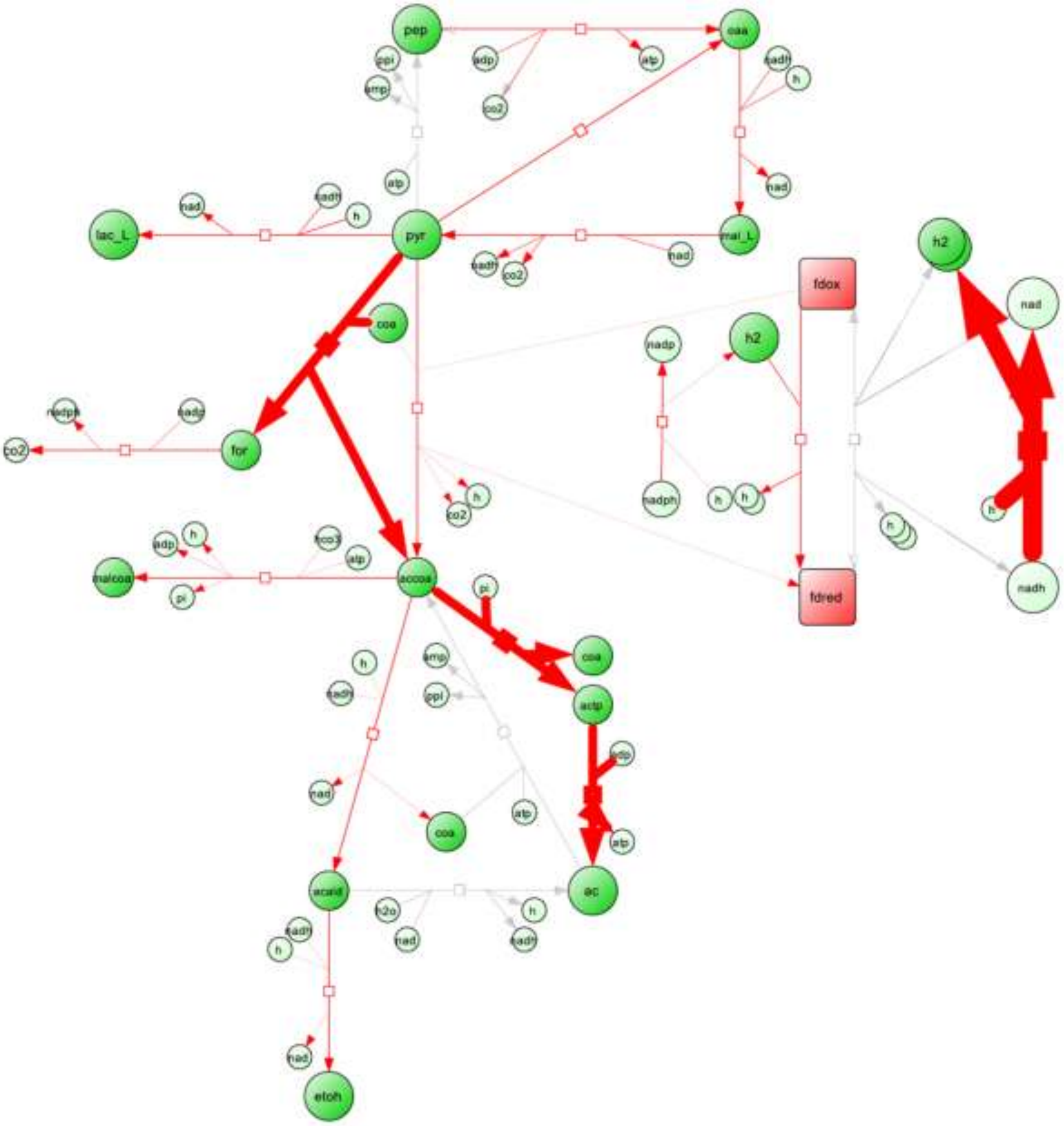
I have shown above how RNAseq data can be usefully incorporated into an *in silico* reconstruction to improve the quality of predictions and shed light on non-optimal phenotypes. Since it is based on direct molecular sequencing and because coverage and sensitivity can be very high for organisms with moderate-size genomes, RNAseq data is especially well suited for providing a confident negative call – i.e., if an mRNA transcript is not detected, there is a high probability that the given gene is not transcribed. The same confidence does not carry over to positive calls, however, as many researchers have noted a consistent lack of reliable quantitative correlation between mRNA and enzyme activity. In an idea experiment, fluxomics data from 13C tracing experiments or other methods would be preferable for positively determining flux measurements, but these experiments are still relatively expensive and difficult. Proteomics datasets could serve as a suitable compromise between these solutions, as they tend to be relatively cost-effective and provide a better positive signal for gene expression. Unlike RNAseq datasets, however, current proteomics technology does not produce reliable negative signals, because the not all peptides are detectable.

One major benefit of the MILP RNAseq integration protocol described above is the flexibility inherent in the types of datasets used to generate a list of gene expression calls. As a case study, the following section describes the integration of a hybrid gene expression dataset with the model. The proteomics dataset is from *Clostridium thermocellum* grown on rich cellobiose medium, and the protein signal rI was normalized to the mass of the peptide Mr(kDa) to give a normalized signal. Gamma values of (0, 0.1) were used for both RNAseq and proteomics

المنارة للاستشارات

www.manaraa.com

datasets. A gene was considered On if the proteomics signal was ≥ 0.1 and Off if the RNAseq signal was ≤ 0. Flux distributions for this analysis using RNAseq data only (Figure 15), proteomics data only (Figure 16), and both RNAseq and proteomics data as described here (Figure 17) are drawn around pyruvate metabolism below. Incorporating both datasets results in a much more disperse flux distribution prediction, which is physiologically more likely.



www.manaraa.com

Figure 15: RNAseq-derived flux distribution surrounding pyruvate metabolism, drawn using CellDesigner (Funahashi et al., 2008).

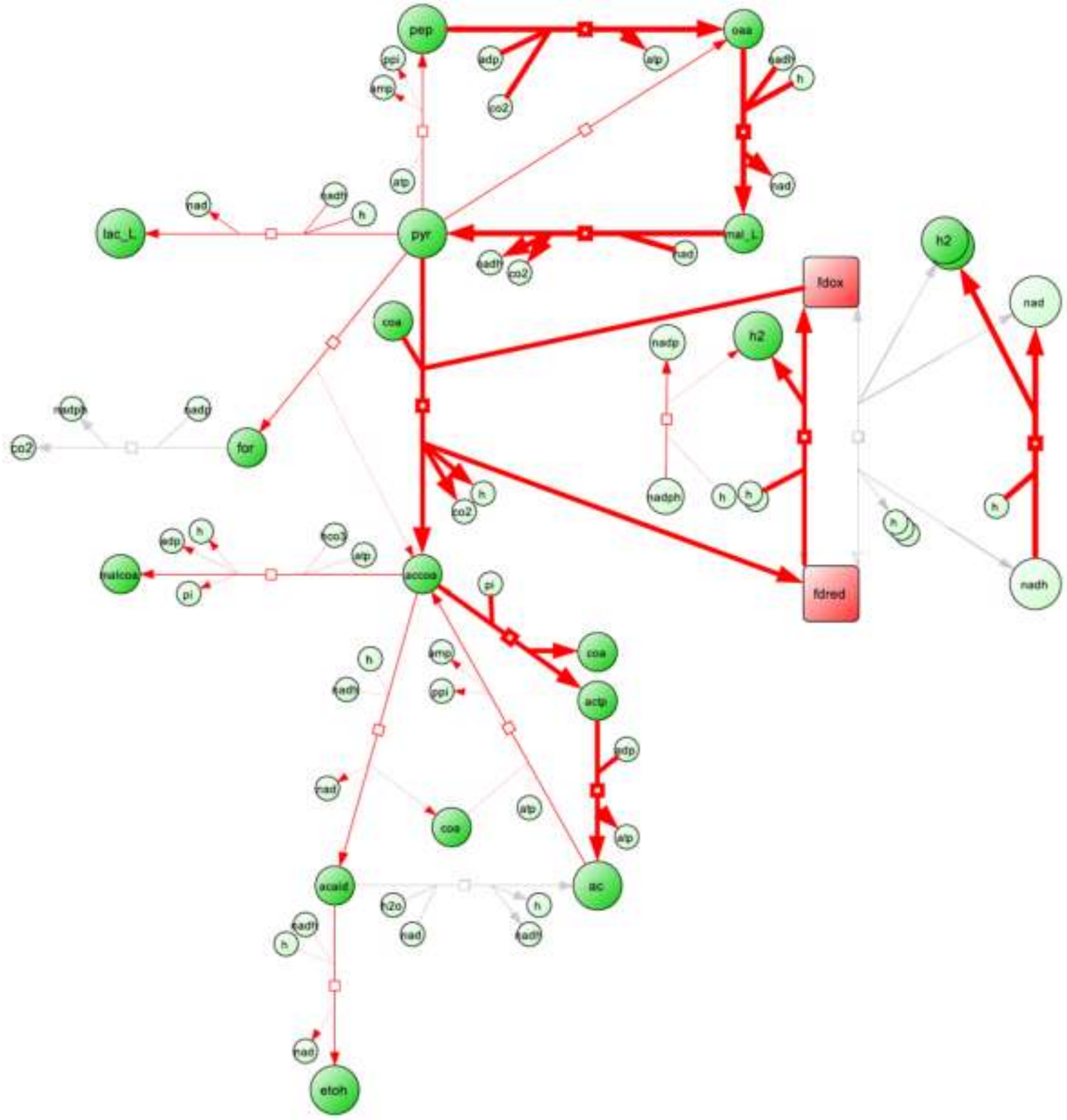


Figure 16: Proteomics-derived flux distribution surrounding pyruvate metabolism, drawn using CellDesigner (Funahashi et al., 2008).

المنارة للاستشارات
www.manaraa.com

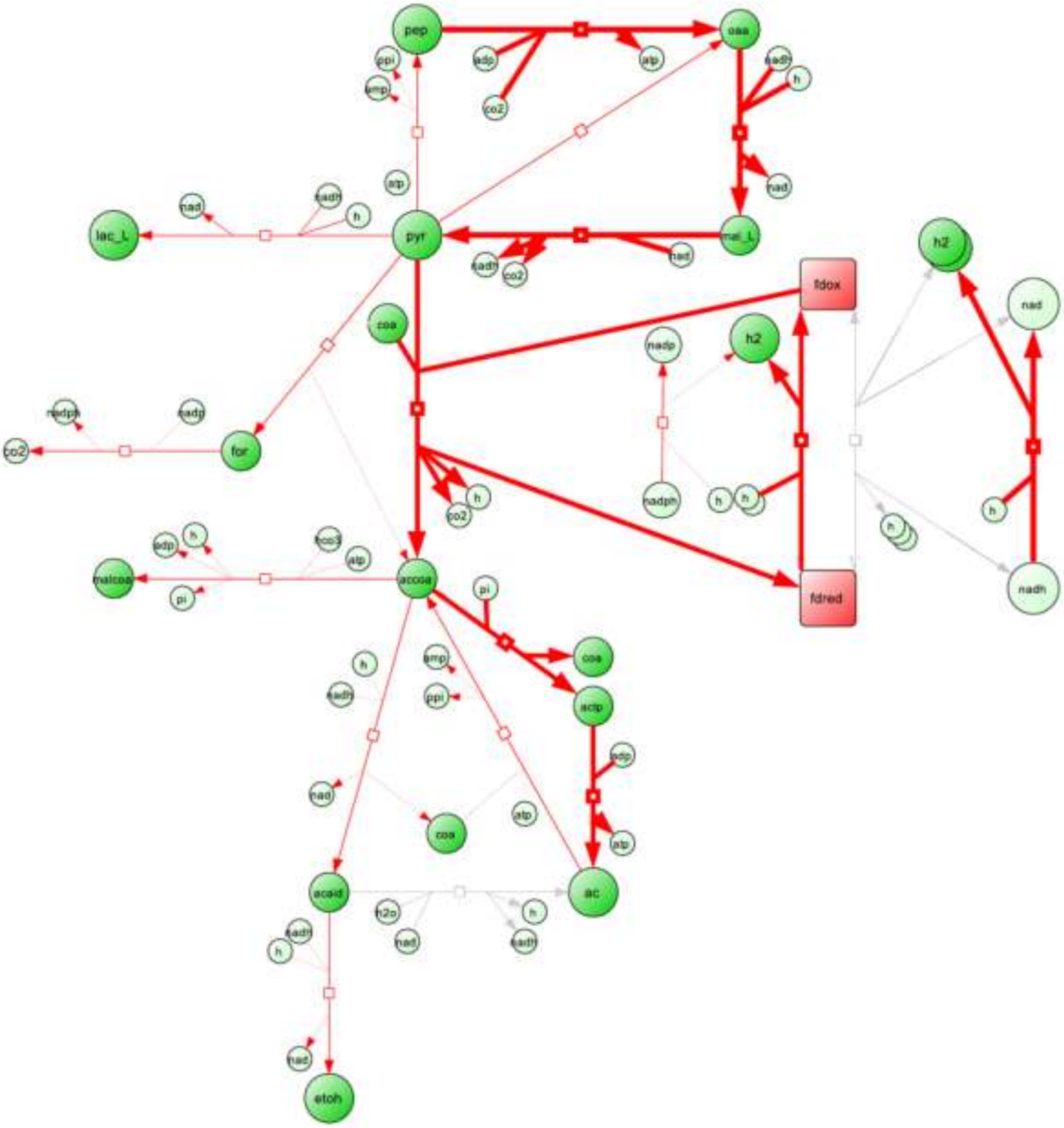


Figure 17: Hybrid RNAseq- and proteomics-derived flux distribution surrounding pyruvate metabolism, drawn using CellDesigner (Funahashi et al., 2008).

D. Discussion

Advancements in systems biology have led to the development of refined high-throughput data for characterization of cellular states and computational methods to analyze and predict biological function. With increasing amounts of data and models becoming available, proper

المنارة للاستشارات

www.manaraa.com

integration of experimental and computational resources may lead to increased diagnostic, analytical, and predictive capabilities that can alleviate some shortcomings of disparate techniques when used in isolation. Experimental data provides information that correlates to a true *in vivo*/*in vitro* state, but it is often difficult to associate data with a broader cellular phenotype. Computational methods can predict functional consequences, but their accuracy and relevance to a true physiological state must be established. Improvements in model accuracy and a clearer picture of the physiological state of *C. thermocellum* demonstrate some of the benefits of algorithmically integrating this data with a genome-scale metabolic model of *C. thermocellum* to produce an experimentally-driven computational flux mapping. Specifically, we found that: 1) RNAseq data is a good experimental data type for algorithmic integration into constraint-based models, 2) quantitative calculation of agreement between model predictions and measured phenotypes show that varying parameters for integrating data can lead to significant changes in outcome, and 3) experimentally-driven computational flux mappings can differ significantly from theoretical, solely computational predictions.

To integrate RNAseq data with a genome-scale metabolic model, we sought to determine the computational state that most closely matched the experimental data where active model reactions matched present mRNA transcripts or inactive model reactions matched cases of absent mRNA transcripts. As we increased values for γ and ε, we allowed for a greater number of genes (in the case of γ) and reactions (in the case of ε) to be left undetermined with no explicit call for presence/absence or activity/inactivity, respectively. Conducting the analysis with varying parameter values was motivated by a desire to not over-constrain or over-specify

www.manaraa.com

the system. However, in conducting the analysis, we found that the best results were obtained from the scenario with the minimum undetermined calls. In this scenario, $\gamma = 0$ meaning all genes were called either present (expression > 0) or absent (expression = 0) and $\varepsilon = 0.0005$ meaning only reactions with a predicted flux value between 0 – 0.0005 were called undetermined. This scenario produced the highest agreement between reaction calls and gene presence/absence and also gave one of the closest functional states to the measured phenotype (Because the simulation described here is independent of a biomass objective, it can more reliably describe a non-optimal growth phenotype. A comparison of the RNAseq-dependent flux distribution with the optimal growth flux distribution reveals several notable differences in phenotype (see Table 6).

Table 5). These results indicate the high degree of confidence that can be associated with RNAseq data for making presence/absence calls. In cases where gene expression data is obtained using hybridization techniques, presence/absence calls can be ambiguous as there is no absolute fluorescence intensity level that demarcates presence from absence. For the RNAseq data, any expression level > 0 (meaning a full mRNA sequencing read was obtained) results in a presence call. Use of RNAseq data for integration with a genome-scale model is ideal from the standpoint of having no ambiguity at the mRNA level, and we have shown that using a clear presence/absence call corresponds to good model integration results.

Effective biological modeling depends on an accurate and complete description of the modeled system. While constraint-based models have been shown to be highly useful for a variety of biological applications (Oberhardt, Palsson, & J. A. Papin, 2009) by producing quantitative

المنارة للاستشارات
www.manaraa.com

phenotype predictions that match well with experimental measurements (Fong & Palsson, 2004), these models are underdetermined so there are numerous ways in which a network can be used to produce an equivalent phenotypic output. Thus, it is often difficult to make detailed predictions at the individual gene or reaction level with a high degree of confidence. Previous efforts to address this problem have sought to elucidate all possible alternate optimal solutions by calculating elementary flux vectors (Urbanczik, 2007) or to calculate the variability of each individual flux for all possible solutions (R Mahadevan, 2003). Here we sought to narrow the range of possible solutions by integrating RNAseq data; however, in doing so, we still generated a variety of possible flux mappings. We sought to maximize the agreement between gene expression and predicted flux activities algorithmically, but this does not guarantee an accurate reflection of the cellular phenotype. After generating our set of possible flux mappings, discrimination of the result that best reflected experimental measurements was obtained by comparison to a vector of cellular phenotype measurements. Only by minimizing the distance (maximizing the agreement) between the predicted flux mapping and the experimental phenotype are we able to identify a flux mapping that should best match the real state of the cell. The final result is a flux mapping that has the highest agreement with gene expression data and with the cellular phenotype. In a sense, the gene expression data could be viewed as training data for our model and the phenotype vector is our objective. By this two-step approach, we are able to more precisely identify a specific flux mapping that correlates to the *in vivo/in vitro* cellular state.

المنارة للاستشارات
www.manaraa.com

The great utility of using constraint-based genome-scale metabolic models is the ability to generate quantitative, detailed predictions for an entire metabolic network based upon genomic and biochemical information. This is a very powerful computational approach to studying biological systems, but it has distinct limitations especially if used in conjunction with flux balance analysis (FBA). Due to the formulation of the computational problem, FBA will calculate optimal solutions to either maximize or minimize a problem. The most common simulation scenario is to maximize growth by stipulating a biomass objective. While we have already discussed the problems associated with having multiple equivalent optimal solutions, there are also potentially problems with computationally deriving optimal solutions in general. Biological systems may not natively function in “optimal” states, meaning that the FBA approach to describing a true cellular state could lead to erroneous predictions. Through the integration of RNAseq data and comparison to a phenotype vector, we were able to generate a flux mapping that closely matches the cellular phenotype of *C. thermocellum* for growth on cellobiose. A normal FBA simulation with the same input constraints predicts optimal biomass production and shows large deviations from the end-product secretion profile that was experimentally observed.

This chapter described the novel application adaptation of a two-step algorithmic approach to integrating RNAseq data with a genome-scale metabolic model to more precisely identify network-wide reaction fluxes in *Clostridium thermocellum* for growth on cellobiose. RNAseq data appears to be ideal for integration of transcriptomic data with constraint-based models and is particularly useful for under-characterized or novel organisms where microarrays are not

www.manaraa.com

available. RNAseq data provided an unambiguous means for determining presence/absence of individual genes to influence activity of computational pathway usage and comparison to a vector of measured cellular phenotypic parameters led to the identification of a flux mapping for the overall system. The end result was a model whose simulation closely matched experimental results and deviated from errors that would have normally been found from a straight FBA optimization simulation. Although this approach was designed for exploring human tissue metabolism, this work represented the first attempt to utilize RNAseq data and to model non-optimal prokaryotic metabolism. Additional contributions were also made regarding the definition and impact of simulation parameters.

www.manaraa.com

V. COMPUTATIONAL STRAIN DESIGN OF *C. THERMOCELLUM* FOR IMPROVED ETHANOL PRODUCTION

A. Introduction

Engineering metabolic pathways for improved production of fuels, chemicals, and recombinant proteins is highly complex and depends on the coordination of regulatory, kinetic, and thermodynamic parameters within the context of the entire network, as well as the careful allocation of energetic and structural resources such as ATP, redox potential, and amino acids. Large-scale *in silico* metabolic reconstructions are powerful tools for managing this complexity and directing engineering decisions. This section will describe some case studies using the *C. thermocellum* model to predict the effects of gene knockouts on metabolic flux distributions and to suggest knockout strain designs to maximize ethanol production. Specifically, the model is used along with Minimization of Metabolic Adjustment (MOMA) as a test case to predict the effects of a pyruvate formate lyase (PFL) knockout on byproduct production, and this approach successfully predicted the qualitative impact on hydrogen, carbon dioxide and formate production. Furthermore, analysis of the model using the OptKnock algorithm to search for gene knockouts that couple improved ethanol secretion with improved growth rate. The model elucidated three general strategies for improved ethanol production, and offered specific gene targets to carry out these strategies. These knockout strains could see as much as 2.6x improvement in ethanol yield compared to current wildtype strains with no decrease in growth rate.

www.manaraa.com

B. Methods

1. Minimization of metabolic adjustment

As described previously, *in silico* models result in highly under-defined solution spaces even when several measured boundary constraints can be applied. The most common approach to probing this solution space, flux balance analysis, can be used to predict the effects of gene deletions at optimal growth by maximizing a biomass equation. Physiologically, however, a knockout strain is in a perturbed state and is unlikely to behave as predicted by FBA. Segre et al. proposed the hypothesis that the metabolic network "responds" to such a perturbation by minimizing the change in the flux vectors of the pre- and post-perturbed cell (Segrè, Vitkup, & Church, 2002). Minimization of metabolic adjustment (MOMA) is a quadratic optimization that minimizes this change, and this approach has been demonstrated as being a better predictor of immediate metabolic effects of genetic perturbation than FBA. The goal of MOMA is to find the new flux vector $\boldsymbol{x} \in \varphi^j$ such that the Euclidean distance

$$D(w, x) = \sqrt{\sum_{i=1}^{N} (w_i - x_i)^2}$$

from the initial flux vector is minimized. Since minimizing the function D is equivalent to minimizing its square and we can omit constant functions, we can reduce the optimization problem to the minimization of the function

$$f(x) = L \cdot x + \frac{1}{2} x^T Q x$$

www.manaraa.com

where the vector $\boldsymbol{L} = -w$ of length N and the matrix **Q** with dimensions N × N define the linear and quadratic parts of the objective function, respectively, and $\mathbf{x}^T$ represents the transpose of **x**. This optimization can be performed generally using any initial flux vector $\boldsymbol{w}$, but MOMA is specifically interested in an initial flux vector associated with optimal growth as determined by FBA.

All modeling simulations in this section were performed in MATLAB using the COBRA Toolbox (Becker et al., 2007). The latest revision of the *C. thermocellum* model was used, and all simulations used the following boundary constraints, unless otherwise noted:

cellobiose[e] uptake ≤ 12.8

cellulose[e] uptake = 0.0

fructose[e] uptake = 0.0

2. *Simulation of PFL-knockout mutant*

Hypophosphite (HPP) has been shown to inhibit the activation of the enzyme pyruvate-formate lyase (PFL) (Brush, Lipsett, & Kozarich, 1988). To simulate a deletion of this enzyme and its perturbation to the metabolic network, *Clostridium thermocellum* was grown to stationary phase in rich media with 0, 3mM, 6mM, and 9mM HPP added. Growth was measured using OD_{600} and cells were observed to grow at the same rate and to the same final cell density in each case. Measurements of hydrogen, carbon dioxide, lactate, acetate, formate, and ethanol production rates were done using HPLC and GC-MS.

www.manaraa.com

3. *Growth-coupled ethanol secretion*

A fundamental challenge to metabolic engineering is addressing the genetic instability of manipulations that reduce growth rate. Cells have been shown to adapt towards an optimal growth rate due to evolutionary pressure during exponential growth (Fong & Palsson, 2004). If a strain design is counter-productive to improved growth and the genetic changes can be easily reversed, many cells will quickly lose the engineered functionality and will outcompete well-behaved individuals. Gene knockouts are especially stable against this problem, and there are a number of strain design algorithms that usefully employ evolutionary adaptation by coupling production of a desired phenotype to optimal growth. The algorithm OptKnock (Burgard, Pharkya, & Maranas, 2003) is one such algorithm that solves a bilevel optimization problem of the form

$$\max_{y_i} v_T$$

s.t.:

$$\max_{v_i} v_{biomass}$$

s.t.:

$$S \cdot v = 0$$

$$a_i \leq v_i \leq b_i \text{ for all reactions i,}$$

$$v_{biomass} \geq v_{target\ biomass}$$

$$y_i = \{0, 1\},$$

$$\sum_{i \in M} (1 - y_i) \leq K$$

where the inner optimization is the same as FBA maximizing biomass, v_T is the target flux, and ($y_i = 1$) is the set of reactions that can be accessed by the model. This results in an optimization

www.manaraa.com

of the target biomass that is concurrent with FBA results for a given set of deletions. This problem can be reduced to a single-level MILP problem that can reach a feasible solution in a reasonable time frame (< 24 hr) for as many as four simultaneous deletions in combination. An implementation of OptKnock is available in the COBRA toolbox for MATLAB. OptKnock was performed on the most recent revision of the *C. thermocellum* model to maximize ethanol production during growth on cellobiose, using up to four deletions.

C. Results

1. Simulation of pfl knockout mutant strain

Persistent difficulties with performing genetic manipulations on *C. thermocellum ATCC 27405* eliminated the possibility of using traditional knockout studies to evaluate and improve the model. In some cases, however, these perturbations to the metabolic network could be mimicked through thermodynamic or chemical means. One such method is treatment with hypophosphite, which has been shown to inactivate the enzyme pyruvate formate lyase (PFL). Fermentation experiments were performed in the presence of 3mM, 6mM, and 9mM hypophosphite to examine the metabolic effects of a PFL knockout experimentally. MOMA analysis of this deletion is seen in Figure 19 and Figure 20 below, and the qualitative results of this analysis correlate well with observations, as shown in Figure 18.

www.manaraa.com

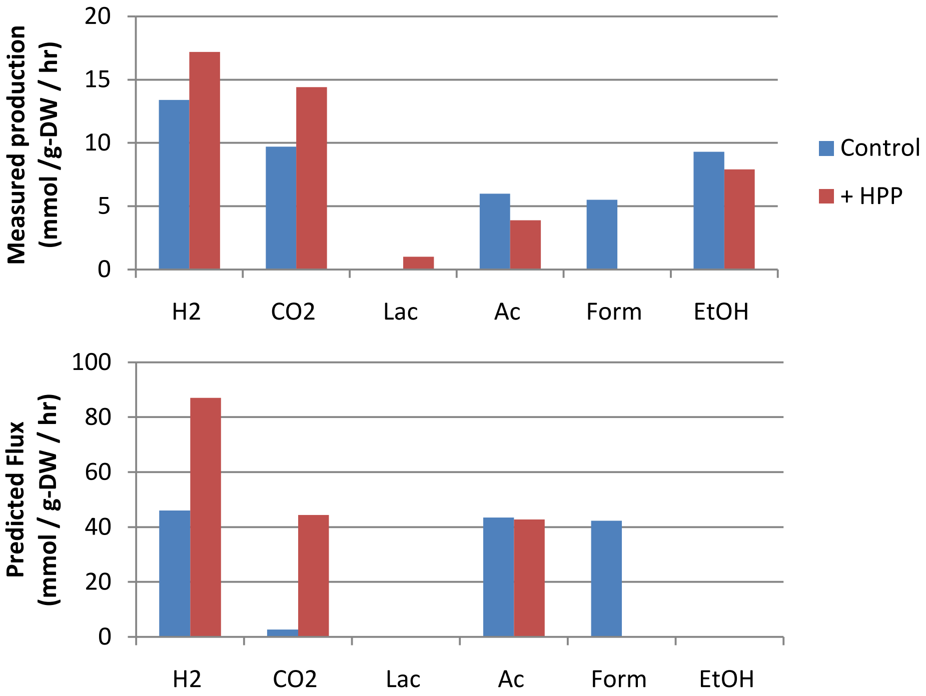


Figure 18: Comparison of MOMA predictions to experimentally measured metabolic response to PFL inactivation.

www.manaraa.com

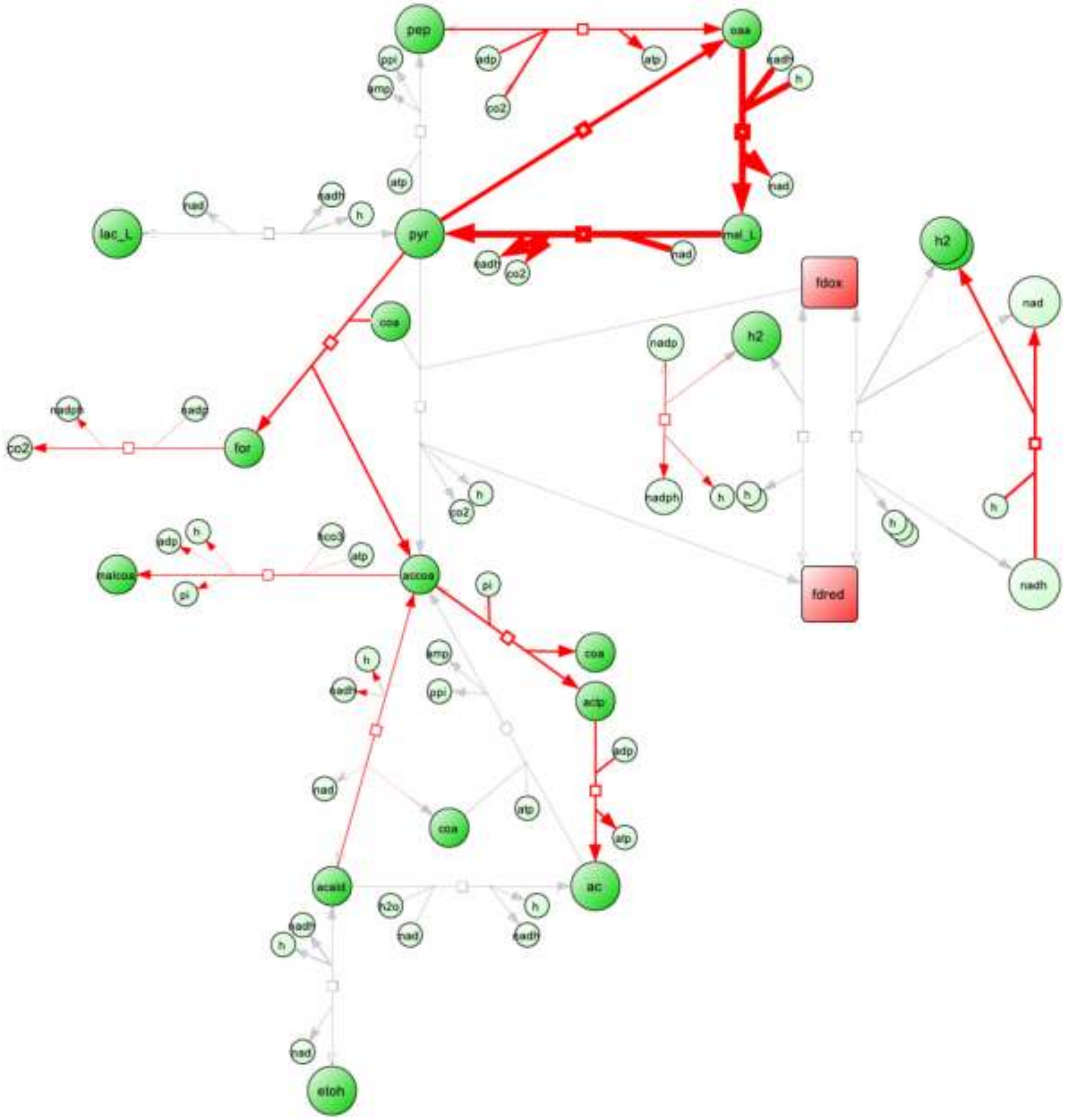


Figure 19: FBA flux distribution prior to pfl knockout, drawn using CellDesigner (Funahashi et al., 2008).

المنارة للاستشارات
www.manaraa.com

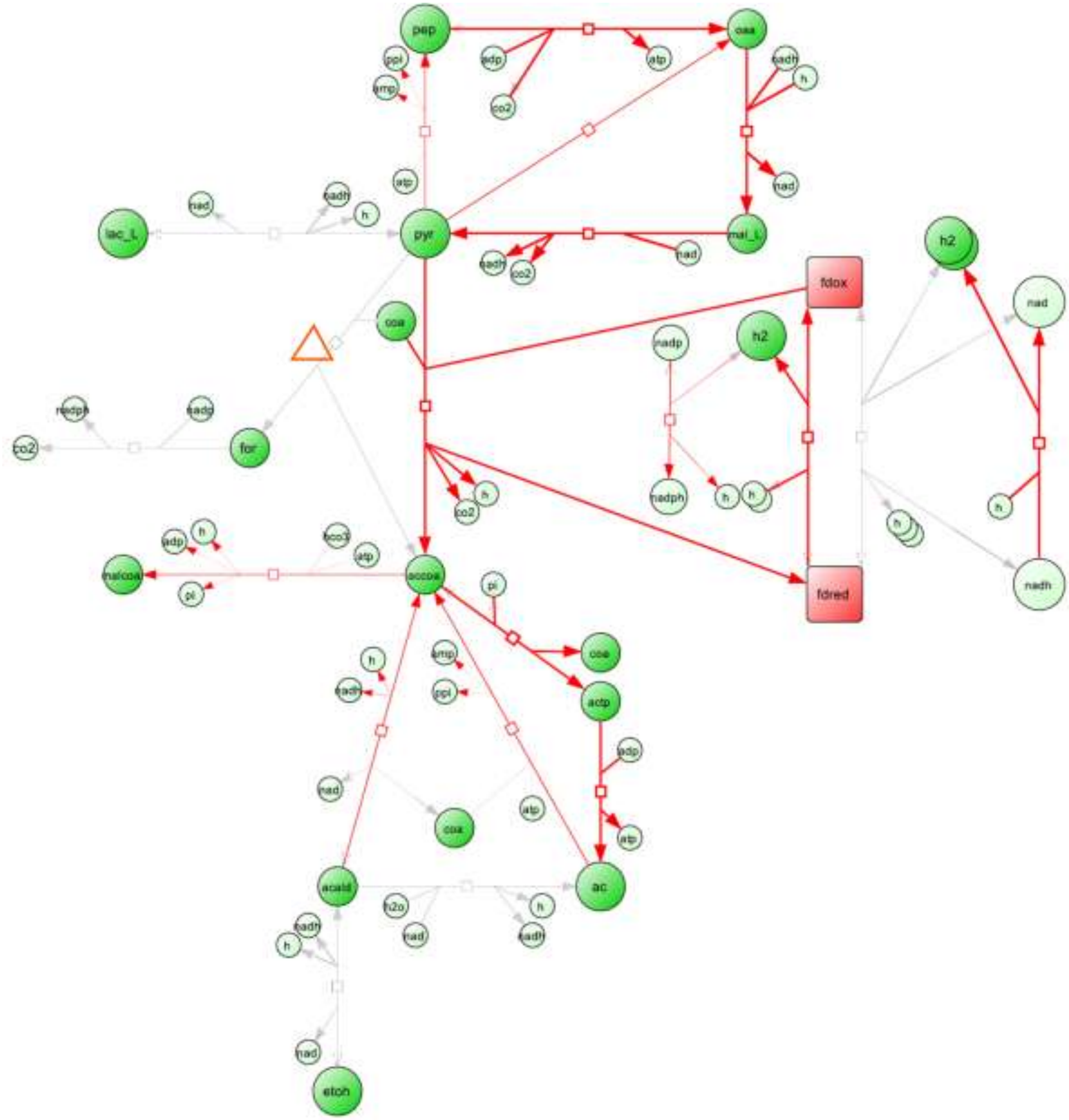


Figure 20: Flux distribution predicted by MOMA for a pfk knockout strain, drawn using CellDesigner (Funahashi et al., 2008). The deletion of PFL is marked by the orange delta symbol.

2. Putative strain designs for improved ethanol production

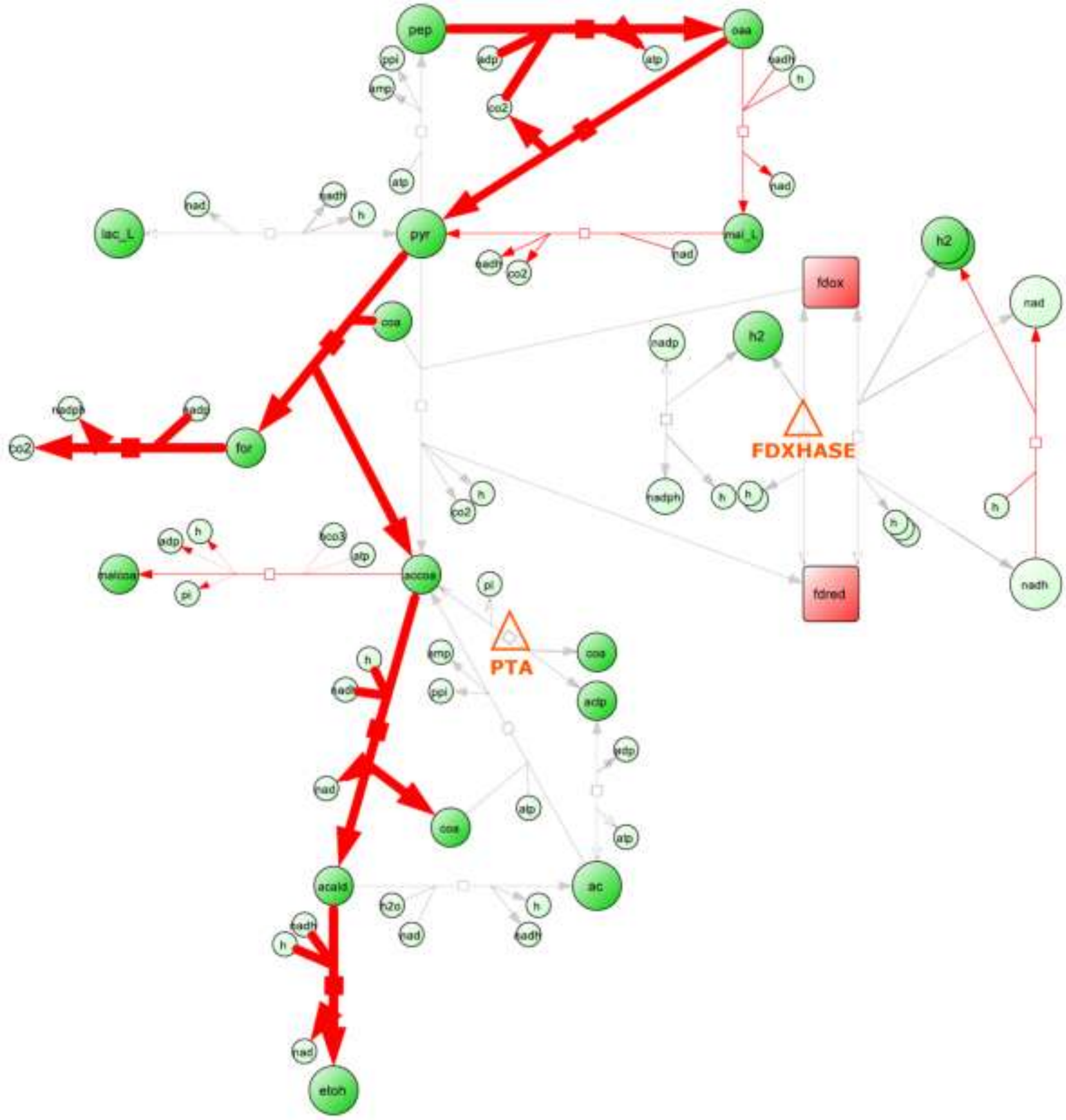
The OptKnock bilevel optimization algorithm was used to derive a set of potential strain designs that maximize ethanol production coupled with optimal growth rate. The resulting designs can serve as a valuable starting point for metabolic engineering. A single-reaction knockout of the

المنارة للاستشارات
www.manaraa.com

hydrogen:NAD+ oxidoreductase(NADOX) results in an ethanol secretion of 25 mmol / g-DW / hr at an optimal growth rate of 0.412 hr^{-1} (see Figure 21). This compares to zero ethanol production at a wild-type optimum, or 17 mmol EtOH / g-DW / hr observed at a non-optimal growth rate of 0.29 hr^{-1}. The reaction NADOX is coded by one of two different protein complexes, (Cthe_0341 and Cthe_0342) or (Cthe_0429 and Cthe_0430), so practical implementation of this reaction knockout may require a minimum of two gene knockouts.

www.manaraa.com

Deleting two reactions concurrently (Figure 22) generates further improvement in optimal ethanol production by blocking Ferredoxin:hydrogenase (FDXHASE) and phosphotransacetylase (PTA), which leads to an ethanol production of 46.9 mmol / g-DW / hr at a growth rate of 0.29 hr^{-1}. The FDXHASE reaction is currently assigned to a six-gene operon, but there are a number of eligible hydrogenases in *C. thermocellum* that may be ferredoxin dependent. Many of these likely have



المنارة للاستشارات
www.manaraa.com

enzyme binding kinetics that favor H_2 uptake, and it remains uncertain which one is primarily responsible for H_2 production. Despite those limitations, both of these reactions could conceivably be blocked by only two gene knockout events, since PTA is coded by a single gene. In both strain designs so far mentioned, some blockage of electron flow to H_2 is used (NADOX and FDXHASE). Electron flux to H_2 as a terminal electron acceptor is energetically favored over flow to ethanol so H_2 production inherently competes with ethanol production, as was explored in Figure 4. Additionally, the double-deletion strain begins to address competing carbon flows by blocking export of acetate through the deletion of PTA. Theoretically, deletion of the subsequent reaction acetate kinase (ACK) serves the same purpose, but practically, PTA is generally preferred because it prevents possible accumulation of the intermediate acetyl phosphate upstream of the deletion.

المنارة للاستشارات
www.manaraa.com

Both the triple- and quadruple-deletion scenarios (

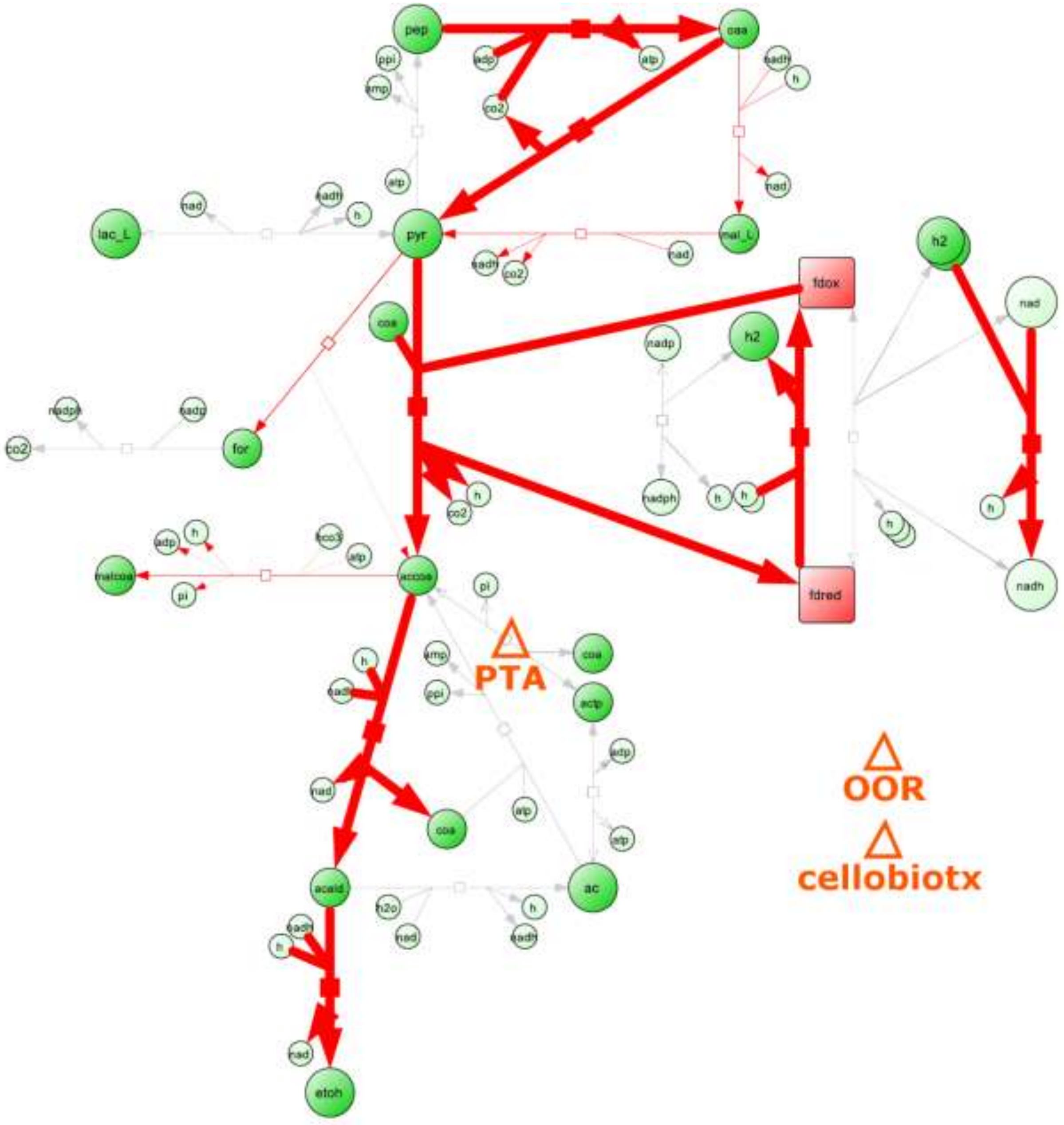


Figure 23 and Figure 24, respectively) begin to target ethanol secretion by changing how ATP is generated in the cell by deleting the cellobiose transporter (cellobtx) and 2-oxoglutarate synthase (OOR) in addition to the previous PTA knockout.

www.manaraa.com

The cellobiose transporter is removed because it requires hydrolysis of one mole of ATP per mole of cellobiose,

cellobtx: cellobiose[e] + h2o[c] + atp[c] -> adp[c] + cellobiose[c] + pi[c] + h[c]

whereas the external beta-glucosidase can hydrolyze cellobiose into two moles of glucose, each of which is transported at the cost of one mole of ATP. The result is a net doubling of ATP cost to import glucose equivalents into the cell:

bglu2_ext: cellobiose[e] + h2o[e] -> 2 glc-d[e]
GLCt5: glc-d[e] + h2o[c] + atp[c] -> adp[c] + glc-d[c] + pi[c] + h[c]

المنارة للاستشارات
www.manaraa.com

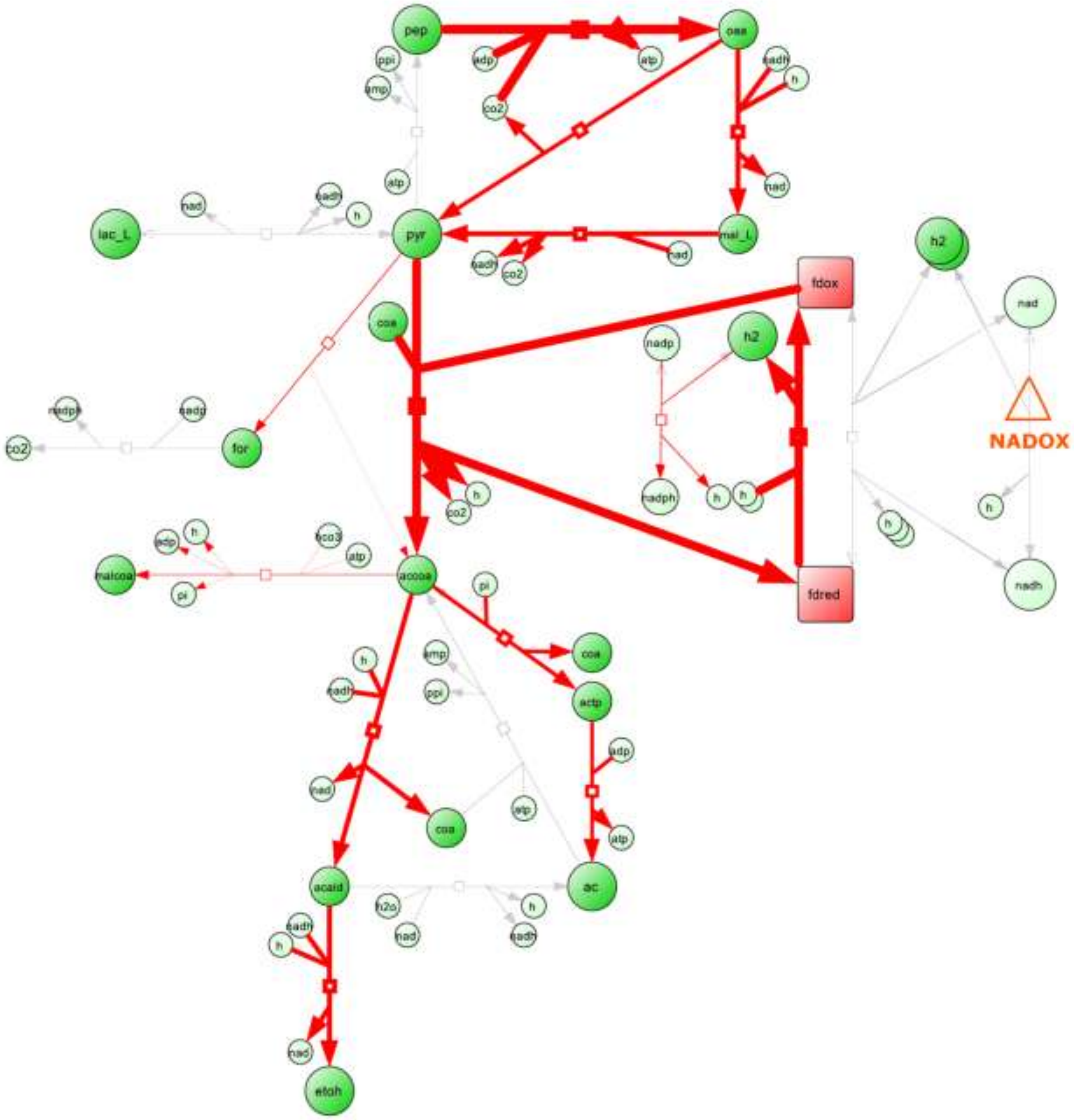


Figure 21: Single-reaction OptKnock proposes deletion of NADOX for improved ethanol production, drawn using CellDesigner (Funahashi et al., 2008).

المنارة للاستشارات

www.manaraa.com

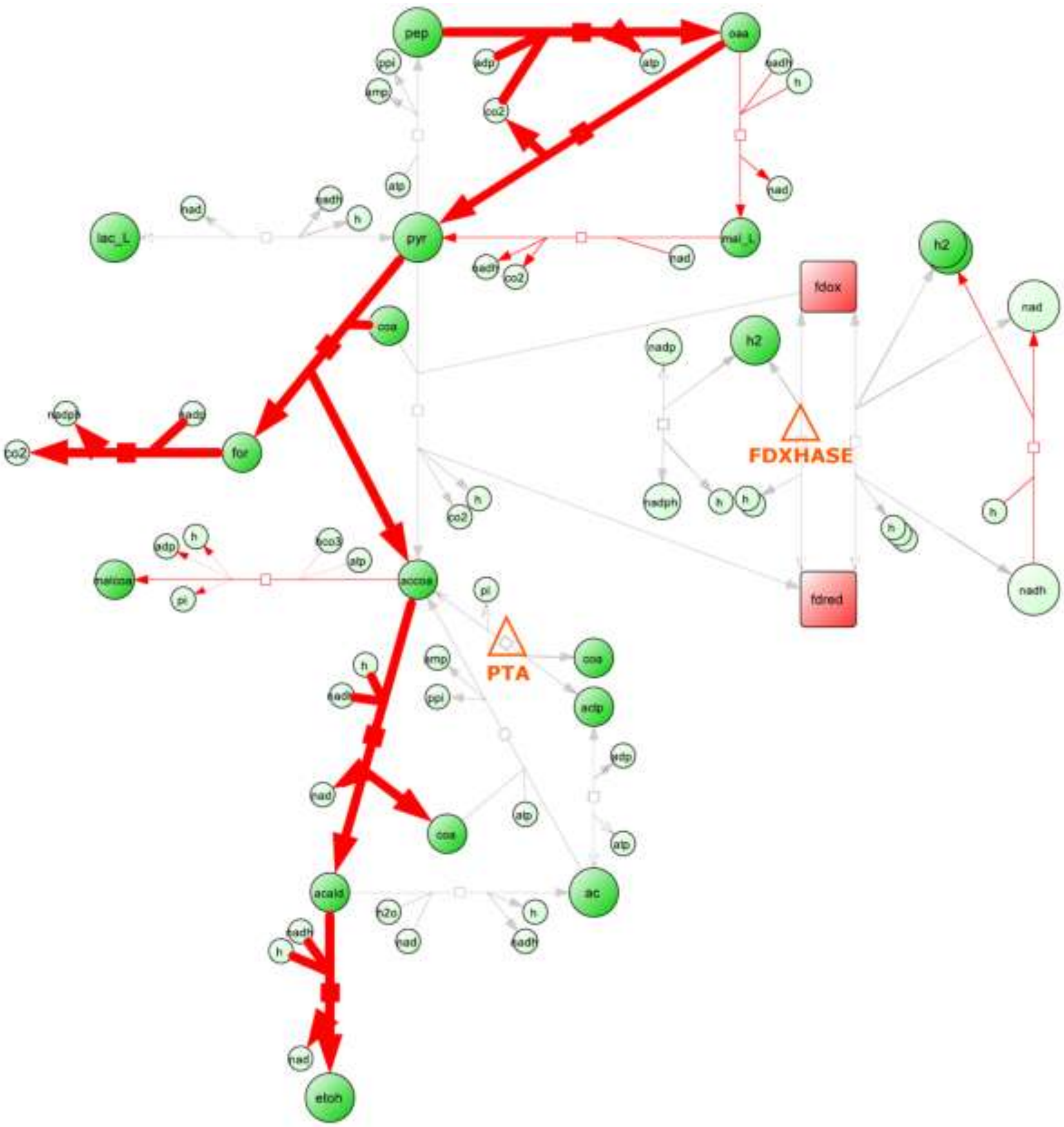


Figure 22: Double-reaction OptKnock proposes deletion of FDXHASE and PTA for improved et

المنارة للاستشارات

www.manaraa.com

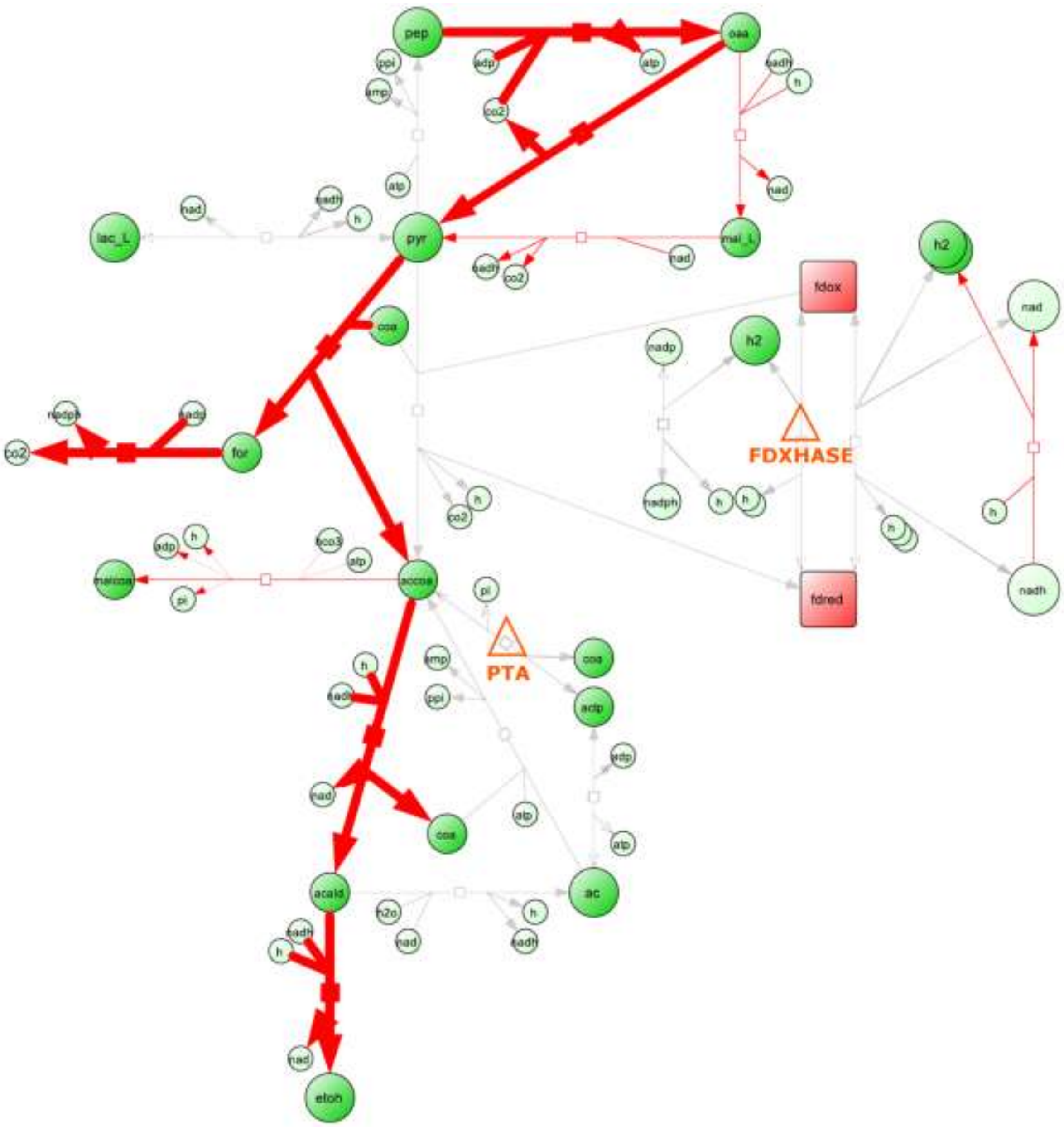


Figure 22hanol production, drawn using CellDesigner (Funahashi et al., 2008).

المنارة للاستشارات

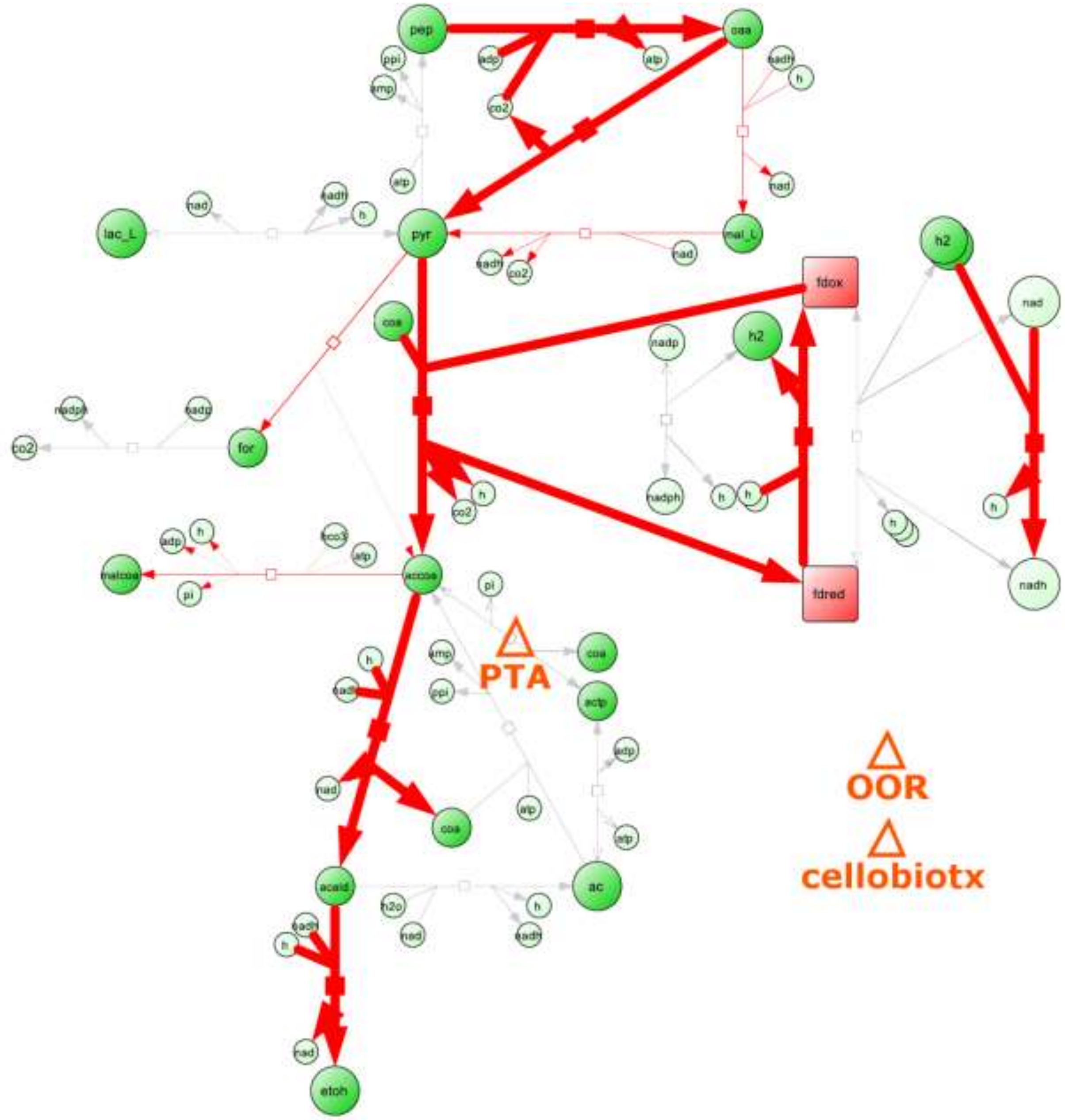


Figure 23: Triple-reaction OptKnock proposes deletion of PTA, OOR, and the cellobiose transporter for improved ethanol production, drawn using CellDesigner (Funahashi et al., 2008).

This deletion results in a lower growth rate, but requires more carbon to be fermented in order to meet the energy needs of the cell. Since this carbon flow is still forced away from acetate via the PTA deletion, the result is a theoretical ethanol production of just over 49 mmol / g-DW / hr at a growth rate of 0.15 hr^{-1}. The third deletion of OOR affects electron flow in the cell, which

www.manaraa.com

allows the reactions FDXHASE and NADOX (this time in reverse) to generate nadh that can be used to reduce acetyl-CoA to ethanol:

FDXHASE: 2 h[c] + 2 fdred[c] <=> 2 fdox[c] + h2[c]

NADOX: nad[c] + h2[c] <=> nadh[c] + h[c]

While the practical implementation of these reaction deletions may be challenging due to incomplete experimental evidence for the gene annotations implied, the strategies used here are non-obvious and could be adopted in other specific reaction deletions to achieve similar results. For instance, it may be a bad strategy to delete the cellobiose uptake for fear of counteracting the natural cellulolytic prowess of *C. thermocellum*, and this disadvantage is not captured in the stoichiometric representation of the reactions used. It could be possible to look for other ways to decrease the ATP yield per carbon uptake.

Four-reaction deletion analysis using OptKnock resulted in an incremental increase in ethanol yield over the triple-deletion, increasing to 49.2 mmol / g-DW / hr, but at a further cost to growth rate, leading to a maximum growth rate of only 0.13 hr^{-1}. This incremental improvement is likely not worth the decrease in cell health, as a lower growth rate would lead to a lower productivity, though it would improve ethanol:carbon yields.

D. Discussion

Successful metabolic engineering demands a careful combination of high-throughput informatics-based approaches with a thoroughly detailed mechanistic understanding of the

www.manaraa.com

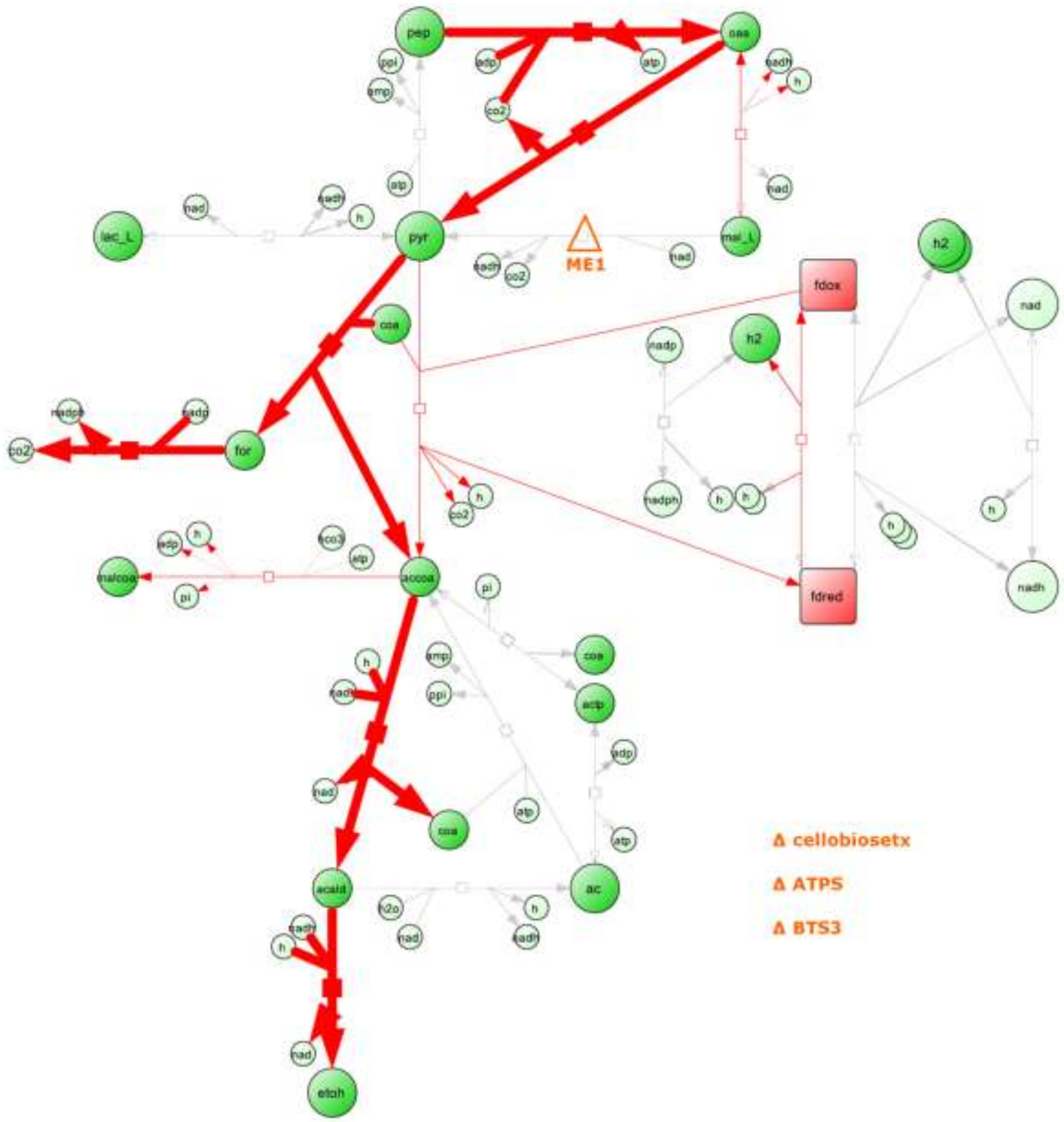


Figure 24: Quadruple-reaction OptKnock proposes deletion of ME1, ATPS, BTS3, and the cellobiose transporter for improved ethanol production, drawn using CellDesigner (Funahashi et al., 2008). Deletion of malic enzyme (ME1), the cellobiose transporter (cellobtx), ATP synthase (ATPS) and biotin synthase (BTS3) are proposed for further incremental improvement of ethanol yield.

biochemical underpinnings of metabolism and genetic control. In this section, I demonstrated the utility of *in silico* constraint-based models for generating optimal strain designs and testing the phenotype responses to genetic manipulation. It also demonstrates the care needed to

المنارة للاستشارات

www.manaraa.com

interpret those predictions in a way that is biologically relevant and practically applicable given the immense complexity of a biological system. With this in mind, a suitable gene knockout strategy for improved ethanol production in *C. thermocellum* might begin with knockouts of PTA and other genes leading to competing carbon end-points like formate and lactate. We have also shown that this strategy could be greatly improved upon by inhibiting electron flow to hydrogen, either by directly deleting the offending hydrogenase enzyme(s) or by disrupting electron transport pathways that ultimately lead to hydrogen. Finally, we have demonstrated the advances that can be achieved by crippling ATP:Carbon yields to force more carbon uptake. This approach must be carefully balanced with cell health to ensure strain stability and high productivity.

In addition to the OptKnock strain design algorithm described here, a number of algorithmic improvements to this deletion analysis have since been developed. OptStrain adds a database of metabolic conversions that could potentially be added to a cell as recombinant enzymes to further enhance productivity or to engineer production of non-native chemicals (Burgard, Pharkya, & Maranas, 2004). More recently, a third algorithm called EMILiO uses successive linear programming to search complex gene manipulations beyond a simple deletion or addition (L. Yang et al., 2011). In other words, EMILiO can search possible up- and down-regulation modifications for optimal strain design, thus fine-tuning strain design predictions and reducing the dependence on adaptive evolution of engineered strains to optimal phenotypes.

المنارة للاستشارات
www.manaraa.com

VI. CONCLUSIONS

Engineering metabolic pathways for improved production of fuels, chemicals, and recombinant proteins is highly complex and involves coordination of regulatory, kinetic, and thermodynamic parameters, as well as the careful allocation of energetic and structural resources such as ATP, redox potential, and amino acids. Large-scale *in silico* metabolic reconstructions are powerful tools for managing this complexity and directing engineering decisions. My dissertation research has been focused on methods for generating, managing, and interpreting all of this data in ways that are meaningful to researchers while maintaining physiological relevance. *In silico* metabolic reconstructions provide the broadest available platform upon which to layer this data where it can be interpreted within a global context.

As a case study for the types of data integrations possible and how they can be used for metabolic engineering, my dissertation research has focused on the construction and analysis of an *in silico* model of the anaerobic cellulolytic bacterium *Clostridium thermocellum*. In this dissertation I have described the model construction process as well as the ongoing refinement to incorporate new physiological findings and to adjust incorrect predictions. The resulting metabolic reconstruction is a functionally complete metabolic model containing 577 enzymatic reactions and 432 associated genes. This model provides a phenotypic context for predicting the annotation of 33 previously unknown or partially-understood genes, successfully mimics growth behavior on cellobiose and fructose in batch and continuous culture, predicts gene knockouts that couple ethanol production with growth rate, provides quantitative explanation

www.manaraa.com

of the bioenergetic tradeoff between hydrogen and ethanol production in *C. thermocellum*, and explores potential alternative media formulations that would improve the overall ethanol flux.

Additionally, I have described the planning, execution, and analysis of an RNAseq experiment for basic growth on a cellobiose minimal medium, explaining how such a dataset can be used to understand quantitative gene expression levels, to detect coexpression of neighboring genes as a cohesive operon, and to predict possible transcription start site locations. I described novel bioinformatics algorithms that utilize RNAseq datasets to quantify gene expression, detect operon structures and predict transcription start sites. Analysis of the RNAseq data from *C. thermocellum* grown on cellobiose minimal medium resulted in the detection of 686 operons in which neighboring genes were coexpressed, and predicted 22 distinct transcription start sites. Furthermore, these algorithms are designed to facilitate integration of multiple analyses in parallel to improve expert interpretation. This data integration is exemplified in Figure 8 where the unified display of disparate analytical techniques was used to more accurately define the structure and transcription start site of the hydrogenase operon beginning at Cthe_0338.

Although the above-mentioned RNAseq analysis techniques were effective at integrating multiple measurements in parallel, they still have the limitation that they examine each gene in isolation from the rest of the genetic network. Understanding the physiological significance of quantitative changes in gene expression levels based solely on this type of analysis is somewhat limited due to the great disconnect between an enzyme's transcript levels and its corresponding metabolic flux. Managing this widely observed phenomenon remains one of the main challenges in engineering metabolism. In Chapter IV, I describe important strides towards

المنارة للاستشارات

www.manaraa.com

addressing this disconnect by integrating an RNAseq dataset with a genome-scale metabolic model in a way that is independent of a biomass objective function. The described method allows a global interpretation of quantitative expression levels within the context of the entire metabolic network and represents a novel application of a previously-described optimization algorithm towards modeling non-optimal prokaryotic metabolism and demonstrates significant improvements in prediction accuracy over both FBA alone and FBA with a Boolean integration of RNAseq data. In *C. thermocellum*, this method was applied to the genome-scale model using data from the RNAseq experiment described in Chapter III as well as proteomics data obtained during growth of *C. thermocellum* on a rich cellobiose medium. I demonstrate how the algorithmic combination of both data types maximizes the respective strengths of the different datasets and improves physiological interpretation.

Finally, Chapter V explored the use of the metabolic model as a tool for metabolic engineering of *C. thermocellum* for improved ethanol production. As a test case study, the method of minimization of metabolic adjustment (MOMA) was used to predict the phenotypic impact of a pyruvate formate lyase deletion. The model predictions qualitatively matched the observed response of a strain of *C. thermocellum* treated with hypophosphite during growth on rich medium and furthermore predicted that such a perturbation would be fatal during growth in a minimal medium without nucleosides available to be scavenged for supplemental formate production. OptKnock simulations were then performed to search for genetic perturbations that would couple improved ethanol production with optimal growth, and these experiments discovered three main strategies for enhancing ethanol production while proposing a number

www.manaraa.com

of candidate genes that carry out these strategies. These knockout strains could result in as much as a 2.6x improvement in ethanol yield compared to the current wildtype strains with no decrease in growth rate. Future efforts to develop genetic tools for *C. thermocellum* will enable these design predictions to be tested, completing the engineering design cycle by validating correct predictions and elucidating problem areas.

The analyses described in this dissertation demonstrate the value of genome-scale metabolic models for integrating disparate bioinformatics datasets and predicting physiological impacts of genetic perturbations, but these tools are only one piece of the growing metabolic engineering toolkit. Beyond merely implementing the strain designs described here, quick advances in synthetic biology are enabling the introduction of non-native enzymatic functions and genetic controls into numerous host organisms. Synthetic pathways have been introduced into *E. coli* for the production of butanol (Atsumi et al., 2008) and into *Corynebacterium glutamicum* and *Clostridium cellulolyticum* for the production of isobutanol (Higashide, Y. Li, Y. Yang, & Liao, 2011; Kevin Michael Smith, K.-M. Cho, & Liao, 2010). These studies depended on intuitive manipulations to metabolism and achieved impressive success, but further improvements in product yields are achievable by manipulations predicted by constraint-based models (Shinfuku et al., 2009). The benefits of genome-scale metabolic analysis will only increase as synthetic pathways become more complex, making intuitive design more difficult. Contributions described in this dissertation will help to make constraint-based metabolic models a core component of the modern synthetic biologist’s toolkit and will greatly assist in the global interpretation of experimental datasets for successful design of synthetic metabolic networks.

المنارة للاستشارات
www.manaraa.com

VII. BIBLIOGRAPHY

Akesson, M., Förster, J., & Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Metabolic engineering*, *6*(4), 285-93. doi:10.1016/j.ymben.2003.12.002

Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, Kevin M, Brynildsen, M. P., Chou, K. J. Y., et al. (2008). Metabolic engineering of Escherichia coli for 1-butanol production. *Metabolic engineering*, *10*(6), 305-11. doi:10.1016/j.ymben.2007.08.003

Beard, D. A., Babson, E., Curtis, E., & Qian, H. (2004). Thermodynamic constraints for biochemical networks. *Journal of theoretical biology*, *228*(3), 327-33. doi:10.1016/j.jtbi.2004.01.008

Becker, S. a, Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., & Herrgård, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, *2*(3), 727-38. doi:10.1038/nprot.2007.99

Bonarius, H. P. J., Schmid, G., & Tramper, J. (1997). Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends in Biotechnology*, *15*(8), 308-314. doi:10.1016/S0167-7799(97)01067-6

Booth, I. R., Ferguson, G. P., Miller, S., Li, C., Gunasekera, B., & Kinghorn, S. (2003). Bacterial production of methylglyoxal: a survival strategy or death by misadventure? *Biochemical Society transactions*, *31*(Pt 6), 1406-8. doi:10.1042/

Brener, D., & Johnson, B. F. (1984). Relationship Between Substrate Concentration and Fermentation Product Ratios in Clostridium thermocellum Cultures. *Appl. Envir. Microbiol.*, *47*(5), 1126-1129. Retrieved from http://aem.asm.org/cgi/content/abstract/47/5/1126

Brush, E. J., Lipsett, K. A., & Kozarich, J. W. (1988). Inactivation of Escherichia coli pyruvate formate-lyase by hypophosphite: evidence for a rate-limiting phosphorus-hydrogen bond cleavage. *Biochemistry*, *27*(6), 2217-2222. doi:10.1021/bi00406a061

Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, *84*(6), 647-657.

Burgard, A. P., Pharkya, P., & Maranas, C. D. (2004). OptStrain: a computational framework for redesign of microbial production systems. *Genome research*, *14*(11), 2367-76. doi:10.1101/gr.2872004

Calusinska, M., Happe, T., Joris, B., & Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology (Reading, England)*, *156*(Pt 6), 1575-88. doi:10.1099/mic.0.032771-0

المنارة للاستشارات

www.manaraa.com

Chan, M., Himes, R. H., & Akagi, J. M. (1971). Fatty Acid Composition of Thermophilic, Mesophilic, and Psychrophilic Clostridia. *J. Bacteriol.*, *106*(3), 876-881. Retrieved from http://jb.asm.org/cgi/content/abstract/106/3/876

Chavali, A. K., Whittemore, J. D., Eddy, J. a, Williams, K. T., & Papin, J. a. (2008). Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Molecular systems biology*, *4*(177), 177. doi:10.1038/msb.2008.15

Chinn, M. S., Nokes, S. E., & Strobel, H. J. (2007). Influence of process conditions on end product formation from Clostridium thermocellum 27405 in solid substrate cultivation on paper pulp sludge. *Bioresource technology*, *98*(11), 2184-93. doi:10.1016/j.biortech.2006.08.033

Cho, B.-K., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., Gao, Y., et al. (2009). The transcription unit architecture of the Escherichia coli genome. *Nature biotechnology*, *27*(11), 1043-9. Nature Publishing Group. doi:10.1038/nbt.1582

Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., Cheng, T.-Y., et al. (2009). Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. (J. A. Papin, Ed.)*PLoS computational biology*, *5*(8), e1000489. Public Library of Science. doi:10.1371/journal.pcbi.1000489

Demain, A. L. (2009). Biosolutions to the energy problem. *Journal of industrial microbiology & biotechnology*, *36*(3), 319-32. doi:10.1007/s10295-008-0521-8

Demain, A. L., Newcomb, M., & Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiology and molecular biology reviews : MMBR*, *69*(1), 124-54. doi:10.1128/MMBR.69.1.124-154.2005

Desvaux, M. (2006). Unravelling carbon metabolism in anaerobic cellulolytic bacteria. *Biotechnology progress*, *22*(5), 1229-38. doi:10.1021/bp060016e

Duarte, N. C., Becker, S. a, Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R., et al. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1777-82. doi:10.1073/pnas.0610772104

Duarte, N. C., Herrgård, M. J., & Palsson, B. Ø. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome research*, *14*(7), 1298-309. doi:10.1101/gr.2250904

Edwards, J. S., Covert, M., & Palsson, B. Ø. (2002). Metabolic modelling of microbes: the Flux-balance approach. *Environmental Microbiology*, *4*(3), 133-140.

المنارة للاستشارات

www.manaraa.com

Feist, A. M., Scholten, J. C. M., Palsson, B. Ø., Brockman, F. J., & Ideker, T. (2006). Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. *Molecular systems biology*, *2*, 2006.0004. doi:10.1038/msb4100046

Fong, S. S., & Palsson, B. Ø. (2004). Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. *Nature genetics*, *36*(10), 1056-8. doi:10.1038/ng1432

Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., & Palsson, B. Ø. (2005). In silico design and adaptive evolution ofescherichia coli for production of lactic acid. *Biotechnology and Bioengineering*, *91*(5), 643-648.

Fong, S. S., Marciniak, J. Y., & Palsson, B. Ø. (2003). Description and interpretation of adaptive evolution of Escherichia coli K-12 MG1655 by using a genome-scale in silico metabolic model. *Journal of bacteriology*, *185*(21), 6400-8. doi:10.1128/JB.185.21.6400

Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., & Kitano, H. (2008). CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. *Proceedings of the IEEE*, *96*(8), 1254-1265. doi:10.1109/JPROC.2008.925458

Gerhard, D. S., Wagner, Lukas, Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., Klein, S. L., et al. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome research*, *14*(10B), 2121-7. doi:10.1101/gr.2596504

Gold, N. D., & Martin, V. J. J. (2007). Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. *Journal of bacteriology*, *189*(19), 6787-95. doi:10.1128/JB.00882-07

Gowen, C. M., & Fong, S. S. (2010). Genome-scale metabolic model integrated with RNAseq data to identify metabolic states of Clostridium thermocellum. *Biotechnology journal*, *5*(7), 759-67. doi:10.1002/biot.201000084

Harbers, M., & Carninci, P. (2005). Tag-based approaches for transcriptome research and genome annotation. *Nature methods*, *2*(7), 495-502. doi:10.1038/nmeth768

Herrero, A. A., Gomez, R. F., & Roberts, M. F. (1982). Ethanol-induced changes in the membrane lipid composition of Clostridium thermocellum. *Biochimica et biophysica acta*, *693*(1), 195-204. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7150588

Herrmann, G., Jayamani, E., Mai, G., & Buckel, W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *Journal of bacteriology*, *190*(3), 784-91. doi:10.1128/JB.01422-07

المنارة للاستشارات

www.manaraa.com

Higashide, W., Li, Y., Yang, Y., & Liao, J. C. (2011). Metabolic Engineering of Clostridium cellulolyticum for Production of Isobutanol from Cellulose. *Applied and environmental microbiology*, *77*(8), 2727-33. doi:10.1128/AEM.02454-10

Hogsett, D. A. (1995). *Cellulose Hydrolysis and Fermentation by Clostridium thermocellum for the Production of Ethanol*. Dartmouth College.

Islam, R., Cicek, N., Sparling, R., & Levin, D. (2006). Effect of substrate loading on hydrogen production during anaerobic fermentation by Clostridium thermocellum 27405. *Applied microbiology and biotechnology*, *72*(3), 576-83. Springer Berlin / Heidelberg. doi:10.1007/s00253-006-0316-7

Jensen, P. A., & Papin, J. A. (2009). A scalable systems analysis approach for regulated metabolic networks. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2009*, 5464-5. doi:10.1109/IEMBS.2009.5334060

Johnson, E. A., Madia, A., & Demain, A. L. (1981). Chemically Defined Minimal Medium for Growth of the Anaerobic Cellulolytic Thermophile Clostridium thermocellum. *Applied and environmental microbiology*, *41*(4), 1060-2. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=243859&tool=pmcentrez&rendertype=abstract

Jol, S. J., Kümmel, A., Hatzimanikatis, V., Beard, D. A., & Heinemann, M. (2010). Thermodynamic Calculations for Biochemical Transport and Reaction Processes in Metabolic Networks. *Biophysical journal*, *99*(10), 3139-3144. doi:10.1016/j.bpj.2010.09.043

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27-30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10592173

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic acids research*, *38*(Database issue), D355-60. doi:10.1093/nar/gkp896

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., et al. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research*, *34*(Database issue), D354-7. doi:10.1093/nar/gkj102

Kun, E., & Abood, L. G. (1949). Colorimetric Estimation of Succinic Dehydrogenase by Triphenyltetrazolium Chloride. *Science (New York, N.Y.)*, *109*(2824), 144-6. doi:10.1126/science.109.2824.144

Lalaurette, E., Thammannagowda, S., Mohagheghi, A., Maness, P.-C., & Logan, B. E. (2009). Hydrogen production from cellulose in a two-stage process combining fermentation and

المنارة للاستشارات

www.manaraa.com

electrohydrogenesis. *International Journal of Hydrogen Energy*, *34*(15), 6201-6210. doi:10.1016/j.ijhydene.2009.05.112

Lamed, R. J., Lobos, J. H., & Su, T. M. (1988). Effects of Stirring and Hydrogen on Fermentation Products of Clostridium thermocellum. *Applied and environmental microbiology*, *54*(5), 1216-21. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=202629&tool=pmcentrez&rendertype=abstract

Lee, J., Yun, H., Feist, A. M., Palsson, B. Ø., & Lee, S. Y. (2008). Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. *Applied microbiology and biotechnology*, *80*(5), 849-62. doi:10.1007/s00253-008-1654-4

Lynd, L. R., & Zhang, Y. (2002). Quantitative determination of cellulase concentration as distinct from cell concentration in studies of microbial cellulose utilization: analytical framework and methodological approach. *Biotechnology and bioengineering*, *77*(4), 467-75. doi:10.1002/bit.10142

Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., Himmel, M., et al. (2008). How biotech can transform biofuels. *Nature biotechnology*, *26*(2), 169-72. doi:10.1038/nbt0208-169

Lynd, L. R., van Zyl, W. H., McBride, J. E., & Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Current opinion in biotechnology*, *16*(5), 577-83. doi:10.1016/j.copbio.2005.08.009

Magrane, M., & Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database*, *2011*, bar009-bar009. doi:10.1093/database/bar009

Mahadevan, R. (2003). The effects of alternate optimal solutions in constraint-Based genome-Scale metabolic models. *Metabolic Engineering*, *5*(4), 264-276.

Mahadevan, Radhakrishnan, Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical journal*, *83*(3), 1331-40. doi:10.1016/S0006-3495(02)73903-9

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). Rna-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*(9), 1509-1517.

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., et al. (2010). The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic acids research*, *38*(Database issue), D382-90. doi:10.1093/nar/gkp887

Maskos, U., & Southern, E. M. (1992). Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides

www.manaraa.com

synthesised in situ. *Nucleic acids research*, *20*(7), 1679-84. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=312256&tool=pmcentrez&rendertype=abstract

Mcbee, R. H. (1954). The Characteristics of Clostridium thermocellum. *Journal of Bacteriology*, *67*(4), 505-506.

Newcomb, M., Chen, C.-Y., & Wu, J. H. D. (2007). Induction of the celC operon of Clostridium thermocellum by laminaribiose. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(10), 3747-52. doi:10.1073/pnas.0700087104

Newcomb, M., Millen, J., Chen, C.-Y., & Wu, J. H. D. (2011). Co-transcription of the celC gene cluster in Clostridium thermocellum. *Applied microbiology and biotechnology*, *90*(2), 625-34. doi:10.1007/s00253-011-3121-x

Oberhardt, M. A., Palsson, B. Ø., & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular systems biology*, *5*, 320. doi:10.1038/msb.2009.77

Oh, Y.-K., Palsson, B. Ø., Park, S. M., Schilling, C. H., & Mahadevan, Radhakrishnan. (2007). Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *The Journal of biological chemistry*, *282*(39), 28791-9. doi:10.1074/jbc.M703759200

Patni, N. J., & Alexander, J. K. (1971). Utilization of glucose by Clostridium thermocellum: presence of glucokinase and other glycolytic enzymes in cell extracts. *Journal of bacteriology*, *105*(1), 220-5. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=248344&tool=pmcentrez&rendertype=abstract

Provot, C., Persuy, M. A., & Mercier, J. C. (1989). Complete nucleotide sequence of ovine beta-casein cDNA: inter-species comparison. *Biochimie*, *71*(7), 827-32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2505862

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, *35*(Database issue), D61-5. doi:10.1093/nar/gkl842

Reed, J. L., & Palsson, B. Ø. (2004). Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome research*, *14*(9), 1797-805. doi:10.1101/gr.2546004

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. Ø. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome biology*, *4*(9), R54. doi:10.1186/gb-2003-4-9-r54

المنارة للاستشارات

www.manaraa.com

Roberts, S. B., Gowen, C. M., Brooks, J. P., & Fong, S. S. (2010). Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. *BMC systems biology*, *4*(1), 31. doi:10.1186/1752-0509-4-31

Rydzak, T., Levin, D. B., Cicek, N., & Sparling, R. (2009). Growth phase-dependant enzyme profile of pyruvate catabolism and end-product formation in Clostridium thermocellum ATCC 27405. *Journal of biotechnology*, *140*(3-4), 169-75. doi:10.1016/j.jbiotec.2009.01.022

Saier, M. H., Yen, M. R., Noto, K., Tamang, D. G., & Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic acids research*, *37*(Database issue), D274-8. doi:10.1093/nar/gkn862

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., et al. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, *265*(5596), 687-95. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/870828

Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15112-7. doi:10.1073/pnas.232349399

Senger, R. S. (2010). Biofuel production improvement with genome-scale models: The role of cell composition. *Biotechnology journal*, *5*(7), 671-85. doi:10.1002/biot.201000007

Senger, R. S., & Papoutsakis, E. T. (2008a). Genome-scale model for Clostridium acetobutylicum: Part I. Metabolic network resolution and analysis. *Biotechnology and bioengineering*, *101*(5), 1036-52. doi:10.1002/bit.22010

Senger, R. S., & Papoutsakis, E. T. (2008b). Genome-scale model for Clostridium acetobutylicum: Part II. Development of specific proton flux states and numerically determined sub-systems. *Biotechnology and bioengineering*, *101*(5), 1053-71. doi:10.1002/bit.22009

Shinfuku, Y., Sorpitiporn, N., Sono, M., Furusawa, C., Hirasawa, T., & Shimizu, H. (2009). Development and experimental verification of a genome-scale metabolic model for Corynebacterium glutamicum. *Microbial cell factories*, *8*, 43. doi:10.1186/1475-2859-8-43

Shlomi, T., Cabili, M. N., Herrgard, M. J., Palsson, B. Ø., & Ruppin, E. (2008). Network-Based prediction of human tissue-Specific metabolism. *Nature Biotechnology*, *26*(9), 1003-1010.

Smith, Kevin Michael, Cho, K.-M., & Liao, J. C. (2010). Engineering Corynebacterium glutamicum for isobutanol production. *Applied microbiology and biotechnology*, *87*(3), 1045-55. doi:10.1007/s00253-010-2522-6

Stephanopoulos, G. (2007). Challenges in engineering microbes for biofuels production. *Science (New York, N.Y.)*, *315*(5813), 801-4. doi:10.1126/science.1139612

المنارة للاستشارات

www.manaraa.com

Sun, M., Zhou, G., Lee, S., Chen, J., Shi, R. Z., & Wang, S. M. (2004). SAGE is far more sensitive than EST for detecting low-abundance transcripts. *BMC Genomics*, *5*(1), 1-4.

Szekely, M. (1977). ΦX174 sequenced. *Nature*, *265*(5596), 685-685. doi:10.1038/265685a0

Taylor, B. L., & Zhulin, I. B. (1999). PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiol. Mol. Biol. Rev.*, *63*(2), 479-506. Retrieved from http://mmbr.asm.org/cgi/content/abstract/63/2/479

Tolman, C. J., Kanodia, S., & Roberts, M. F. (1987). 31P and 13C NMR analyses of the energy metabolism of the thermophilic anaerobe Clostridium thermocellum. *The Journal of biological chemistry*, *262*(23), 11088-96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3611106

Urbanczik, R. (2007). Enumerating constrained elementary flux vectors of metabolic networks. *IET systems biology*, *1*(5), 274-9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17907675

Varma, A., & Palsson, B. Ø. (1993). Metabolic Capabilities of Escherichia coli II. Optimal Growth Patterns. *Journal of Theoretical Biology*, *165*(4), 503-522. doi:10.1006/jtbi.1993.1203

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, *10*(1), 57-63. Nature Publishing Group. doi:10.1038/nrg2484

Wilhelm, B. T., & Landry, J.-R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods (San Diego, Calif.)*, *48*, 249-257. doi:10.1016/j.ymeth.2009.03.016

Yang, L., Cluett, W. R., & Mahadevan, Radhakrishnan. (2011). EMILiO: A Fast Algorithm for Genome-Scale Strain Design. *Metabolic Engineering*. doi:10.1016/j.ymben.2011.03.002

Zhang, Y., & Lynd, L. R. (2003). Quantification of Cell and Cellulase Mass Concentrations during Anaerobic Cellulose Fermentation: Development of an Enzyme-Linked Immunosorbent Assay-Based Method with Application to Clostridium t hermocellum Batch Cultures. *Analytical Chemistry*, *75*(2), 219-227. American Chemical Society. doi:10.1021/ac020271n

Zhang, Z., Schwartz, S., Wagner, L, & Miller, W. (n.d.). A greedy algorithm for aligning DNA sequences. *Journal of computational biology : a journal of computational molecular cell biology*, *7*(1-2), 203-14. doi:10.1089/10665270050081478

المنارة للاستشارات

www.manaraa.com

APPENDIX 1: MODEL CONTENTS

A. Reaction List

Reaction ID	Reaction Name	Pathway	EC	Equation
R_OIVD2	R_2_oxoisovalerate_dehydrogenase__acylating_3_methyl_2_oxobutanoate_	Valine, Leucine, and Isoleucine Metabolism	1.2.4.4	[c] : 3mob + coa + nad --> co2 + ibcoa + nadh
R_OIVD3	R_2_oxoisovalerate_dehydrogenase__acylating_3_methyl_2_oxopentanoate_	Valine, Leucine, and Isoleucine Metabolism	1.2.4.4	[c] : 3mop + coa + nad --> 2mbcoa + co2 + nadh
R_OIVD1	R_2_oxoisovalerate_dehydrogenase__acylating_4_methyl_2_oxopentanoate_	Valine, Leucine, and Isoleucine Metabolism	1.2.4.4	[c] : 4mop + coa + nad <==> co2 + ivcoa + nadh
R_ALDD2x	aldehydedehydrogenase(acetaldehyde,NAD)	Pyruvate Metabolism	1.2.1.3	[c] : acald + h2o + nad --> ac + 2 h + nadh
R_MME	methylmalonyl-CoAepimerase	Miscellaneous	5.1.99.1	[c] : mmcoa-r <==> mmcoa-s
R_PRPPS	phosphoribosylpyrophosphatesynthetase	Pentose Phosphate Pathway	2.7.6.1	[c] : atp + r5p <==> amp + h + prpp
R_TECA2S45	R_glycerol_teichoic_acid__n45___unlinked__D_ala_substituted	Cell Envelope Biosynthesis	.	[c] : 45 ala-d + 45 atp + 45 cdpglyc + 45 h2o + uacgam + uacmam --> amp + cmp + gtca2_45_ct + h + ppi + udp + ump
R_IPMD	_3-isopropylmalatedehydrogenase	Valine, Leucine, and Isoleucine Metabolism	1.1.1.85	[c] : 3c2hmp + nad --> 3c4mop + h + nadh
R_PFK	phosphofructokinase	Glycolysis/Gluconeogenesis	2.7.1.11	[c] : atp + f6p --> adp + fdp + h
R_PANTS	pantothenatesynthase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	6.3.2.1	[c] : ala-b + atp + pant-r --> amp + h + pnto-r + ppi
R_ALATRS	Alanyl-tRNAsynthetase	Alanine and Aspartate Metabolism	6.1.1.7	[c] : ala-l + atp + trnaala --> alatrna + amp + ppi
R_H2CO3D	carboxylicaciddissociation	Miscellaneous	4.2.1.1	[c] : co2 + h2o <==> h2co3
R_GLYCK	glyceratekinase	Glycolysis/Gluconeogenesis	2.7.1.31	[c] : atp + glyc-r --> 3pg + adp + h
R_ADCL	_4-aminobenzoatesynthase	Folate Metabolism	4.1.3.38	[c] : 4adcho --> 4abz + h + pyr
R_CTPS1	CTPsynthase(NH3)	Pyrimidine Metabolism	6.3.4.2	[c] : atp + nh4 + utp --> adp + ctp + 2 h + pi
R_NH4t	ammoniareversibletransport	Transport, Extracellular	.	nh4[e] <==> nh4[c]
R_FACOAL150_ISO_	R_fatty_acid__CoA_ligase__Iso_C150_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa3 <==> amp + fa3coa + ppi
R_SPMS	spermidinesynthase	Urea Cycle/Amino Group Metabolism	2.5.1.16	[c] : ametam + ptrc --> 5mta + h + spmd
R_PTHPS	_6-pyruvoyltetrahydropterinsynthase	Folate Metabolism	4.2.3.12	[c] : ahdt --> 6pthp + pppi
R_SERt2r	L-serinereversibletransportvia	Transport, Extracellular	.	h[e] + ser-l[e] <==> h[c] + ser-l[c]

www.manaraa.com

	protonsymport			
R_TRPt2r	L-tryptophanreversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + trp-l[e] <==> h[c] + trp-l[c]
R_HPPK	_2-amino-4-hydroxy-6-hydroxymethyldihydropteridinediphosphokinase	Folate Metabolism	2.7.6.3	[c] : 2ahhmp + atp --> 2ahhmd + amp + h
R_bglu2	betaglucosidase	Cellulose Metabolism	3.2.1.21	[c] : cellobiose + h2o --> 2 glc-d
R_G1PCTR	glucose-1-phosphatecytidylyltransferase	Sugar Metabolism	2.7.7.33	[c] : ctp + g1p + h <==> cdpglu + ppi
R_SERTRS	Seryl-tRNAsynthetase	Glycine, Serine, and Threonine Metabolism	6.1.1.11	[c] : atp + ser-l + trnaser --> amp + ppi + sertrna
R_NNATr	nicotinate-nucleotideadenylyltransferase	Vitamins and Cofactors - NAD Metabolism	2.7.7.18	[c] : atp + h + nicrnt <==> dnad + ppi
R_DHORD10	dihydooroticaciddehydrogenase	Pyrimidine Metabolism	1.3.3.1	[c] : dhor-s + nad --> orot + nadh + h
R_LIPO4S24_CT	R_lipoteichoic_acid_synthesis__n24___linked__unsubstituted	Cell Envelope Biosynthesis	.	[c] : 24 cdpglyc + d12dg_ct --> cmp + h + lipo4_24_ct
R_ASNTRS	Asparaginyl-tRNAsynthetase	Alanine and Aspartate Metabolism	6.1.1.22	[c] : asn-l + atp + trnaasn --> amp + asntrna + ppi
R_ACS2	acetyl-CoAsynthase(propionate)	Miscellaneous	6.2.1.1	[c] : atp + coa + ppa --> amp + ppcoa + ppi
R_CA2abc	calciumtransportviaABCsystem	Transport, Extracellular	.	atp[c] + ca2[e] + h2o[c] --> adp[c] + ca2[c] + h[c] + pi[c]
R_DGK1	deoxyguanylatekinase(dGMP:ATP)	Nucleobase Metabolism	2.7.4.8	[c] : atp + dgmp <==> adp + dgdp
R_cellobtx	cellobiosetransport	Transport, Extracellular	.	cellobiose[e] + atp[c] + h2o[c] --> cellobiose[c] + adp[c] + pi[c] + h[c]
R_PRAGSr	phosphoribosylglycinamidesynthase	Purine Metabolism	6.3.4.13	[c] : atp + gly + pram <==> adp + gar + h + pi
R_CHTBNASE	beta-N-acetylhexosaminidase	Aminosugar Metabolism	3.2.1.52	[c] : ctobiose + h2o --> 2 acgam
R_PPM2	phosphopentomutase2(deoxyribose)	Pentose Phosphate Pathway	5.4.2.7	[c] : 2dr1p <==> 2dr5p
R_ACGAMK	N-acetylglucosaminekinase	Aminosugar Metabolism	2.7.1.59	[c] : acgam + atp --> acgam6p + adp + h
R_ARBabc	L-arabinosetransportviaABCsystem	Transport, Extracellular	3.6.3.17	arab-l[e] + atp[c] + h2o[c] --> adp[c] + arab-l[c] + h[c] + pi[c]
R_FACOAL141	R_fatty_acid__CoA_ligase__tetradecenoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + ttdcea <==> amp + ppi + tdecoa
R_FACOAL140	R_fatty_acid__CoA_ligase__tetradecanoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + ttdca <==> amp + ppi + tdcoa
R_CYSt2r	L-cysteinereversibletransportviaprotonsymport	Transport, Extracellular	.	cys-l[e] + h[e] <==> cys-l[c] + h[c]
R_TRPTRS	Tryptophanyl-tRNAsynthetase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.1.1.2	[c] : atp + trnatrp + trp-l --> amp + ppi + trptrna
R_PAPSR	phosphoadenylyl-sulfatereductase(thioredoxin)	Cysteine Metabolism	1.8.4.8	[c] : paps + trdrd --> 2 h + pap + so3 + trdox
R_FACOAL171_anteiso_	R_fatty_acid__CoA_ligase__anteiso_C171_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa10 <==> amp + fa10coa + ppi

www.manaraa.com

R_NTP7	nucleoside-triphosphatase(UTP)	Pyrimidine Metabolism	3.6.1.15	[c] : h2o + utp --> h + pi + udp
R_NTD9	_5'-nucleotidase(GMP)	Nucleotides	3.1.3.5	[c] : gmp + h2o --> gsn + pi
R_NTD8	_5'-nucleotidase(dGMP)	Nucleotides	3.1.3.5	[c] : dgmp + h2o --> dgsn + pi
R_ACODA	acetylornithinedeacetylase	Urea Cycle/Amino Group Metabolism	3.5.1.14	[c] : acorn + h2o --> ac + orn
R_cellpp	cellopentaosephosphorylase	Cellulose Metabolism	2.4.1.49	[c] : cellopentaose + pi --> g1p + cellotetraose
R_NTD7	_5'-nucleotidase(AMP)	Nucleotides	3.1.3.5	[c] : amp + h2o --> adn + pi
R_NTD6	_5'-nucleotidase(dAMP)	Nucleotides	3.1.3.5	[c] : damp + h2o --> dad-2 + pi
R_ALAR	alanineracemase	Alanine and Aspartate Metabolism	5.1.1.1	[c] : ala-l <==> ala-d
R_GUAD	guaninedeaminase	Purine Metabolism	3.5.4.3	[c] : gua + h + h2o --> nh4 + xan
R_GLUDy	glutamatedehydrogenase(NADP)	Glutamate Metabolism	1.4.1.4	[c] : glu-l + h2o + nadp <==> akg + h + nadph + nh4
R_PYRt2r	pyruvatereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + pyr[e] <==> h[c] + pyr[c]
R_PRAMPC	phosphoribosyl-AMPcyclohydrolase	Histidine Metabolism	3.5.4.19	[c] : h2o + prbamp --> prfp
R_ALAt2r	L-alaninereversibletransportviaprotonsymport	Transport, Extracellular	.	ala-l[e] + h[e] <==> ala-l[c] + h[c]
R_G5SADr	L-glutamate5-semialdehydedehydratase,reversible	Arginine and Proline Metabolism	.	[c] : glu5sa <==> 1pyr5c + h + h2o
R_cellulase3	cellulose_breakdown	Cellulose Metabolism	3.2.1.4 3.2.1.91	[e] : 3 cellulose + h2o --> cellotriose
R_cellulase2	cellulose_breakdown	Cellulose Metabolism	3.2.1.4 3.2.1.91	[e] : 2 cellulose + h2o --> cellobiose
R_cellulase5	cellulose_breakdown	Cellulose Metabolism	3.2.1.4 3.2.1.91	[e] : 5 cellulose + h2o --> cellopentaose
R_cellulase4	cellulose_breakdown	Cellulose Metabolism	3.2.1.4 3.2.1.91	[e] : 4 cellulose + h2o --> cellotetraose
R_cellulase6	cellulose_breakdown	Cellulose Metabolism	3.2.1.4 3.2.1.91	[e] : 6 cellulose + h2o --> cellohexaose
R_TEICH45	R_teichuronic_acid__n45__unlinked__GalNAc_GlcA_repeated	Cell Envelope Biosynthesis	.	[c] : 45 udpacgal + 45 udpglcur <==> 45 h + teich_45_ct + 45 udp + 45 ump
R_GLUt2r	L-glutamatetransportviaprotonsymport,reversible	Transport, Extracellular	.	glu-l[e] + h[e] <==> glu-l[c] + h[c]
R_ASP1DC	aspartate1-decarboxylase	Alanine and Aspartate Metabolism	4.1.1.11	[c] : asp-l + h --> ala-b + co2
R_GALabc	D-galactosetransportviaABCsystem	Transport, Extracellular	3.6.3.17	atp[c] + gal[e] + h2o[c] --> adp[c] + gal[c] + h[c] + pi[c]
R_ACHBS	_2-aceto-2-hydroxybutanoatesynthase	Valine, Leucine, and Isoleucine Metabolism	2.2.1.6	[c] : 2obut + h + pyr --> 2ahbut + co2
R_UDPG4E	UDPglucose4-epimerase	Nucleotide Sugar Metabolism	5.1.3.2	[c] : udpg <==> udpgal
R_FMETTRS	Methionyl-tRNAformyltransferase	Methionine Metabolism	2.1.2.9	[c] : 10fthf + mettrna --> fmettrna + h + thf
R_GLUTRS	Glutamyl-tRNAsynthetase	Glutamate Metabolism	6.1.1.17	[c] : atp + glu-l + trnaglu --> amp + glutrna + ppi
R_GLUTRR	glutamyl-tRNAreductase	Vitamins and Cofactors - Porphyrin Metabolism	1.2.1.70	[c] : glutrna + h + nadph --> glu1sa + nadp + trnaglu

المنارة للاستشارات

www.manaraa.com

R_ASPCT	aspartatecarbamoyltransferase	Pyrimidine Metabolism	2.1.3.2	[c] : asp-l + cbp --> cbasp + h + pi
R_G5SD	glutamate-5-semialdehydedehydrogenase	Arginine and Proline Metabolism	1.2.1.41	[c] : glu5p + h + nadph --> glu5sa + nadp + pi
R_GLYK	glycerolkinase	Glycolysis/Gluconeogenesis	2.7.1.30	[c] : atp + glyc --> adp + glyc3p + h
R_ANS	anthranilatesynthase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.1.3.27	[c] : chor + gln-l --> anth + glu-l + h + pyr
R_HSTPT	histidinol-phosphatetransaminase	Histidine Metabolism	2.6.1.9	[c] : glu-l + imacp --> akg + hisp
R_TRDR	thioredoxinreductase(NADPH)	Pyrimidine Metabolism	1.8.1.9	[c] : h + nadph + trdox --> nadp + trdrd
R_cellottx	cellotreosetransport	Transport, Extracellular	.	cellotriose[e] + atp[c] + h2o[c] --> cellotriose[c] + adp[c] + pi[c] + h[c]
R_ADSL2r	adenylosuccinatelyase	Purine Metabolism	4.3.2.2	[c] : 25aics <==> aicar + fum
R_THRTRS	Threonyl-tRNAsynthetase	Threonine and Lysine Metabolism	6.1.1.3	[c] : atp + thr-l + trnathr --> amp + ppi + thrtrna
R_MALt2r	L-malatereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + mal-l[e] <==> h[c] + mal-l[c]
R_FACOAL150_anteiso_	R_fatty_acid__CoA_ligase__anteiso_C150_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa4 <==> amp + fa4coa + ppi
R_GLCP	glycogenphosphorylase	Glycogen Metabolism	2.4.1.1	[c] : glycogen + pi --> g1p
R_TMDS	thymidylatesynthase	Pyrimidine Metabolism	2.1.1.45	[c] : dump + mlthf --> dhf + dtmp
R_DPCOAK	dephospho-CoAkinase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	2.7.1.24	[c] : atp + dpcoa --> adp + coa + h
R_DRPA	deoxyribose-phosphatealdolase	Pentose Phosphate Pathway	4.1.2.4	[c] : 2dr5p --> acald + g3p
R_FRUK	fructose-1-phosphatekinase	Glycolysis/Gluconeogenesis	2.7.1.56	[c] : atp + f1p --> adp + fdp + h
R_UDCPKr	undecaprenolkinase(reversible)	Cell Envelope Biosynthesis	2.7.1.66	[c] : atp + udcp <==> adp + h + udcpp
R_DMATT	dimethylallyltranstransferase	Sterol Metabolism	2.5.1.10	[c] : dmpp + ipdp --> grdp + ppi
R_POR2_i	pyruvatesynthase	Pyruvate Metabolism	1.2.7.1	[c] : coa + 2 fdox + pyr --> accoa + co2 + 2 fdred + h
R_UAGCVT	UDP-N-acetylglucosamine1-carboxyvinyltransferase	Cell Envelope Biosynthesis	2.5.1.7	[c] : pep + uacgam --> pi + uaccg
R_cellohtx	cellohexaosetransport	Transport, Extracellular	.	cellohexaose[e] + atp[c] + h2o[c] --> cellohexaose[c] + adp[c] + pi[c] + h[c]
R_DHDPRy	dihydrodipicolinatereductase(NADPH)	Threonine and Lysine Metabolism	1.3.1.26	[c] : 23dhdp + h + nadph --> nadp + thdp
R_GF6PTA	glutamine-fructose-6-phosphatetransaminase	Aminosugar Metabolism	2.6.1.16	[c] : f6p + gln-l --> gam6p + glu-l
R_NO3t7	nitratetransportinvianitriteantiport	Transport, Extracellular	.	no2[c] + no3[e] --> no2[e] + no3[c]
R_PROTRS	Prolyl-tRNAsynthetase	Arginine and Proline Metabolism	6.1.1.15	[c] : atp + pro-l + trnapro --> amp + ppi + protrna
R_HSK	homoserinekinase	Glycine, Serine, and Threonine Metabolism	2.7.1.39	[c] : atp + hom-l --> adp + h + phom
R_PHEt2r	L-phenylalaninereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + phe-l[e] <==> h[c] + phe-l[c]
R_ACCOAC	R_acetyl_CoA_carboxylase	Fatty Acid Metabolism	6.4.1.2 6.3.4.14	[c] : accoa + atp + hco3 --> adp + h + malcoa + pi

المنارة للاستشارات

www.manaraa.com

R_TPI	triose-phosphateisomerase	Glycolysis/Gluconeogenesis	5.3.1.1	[c] : dhap <==> g3p
R_GRTT	geranyltranstransferase	Sterol Metabolism	2.5.1.10	[c] : grdp + ipdp --> frdp + ppi
R_DDPGA	_2-dehydro-3-deoxy-phosphogluconatealdolase	Miscellaneous	4.1.1.3	[c] : 4h2oglt <==> glx + pyr
R_ASPTA	aspartatetransaminase	Alanine and Aspartate Metabolism	2.6.1.1	[c] : akg + asp-l <==> glu-l + oaa
R_NDPK8	nucleoside-diphosphatekinase(ATP:dADP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + dadp <==> adp + datp
R_NDPK7	nucleoside-diphosphatekinase(ATP:dCDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + dcdp <==> adp + dctp
R_NDPK5	nucleoside-diphosphatekinase(ATP:dGDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + dgdp <==> adp + dgtp
R_NDPK4	nucleoside-diphosphatekinase(ATP:dTDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + dtdp <==> adp + dttp
R_NDPK3	nucleoside-diphosphatekinase(ATP:CDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + cdp <==> adp + ctp
R_NDPK2	nucleoside-diphosphatekinase(ATP:UDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + udp <==> adp + utp
R_NDPK1	nucleoside-diphosphatekinase(ATP:GDP)	Nucleobase Metabolism	2.7.4.6	[c] : atp + gdp <==> adp + gtp
R_DUTPDP	dUTPdiphosphatase	Pyrimidine Metabolism	3.6.1.23	[c] : dutp + h2o --> dump + h + ppi
R_GGTT	R_geranylgeranyltranstransferase	Sterol Metabolism	.	[c] : ggdp + ipdp --> pendp + ppi
R_FORt3	formatetransportoutviaprotonantiport	Transport, Extracellular	.	for[c] + h[e] --> for[e] + h[c]
R_DAGK_CT	R_diacylglycerol_kinase	Cell Envelope Biosynthesis	2.7.1.107	[c] : 12dgr_ct + atp --> 12dag3p_ct + adp + h
R_LEUTAi	leucinetransaminase(irreversible)	Valine, Leucine, and Isoleucine Metabolism	2.6.1.42	[c] : 4mop + glu-l --> akg + leu-l
R_LACZ	b-galactosidase	Sugar Metabolism	3.2.1.108	[c] : h2o + lcts --> gal + glc-d
R_ENO	enolase	Glycolysis/Gluconeogenesis	4.2.1.11	[c] : 2pg <==> h2o + pep
R_DAPabc	M-diaminopimelicacidABCtransport	Transport, Extracellular	.	26dap-m[e] + atp[c] + h2o[c] --> 26dap-m[c] + adp[c] + h[c] + pi[c]
R_G3POA_CT	R_glycerol_3_phosphate_O_acyltransferase	Cell Envelope Biosynthesis	.	[c] : glyc3p + 0.051 dcacoa + 0.022 trdacoa + 0.023 fa1coa + 0.0092 tdcoa + 0.022 tdecoa + 0.323 fa6coa + 0.2293 pmtcoa + 0.1334 fa12coa + 0.1255 hdcoa + 0.0479 stcoaiso + 0.0137 stcoa --> 1ag3p_ct + coa
R_TYRTA	tyrosinetransaminase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.6.1.1 2.6.1.9	[c] : akg + tyr-l <==> 34hpp + glu-l
R_ASAD	aspartate-semialdehydedehydrogenase	Threonine and Lysine Metabolism	1.2.1.11	[c] : aspsa + nadp + pi <==> 4pasp + h + nadph
R_GLUR	glutamateracemase	Cell Envelope Biosynthesis	5.1.1.3	[c] : glu-d <==> glu-l
R_DCTPD2	dCTPdeaminase	Pyrimidine Metabolism	3.5.4.13	[c] : ctp + h + h2o --> nh4 + utp
R_AMAOTr	adenosylmethionine-8-amino-7-oxononanoatetransaminase	Vitamins and Cofactors - Biotin Metabolism	2.6.1.62	[c] : 8aonn + amet <==> amob + dann

المنارة للاستشارات

www.manaraa.com

R_RNTR2	ribonucleoside-triphosphatereductase(GTP)	Purine Metabolism	1.17.4.2	[c] : gtp + trdrd --> dgtp + h2o + trdox
R_RNTR3	ribonucleoside-triphosphatereductase(CTP)	Pyrimidine Metabolism	1.17.4.2	[c] : ctp + trdrd --> dctp + h2o + trdox
R_UGT_CT	R_UDP_glucosyltransferase__diglucosyl_	Cell Envelope Biosynthesis	.	[c] : 12dgr_ct + 2 udpg --> d12dg_ct + 2 h + 2 udp
R_KAS180iso	R_b_ketoacyl_synthetase__Iso_C180_	Fatty Acid Metabolism	2.3.1.41	[c] : 20 h + ibcoa + 7 malcoa + 14 nadph --> 7 co2 + 8 coa + ocdcaiso + 6 h2o + 14 nadp
R_ADSL1r	adenylsuccinatelyase	Purine Metabolism	4.3.2.2	[c] : dcamp <==> amp + fum
R_PTAr	phosphotransacetylase	Pyruvate Metabolism	2.3.1.8	[c] : accoa + pi <==> actp + coa
R_OCBT	ornithinecarbamoyltransferase	Urea Cycle/Amino Group Metabolism	2.1.3.3	[c] : cbp + orn <==> citr-l + h + pi
R_CYTDK1	cytidinekinase(ATP)	Pyrimidine Metabolism	2.7.1.48	[c] : atp + cytd --> adp + cmp + h
R_CYTDK2	cytidinekinase(GTP)	Pyrimidine Metabolism	2.7.1.48	[c] : cytd + gtp --> cmp + gdp + h
R_IPDDI	isopentenyl-diphosphateD-isomerase	Sterol Metabolism	5.3.3.2	[c] : ipdp <==> dmpp
R_ACBIPGT	Adenosylcobinamidephosphateguanyltransferase	Vitamins and Cofactors - Porphyrin Metabolism	2.7.7.62	[c] : adocbip + gtp + h --> agdpcbi + ppi
R_GMPS	GMPsynthase	Purine Metabolism	6.3.4.1	[c] : atp + nh4 + xmp --> amp + gmp + 2 h + ppi
R_CO2t	CO2transporterviadiffusion	Transport, Extracellular	.	co2[e] <==> co2[c]
R_GBEZ	_1,4-alpha-glucanbranchingenzyme	Glycogen Metabolism	2.4.1.18	[c] : 14glun --> glycogen + h2o
R_ASPTRS	Aspartyl-tRNAsynthetase	Alanine and Aspartate Metabolism	6.1.1.12	[c] : asp-l + atp + trnaasp --> amp + asptrna + ppi
R_OORr	_2-oxoglutaratesynthase(reversible)	Citric Acid Cycle	1.2.7.3	[c] : akg + coa + 2 fdox <==> co2 + 2 fdred + h + succoa
R_KAS100iso	R_b_ketoacyl_synthetase__branched(iso)_C100	Fatty Acid Metabolism	2.3.1.41	[c] : 8 h + ibcoa + 3 malcoa + 6 nadph --> 3 co2 + 4 coa + dca + 2 h2o + 6 nadp
R_H2td	hydrogentransport	Transport, Extracellular	.	h2[c] <==> h2[e]
R_HSST	homoserineO-succinyltransferase	Methionine Metabolism	2.3.1.46	[c] : hom-l + succoa --> coa + suchms
R_GLGC	glucose-1-phosphateadenylyltransferase	Glycolysis/Gluconeogenesis	2.7.7.27	[c] : atp + g1p + h --> adpglc + ppi
R_FOLD3	dihydropteroatesynthase	Folate Metabolism	2.5.1.15	[c] : 2ahhmd + 4abz --> dhpt + ppi
R_PTA2	Phosphateacetyltransferase	Pyruvate Metabolism	2.3.1.8	[c] : pi + ppcoa --> coa + ppap
R_FDH	formatedehydrogenase	Pyruvate Metabolism	1.2.1.2	[c] : for + nad --> co2 + nadh
R_PRMICIi	_1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino)imidazole-4-carboxamideisomerase(irreversible)	Histidine Metabolism	5.3.1.16	[c] : prfp --> prlp
R_H2Ot	H2Otransportviadiffusion	Transport, Extracellular	.	h2o[e] <==> h2o[c]
R_ASPO5	L-aspartateoxidase	Vitamins and Cofactors - NAD Metabolism	1.4.3.16	[c] : asp-l + fum --> iasp + succ
R_MGt5	magnesiumtransportin/outviapermease(noH+)	Transport, Extracellular	.	mg2[c] <==> mg2[e]
R_UGMDDS	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alaninesynthetase	Cell Envelope Biosynthesis	6.3.2.10	[c] : alaala + atp + ugmd --> adp + h + pi + ugmda

www.manaraa.com

R_VALTA	valinetransaminase	Valine, Leucine, and Isoleucine Metabolism	2.6.1.42	[c] : akg + val-l <==> 3mob + glu-l
R_DAPDC	diaminopimelatedecarboxylase	Threonine and Lysine Metabolism	4.1.1.20	[c] : 26dap-m + h --> co2 + lys-l
R_PHETA1	phenylalaninetransaminase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.6.1.57	[c] : akg + phe-l <==> glu-l + phpyr
R_UPP3S	uroporphyrinogen-IIIsynthase	Porphyrin Metabolism	4.2.1.75	[c] : hmbil --> h2o + uppg3
R_IGPS	indole-3-glycerol-phosphatesynthase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.1.1.48	[c] : 2cpr5p + h --> 3ig3p + co2 + h2o
R_PPTGS_CT	R_Peptidoglycan_subunit_synthesis	Cell Envelope Biosynthesis	.	[c] : uaagmda --> h + peptido_ct + udcpdp
R_ARGDC	argininedecarboxylase	Urea Cycle/Amino Group Metabolism	4.1.1.19	[c] : arg-l + h --> agm + co2
R_FACOAL160_ISO_	R_fatty_acid__CoA_ligase__Iso_C160_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa6 <==> amp + fa6coa + ppi
R_ILETA	isoleucinetransaminase	Valine, Leucine, and Isoleucine Metabolism	2.6.1.42	[c] : akg + ile-l <==> 3mop + glu-l
R_PGM	phosphoglyceratemutase	Glycolysis/Gluconeogenesis	5.4.2.1	[c] : 2pg <==> 3pg
R_PRO1x_r	prolineoxidase(L-proline,NAD)	Arginine and Proline Metabolism	1.5.1.2	[c] : nad + pro-l <==> 1pyr5c + 2 h + nadh
R_PGK	phosphoglyceratekinase	Glycolysis/Gluconeogenesis	2.7.2.3	[c] : 3pg + atp <==> 13dpg + adp
R_PGI	glucose-6-phosphateisomerase	Glycolysis/Gluconeogenesis	5.3.1.9	[c] : g6p <==> f6p
R_FMNAT	FMNadenylyltransferase	Vitamins and Cofactors - Riboflavin Metabolism	2.7.7.2	[c] : atp + fmn + h --> fad + ppi
R_TRPS3	tryptophansynthase(indoleglycerolphosphate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.1.20	[c] : 3ig3p --> g3p + indole
R_TRPS2	tryptophansynthase(indole)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.1.20	[c] : indole + ser-l --> h2o + trp-l
R_TRPS1	tryptophansynthase(indoleglycerolphosphate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.1.20	[c] : 3ig3p + ser-l --> g3p + h2o + trp-l
R_Kabc	PotassiumABCtransporter	Transport, Extracellular	.	atp[c] + h2o[c] + k[e] --> adp[c] + h[c] + k[c] + pi[c]
R_FTHFCL	_5-formethyltetrahydrofolatecyclo-ligase	Folate Metabolism	6.3.3.2	[c] : 5fthf + atp --> adp + methf + pi
R_HISDC	histidinedecarboxylase	Histidine Metabolism	4.1.1.22	[c] : h + his-l --> co2 + hista
R_KAS6	R_b_ketoacyl_synthetase__Iso_C160_	Fatty Acid Metabolism	2.3.1.41	[c] : 17 h + ibcoa + 6 malcoa + 12 nadph --> 6 co2 + 7 coa + fa6 + 5 h2o + 12 nadp
R_KAS7	R_b_ketoacyl_synthetase__n_C161_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 19 h + 7 malcoa + 13 nadph --> 7 co2 + 8 coa + 6 h2o + hdcea + 13 nadp
R_KAS4	R_b_ketoacyl_synthetase__Anteiso_C150_	Fatty Acid Metabolism	2.3.1.41	[c] : 2mbcoa + 14 h + 5 malcoa + 10 nadph --> 5 co2 + 6 coa + fa4 + 4 h2o + 10 nadp
R_KAS5	R_b_ketoacyl_synthetase__Iso_C161_	Fatty Acid Metabolism	2.3.1.41	[c] : 16 h + ibcoa + 6 malcoa + 11 nadph --> 6 co2 + 7 coa + fa5 + 5 h2o + 11 nadp
R_KAS2	R_b_ketoacyl_synthetase__Iso_C180_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 17 h + 6 malcoa + 12 nadph --> 6 co2 + 7 coa + 5 h2o + 12 nadp + ttdca
R_KAS3	R_b_ketoacyl_synthetase__I	Fatty Acid Metabolism	2.3.1.41	[c] : 14 h + ivcoa + 5 malcoa + 10 nadph --> 5 co2 + 6 coa

المنارة للاستشارات

www.manaraa.com

	so_C150_			+ fa3 + 4 h2o + 10 nadp
R_KAS1	R_b_ketoacyl_synthetase__Iso_C140_	Fatty Acid Metabolism	2.3.1.41	[c] : 14 h + ibcoa + 5 malcoa + 10 nadph --> 5 co2 + 6 coa + fa1 + 4 h2o + 10 nadp
R_ARGSL	argininosuccinatelyase	Alanine and Aspartate Metabolism	4.3.2.1	[c] : argsuc <==> arg-l + fum
R_SHKK	shikimatekinase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.7.1.71	[c] : atp + skm --> adp + h + skm5p
R_KAS8	R_b_ketoacyl_synthetase__palmitate__n_C160_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 20 h + 7 malcoa + 14 nadph --> 7 co2 + 8 coa + 6 h2o + hdca + 14 nadp
R_KAS9	R_b_ketoacyl_synthetase__Iso_C171_	Fatty Acid Metabolism	2.3.1.41	[c] : 16 h + ivcoa + 6 malcoa + 11 nadph --> 6 co2 + 7 coa + fa9 + 5 h2o + 11 nadp
R_FRUpts2	FructosetransportviaPEP:PyrPTS(f6pgenerating)	Transport, Extracellular	.	fru[e] + pep[c] --> f6p[c] + pyr[c]
R_AHSERL	O-acetylhomoserine(thiol)-lyase	Methionine Metabolism	2.5.1.49	[c] : achms + ch4s --> ac + h + met-l
R_THRt2r	L-threoninereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + thr-l[e] <==> h[c] + thr-l[c]
R_KAS130iso	R_b_ketoacyl_synthetase__branched(iso)_C130	Fatty Acid Metabolism	2.3.1.41	[c] : 11 h + ivcoa + 4 malcoa + 8 nadph --> 4 co2 + 5 coa + trdca + 3 h2o + 8 nadp
R_OAAdecarb	oxaloacetatedecarboxylase	Pyruvate Metabolism	4.1.1.3	[c] : oaa <==> pyr + co2
R_MAN6PI	mannose-6-phosphateisomerase	Sugar Metabolism	5.3.1.8	[c] : man6p <==> f6p
R_PPNCL3	phosphopantothenate-cysteineligase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	6.3.2.5	[c] : 4ppan + atp + cys-l --> 4ppcys + amp + h + ppi
R_UAG4E	R_UDP_N_acetylglucosamine_4_epimerase	Aminosugar Metabolism	5.1.3.7	[c] : uacgam <==> udpacgal
R_LDH_L	L-lactatedehydrogenase	Glycolysis/Gluconeogenesis	1.1.1.27	[c] : lac-l + nad <==> h + nadh + pyr
R_ASPKi	aspartatekinase,irreversible	Alanine and Aspartate Metabolism	2.7.2.4	[c] : asp-l + atp --> 4pasp + adp
R_FACOAL171_ISO_	R_fatty_acid__CoA_ligase__Iso_C171_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa9 <==> amp + fa9coa + ppi
R_APRAUR	_5-amino-6-(5-phosphoribosylamino)uracilreductase	Vitamins and Cofactors - Riboflavin Metabolism	1.1.1.193	[c] : 5apru + h + nadph --> 5aprbu + nadp
R_FRTT	R_farnesyltranstransferase	Sterol Metabolism	.	[c] : frdp + ipdp --> ggdp + ppi
R_PPTT	R_trans_pentaprenyltranstransferase	Sterol Metabolism	.	[c] : ipdp + pendp --> hexdp + ppi
R_EHGLAT	L-erythro-4-Hydroxyglutamate:2-oxoglutarateaminotransferase	Arginine and Proline Metabolism	2.6.1.1	[c] : akg + e4hglu --> 4h2oglt + glu-l
R_ASPt2r	L-aspartatereversibletransportviaprotonsymport	Transport, Extracellular	.	asp-l[e] + h[e] <==> asp-l[c] + h[c]
R_NADS1	NADsynthase(nh3)	Vitamins and Cofactors - NAD Metabolism	6.3.1.5	[c] : atp + dnad + nh4 --> amp + h + nad + ppi
R_AGMT	agmatinase	Urea Cycle/Amino Group Metabolism	3.5.3.11	[c] : agm + h2o --> ptrc + urea
R_PSERT	phosphoserinetransaminase	Glycine, Serine, and Threonine Metabolism	2.6.1.52	[c] : 3php + glu-l --> akg + pser-l

www.manaraa.com

R_PSDC_CT	R_phosphatidylserine_decarboxylase	Cell Envelope Biosynthesis	4.1.1.65	[c] : ps_ct --> co2 + psetha_ct
R_DB4PS	_3,4-Dihydroxy-2-butanone-4-phosphatesynthase	Vitamins and Cofactors - Riboflavin Metabolism	.	[c] : ru5p-d --> db4p + for + h
R_PPCOAC	Propionyl-CoAcarboxylase,mitochondrial	Valine, Leucine, and Isoleucine Metabolism	6.4.1.3	[c] : atp + hco3 + ppcoa --> adp + h + mmcoa-s + pi
R_ADPRDP	ADPribosediphosphatase	Purine Metabolism	3.6.1.13	[c] : adprib + h2o --> amp + 2 h + r5p
R_TALA	transaldolase	Pentose Phosphate Pathway	2.2.1.2	[c] : g3p + s7p <==> e4p + f6p
R_IMPD	IMPdehydrogenase	Purine Metabolism	1.1.1.205	[c] : h2o + imp + nad --> h + nadh + xmp
R_TMDPP	thymidinephosphorylase	Pyrimidine Metabolism	2.4.2.4	[c] : pi + thymd <==> 2dr1p + thym
R_H2St	h2stransport(diffusion)	Transport, Extracellular	.	h2s[e] <==> h2s[c]
R_IMPC	IMPcyclohydrolase	Purine Metabolism	3.5.4.10	[c] : h2o + imp <==> fprica
R_UPPRT	uracilphosphoribosyltransferase	Pyrimidine Metabolism	2.4.2.9	[c] : prpp + ura --> ppi + ump
R_DHORTS	dihydroorotase	Pyrimidine Metabolism	3.5.2.3	[c] : dhor-s + h2o <==> cbasp + h
R_MTRI	_5-methylthioribose-1-phosphateisomerase	Methionine Metabolism	5.3.1.23	[c] : 5mdr1p <==> 5mdru1p
R_ACONT	aconitase	Citric Acid Cycle	4.2.1.3	[c] : cit <==> icit
R_HEX1	hexokinase(D-glucose:ATP)	Glycolysis/Gluconeogenesis	2.7.1.2	[c] : atp + glc-d --> adp + g6p + h
R_HEX7	hexokinase(D-fructose:ATP)	Sugar Metabolism	2.7.1.4	[c] : atp + fru --> adp + f6p + h
R_PPNDH	prephenatedehydratase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.1.51	[c] : h + pphn --> co2 + h2o + phpyr
R_TMDPK	thiaminediphosphokinase	Vitamins and Cofactors - Thiamine Metabolism	2.7.6.2	[c] : atp + thm --> amp + h + thmpp
R_UAMAS	UDP-N-acetylmuramoyl-L-alaninesynthetase	Cell Envelope Biosynthesis	6.3.2.8	[c] : ala-l + atp + uamr --> adp + h + pi + uama
R_PUNP5	purine-nucleosidephosphorylase(Inosine)	Purine Metabolism	2.4.2.1	[c] : ins + pi <==> hxan + r1p
R_PUNP4	purine-nucleosidephosphorylase(Deoxyguanosine)	Purine Metabolism	2.4.2.1 2.4.2.4	[c] : dgsn + pi <==> 2dr1p + gua
R_PUNP7	purine-nucleosidephosphorylase(Xanthosine)	Purine Metabolism	2.4.2.1	[c] : pi + xtsn <==> r1p + xan
R_PUNP6	purine-nucleosidephosphorylase(Deoxyinosine)	Purine Metabolism	2.4.2.1 2.4.2.4	[c] : din + pi <==> 2dr1p + hxan
R_PUNP1	purine-nucleosidephosphorylase(Adenosine)	Purine Metabolism	2.4.2.1	[c] : adn + pi <==> ade + r1p
R_PUNP3	purine-nucleosidephosphorylase(Guanosine)	Purine Metabolism	2.4.2.1	[c] : gsn + pi <==> gua + r1p
R_NNDMBRT	Nicotinate-nucleotidedimethylbenzimidazolephosphoribosyltransferase	Vitamins and Cofactors - Riboflavin Metabolism	2.4.2.21	[c] : dmbzid + nicrnt --> 5prdmbz + h + nac
R_XANt2r	xanthinetransportinviaproto	Transport, Extracellular	.	h[e] + xan[e] <==> h[c] + xan[c]

المنارة للاستشارات

www.manaraa.com

	nsymport, reversible			
R_LIPO2S24_CT	R_lipoteichoic_acid_synthesis__n24___linked__N_acetylglucosamine_substituted	Cell Envelope Biosynthesis	.	[c] : 24 cdpglyc + d12dg_ct + 24 uacgam --> 24 cmp + 48 h + lipo2_24_ct + 24 udp
R_KAS141	R_b_ketoacyl_synthetase__n_C141_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 16 h + 6 malcoa + 11 nadph --> 6 co2 + 7 coa + 5 h2o + 11 nadp + ttdcea
R_L_DASH_LACt4r	L-lactatereversibletransportviasodiumsymport	Transport, Extracellular	.	lac-l[e] + na1[e] <==> lac-l[c] + na1[c]
R_BIOMASS	Total biomass flux including extracellular protein	Biomass	.	cellmass[c] + 1.3 cellome[e] --> biomass[c]
R_PACCOAL2	phenylacetate-CoAligase(hydroxyphenylacetate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.2.1.30	[c] : 4hphac + atp + coa --> amp + hphaccoa + ppi
R_PACCOAL3	phenylacetate-CoAligase(indoleacetate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.2.1.30	[c] : atp + coa + ind3ac --> amp + indaccoa + ppi
R_UREAabc	ureatransportviaABCsystem	Transport, Extracellular	.	atp[c] + h2o[c] + urea[e] --> adp[c] + h[c] + urea[c] + pi[c]
R_AMAA	N-acetylmuramoyl-L-alanineamidase	Cell Envelope Biosynthesis	3.5.1.28	[c] : acmama + h2o --> acmam + ala-l
R_CDPMEK	_4-(cytidine5'-diphospho)-2-C-methyl-D-erythritolkinase	Sterol Metabolism	2.7.1.148	[c] : 4c2me + atp --> 2p4c2me + adp + h
R_UAMAGS	UDP-N-acetylmuramoyl-L-alanyl-D-glutamatesynthetase	Cell Envelope Biosynthesis	6.3.2.9	[c] : atp + glu-d + uama --> adp + h + pi + uamag
R_MTHFR3	_5,10-methylenetetrahydrofolatereductase(NADPH)	Folate Metabolism	1.5.1.20	[c] : 2 h + mlthf + nadph --> 5mthf + nadp
R_RNTR1	ribonucleoside-triphosphatereductase(ATP)	Purine Metabolism	1.17.4.2	[c] : atp + trdrd --> datp + h2o + trdox
R_ARGSSr	argininosuccinatesynthase,reversible	Urea Cycle/Amino Group Metabolism	6.3.4.5	[c] : asp-l + atp + citr-l <==> amp + argsuc + h + ppi
R_CODHr	carbonmonoxidedehydrogenase/acetyl-CoAsynthase	Miscellaneous	1.2.99.2	[c] : accoa + 2 fdox + h2o + h4spt <==> co2 + coa + 2 fdred + 2 h + mh4spt
R_SPTi	serine-pyruvateaminotransferase(irreversible),peroxisomal	Glycine, Serine, and Threonine Metabolism	2.6.1.51	[c] : pyr + ser-l <==> ala-l + hpyr
R_GLCt5	glucose transport via atp	Transport, Extracellular	.	glc-d[e] + atp[c] + h2o[c] --> glc-d[c] + adp[c] + pi[c] + h[c]
R_GAPD	glyceraldehyde-3-phosphatedehydrogenase	Glycolysis/Gluconeogenesis	1.2.1.12	[c] : g3p + nad + pi <==> 13dpg + h + nadh
R_CYSTGL	cystathionineg-lyase	Cysteine Metabolism	4.4.1.1	[c] : cyst-l + h2o --> 2obut + cys-l + nh4
R_LIPO1S24_CT	R_lipoteichoic_acid_synthesis__n24___linked__glucose_substituted	Cell Envelope Biosynthesis	.	[c] : 24 cdpglyc + d12dg_ct + 24 udpg --> 24 cmp + 48 h + lipo1_24_ct + 24 udp
R_RNDR4	ribonucleoside-diphosphatereductase(UDP)	Pyrimidine Metabolism	1.17.4.1	[c] : trdrd + udp --> dudp + h2o + trdox
R_RNDR2	ribonucleoside-diphosphatereductase(GDP)	Purine Metabolism	1.17.4.1	[c] : gdp + trdrd --> dgdp + h2o + trdox
R_FACOAL140_ISO_	R_fatty_acid__CoA_ligase__Iso_C140_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa1 <==> amp + fa1coa + ppi
R_RNDR1	ribonucleoside-diphosphatereductase(ADP)	Purine Metabolism	1.17.4.1	[c] : adp + trdrd --> dadp + h2o + trdox
R_PABB	pabb	Folate Metabolism	2.6.1.85	[c] : chor + nh4 --> 4adcho + h2o

المنارة للاستشارات
www.manaraa.com

R_SHSL4r	O-succinylhomoserinelyase(elimination),reversible	Methionine Metabolism	2.5.1.48	[c] : h2o + suchms <==> 2obut + h + nh4 + succ
R_GARFTi	phosphoribosylglycinamideformyltransferase,irreversible	Purine Metabolism	2.1.2.2	[c] : 10fthf + gar --> fgam + h + thf
R_CHORS	chorismatesynthase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.3.5	[c] : 3psme --> chor + pi
R_ACGK	acetylglutamatekinase	Urea Cycle/Amino Group Metabolism	2.7.2.8	[c] : acglu + atp --> acg5p + adp
R_PGAMT	phosphoglucosaminemutase	Cell Envelope Biosynthesis	5.4.2.10	[c] : gam1p <==> gam6p
R_ACGS	N-acetylglutamatesynthase	Urea Cycle/Amino Group Metabolism	2.3.1.1	[c] : accoa + glu-l --> acglu + coa + h
R_PIt2r	phosphatereversibletransportviasymport	Transport, Extracellular	.	h[e] + pi[e] <==> h[c] + pi[c]
R_ATPPRT	ATPphosphoribosyltransferase	Histidine Metabolism	2.4.2.17	[c] : atp + prpp --> ppi + prbatp
R_RBFSb	riboflavinsynthase	Vitamins and Cofactors - Riboflavin Metabolism	2.5.1.9	[c] : 2 dmlz --> 4r5au + ribflv
R_RBFSa	riboflavinsynthase	Vitamins and Cofactors - Riboflavin Metabolism	2.5.1.9	[c] : 4r5au + db4p --> dmlz + 2 h2o + pi
R_IOR	indolepyruvateferredoxinoxidoreductase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	1.2.7.8	[c] : coa + 2 fdox + indpyr <==> co2 + 2 fdred + h + indaccoa
R_GLYAT	glycineC-acetyltransferase	Glycine, Serine, and Threonine Metabolism	2.3.1.29	[c] : accoa + gly <==> 2aobut + coa
R_CODH2	carbonmonoxidedehydrogenase/acetyl-CoAsynthase2	Miscellaneous	1.2.99.2	[c] : co + 2 fdox + h2o --> co2 + 2 fdred + 2 h
R_PPS	phosphoenolpyruvatesynthase	Glycolysis/Gluconeogenesis	2.7.9.2	[c] : atp + h2o + pyr --> amp + 2 h + pep + pi
R_G3PCT	R_glycerol_3_phosphate_cytidylyltransferase	Cell Envelope Biosynthesis	2.7.7.39	[c] : ctp + glyc3p + h --> cdpglyc + ppi
R_FUM	fumarase	Citric Acid Cycle	4.2.1.2	[c] : fum + h2o <==> mal-l
R_ORNTAC	ornithinetransacetylase	Urea Cycle/Amino Group Metabolism	2.3.1.35	[c] : acorn + glu-l <==> acglu + orn
R_PTPATi	pantetheine-phosphateadenylyltransferase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	2.7.7.3	[c] : atp + h + pan4p --> dpcoa + ppi
R_PPA	inorganicdiphosphatase	Miscellaneous	3.6.1.1	[c] : h2o + ppi --> h + 2 pi
R_PPCDC	phosphopantothenoylcysteinedecarboxylase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	4.1.1.36	[c] : 4ppcys + h --> co2 + pan4p
R_PPM	phosphopentomutase	Pentose Phosphate Pathway	5.4.2.2	[c] : r1p <==> r5p
R_IOR3	indolepyruvateferredoxinoxidoreductase(hydroxyphenylpyvurate)	Tyrosine, Tryptophan, and Phenylalanine Metabolism	1.2.7.8	[c] : 34hpp + coa + 2 fdox <==> co2 + 2 fdred + h + hphaccoa
R_AHSERL2	O-acetylhomoserine(thiol)-lyase	Methionine Metabolism	2.5.1.49	[c] : achms + h2s --> ac + h + hcys-l
R_NTP3e	nucleoside-triphosphatase(GTP)	Purine Metabolism	3.6.1.15	[e] : gtp + h2o --> gdp + h + pi
R_LYSTRS	Lysyl-tRNAsynthetase	Threonine and Lysine Metabolism	6.1.1.6	[c] : atp + lys-l + trnalys --> amp + lystrna + ppi
R_ADSK	adenylyl-sulfatekinase	Purine Metabolism	2.7.1.25	[c] : aps + atp --> adp + h + paps

المنارة للاستشارات
www.manaraa.com

R_bglu1	betaglucosidase,cellulose	Cellulose Metabolism	3.2.1.21	[e] : cellulose + h2o --> glc-d
R_ADSS	adenylosuccinatesynthase	Purine Metabolism	6.3.4.4	[c] : asp-l + gtp + imp --> dcamp + gdp + 2 h + pi
R_bglu3	betaglucosidase	Cellulose Metabolism	3.2.1.21	[c] : cellotriose + h2o --> glc-d + cellobiose
R_bglu4	betaglucosidase	Cellulose Metabolism	3.2.1.21	[c] : cellotetraose + h2o --> glc-d + cellotriose
R_bglu5	betaglucosidase	Cellulose Metabolism	3.2.1.21	[c] : cellopentaose + h2o --> glc-d + cellotetraose
R_bglu6	betaglucosidase	Cellulose Metabolism	3.2.1.21	[c] : cellohexaose + h2o --> glc-d + cellopentaose
R_BACCL	biotin-[acetyl-CoA-carboxylase]ligase	Vitamins and Cofactors - Biotin Metabolism	6.3.4.15	[c] : atp + btn + h --> btamp + ppi
R_DTMPK	dTMPkinase	Nucleobase Metabolism	2.7.4.9	[c] : atp + dtmp <==> adp + dtdp
R_DHNPA	dihydroneopterinaldolase	Folate Metabolism	4.1.2.25	[c] : dhnpt --> 2ahhmp + gcald + h
R_NTP5	nucleoside-triphosphatase(CTP)	Pyrimidine Metabolism	3.6.1.15	[c] : ctp + h2o --> cdp + h + pi
R_NTP4	nucleoside-triphosphatase(dGTP)	Purine Metabolism	3.6.1.15	[c] : dgtp + h2o --> dgdp + h + pi
R_ACOTA	acetylornithinetransaminase	Urea Cycle/Amino Group Metabolism	2.6.1.11	[c] : acorn + akg <==> acg5sa + glu-l
R_NTP6	nucleoside-triphosphatase(dCTP)	Pyrimidine Metabolism	3.6.1.15	[c] : dctp + h2o --> dcdp + h + pi
R_NTP1	nucleoside-triphosphatase(ATP)	Purine Metabolism	3.6.1.15	[c] : atp + h2o --> adp + h + pi
R_SHK3D	shikimatedehydrogenase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	1.1.1.25	[c] : 3dhsk + h + nadph --> nadp + skm
R_NTP3	nucleoside-triphosphatase(GTP)	Purine Metabolism	3.6.1.5	[c] : gtp + h2o --> gdp + h + pi
R_NTP2	nucleoside-triphosphatase(dATP)	Purine Metabolism	3.6.1.15	[c] : datp + h2o --> dadp + h + pi
R_DHQD	_3-dehydroquinatedehydratase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	4.2.1.10	[c] : 3dhq <==> 3dhsk + h2o
R_LYSt2r	L-lysinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + lys-l[e] <==> h[c] + lys-l[c]
R_NTP9	nucleoside-triphosphatase(dTTP)	Pyrimidine Metabolism	3.6.1.15	[c] : dttp + h2o --> dtdp + h + pi
R_NTP8	nucleoside-triphosphatase(dUTP)	Pyrimidine Metabolism	3.6.1.15	[c] : dutp + h2o --> dudp + h + pi
R_MOHMT	_3-methyl-2-oxobutanoatehydroxymethyltransferase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	2.1.2.11	[c] : 3mob + h2o + mlthf --> 2dhp + thf
R_ACLS	acetolactatesynthase	Valine, Leucine, and Isoleucine Metabolism	2.2.1.6	[c] : h + 2 pyr --> alac-s + co2
R_PAP_CT	R_phosphatidic_acid_phosphatase	Cell Envelope Biosynthesis	3.1.3.4	[c] : 12dag3p_ct + h2o --> 12dgr_ct + pi
R_ASNt2r	L-asparaginereversibletransportviaprotonsymport	Transport, Extracellular	.	asn-l[e] + h[e] <==> asn-l[c] + h[c]
R_2HBO	_2-Hydroxybutyrate:NAD+oxidoreductase	Miscellaneous	1.1.1.27	[c] : 2hb + nad <==> 2obut + h + nadh
R_FACOAL170_anteiso_	R_fatty_acid__CoA_ligase__anteiso_C170_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa12 <==> amp + fa12coa + ppi
R_DHQS	_3-dehydroquinatesynthase	Tyrosine, Tryptophan,	4.2.3.4	[c] : 2dda7p --> 3dhq + pi

المنارة للاستشارات
www.manaraa.com

		and Phenylalanine Metabolism		
R_GLCS1	glycogensynthase(ADPGlc)	Glycogen Metabolism	2.4.1.21	[c] : adpglc --> adp + glycogen + h
R_ARGTRS	Arginyl-tRNAsynthetase	Arginine and Proline Metabolism	6.1.1.19	[c] : arg-l + atp + trnaarg --> amp + argtrna + ppi
R_GTPDPK	GTPdiphosphokinase	Purine Metabolism	2.7.6.5	[c] : atp + gtp --> amp + gdptp + h
R_ATPS2_i	ATPsynthase(twoprotonsforoneATP)	Miscellaneous	3.6.3.14	atp[c] + 2 h[c] + h2o[c] --> adp[c] + 2 h[e] + pi[c]
R_ILEt2r	L-isoleucinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + ile-l[e] <==> h[c] + ile-l[c]
R_TECA4S_CT	R_minor_teichoic_acid_synthesis__n30_	Cell Envelope Biosynthesis	.	[c] : 30 h2o + 30 udpacgal + 30 udpg --> 60 h + tcam_ct + 30 udp + 30 ump
R_PAPPT3	phospho-N-acetylmuramoyl-pentapeptide-transferase(meso-2,6-diaminopimelate)	Cell Envelope Biosynthesis	2.7.8.13	[c] : udcpp + ugmda --> uagmda + ump
R_SPMDabc	spermidinetransportviaABCsystem	Transport, Extracellular	3.6.3.31	atp[c] + h2o[c] + spmd[e] --> adp[c] + h[c] + pi[c] + spmd[c]
R_PPBNGS	porphobilinogensynthase	Porphyrin Metabolism	4.2.1.24	[c] : 2 5aop --> h + 2 h2o + ppbng
R_CYTK1	cytidylatekinase(CMP)	Nucleobase Metabolism	2.7.4.14	[c] : atp + cmp <==> adp + cdp
R_HMBS	hydroxymethylbilanesynthase	Porphyrin Metabolism	2.5.1.61	[c] : h2o + 4 ppbng --> hmbil + 4 nh4
R_URAt2	uraciltransportinviaprotonsymport	Transport, Extracellular	.	h[e] + ura[e] --> h[c] + ura[c]
R_ADOCBLS	Adenosylcobalamin5'-phosphatesynthase	Vitamins and Cofactors - Porphyrin Metabolism	2.7.8.26	[c] : agdpcbi + rdmbzi --> adocbl + gmp + h
R_URIDK1r	uridylatekinase(UMP)	Nucleobase Metabolism	2.7.4.22	[c] : atp + ump <==> adp + udp
R_UPP3MT	uroporphyrinogenmethyltransferase	Porphyrin Metabolism	2.1.1.107	[c] : 2 amet + uppg3 --> 2 ahcys + dscl + h
R_THRS	threoninesynthase	Glycine, Serine, and Threonine Metabolism	4.2.3.1	[c] : h2o + phom --> pi + thr-l
R_P5CR	pyrroline-5-carboxylatereductase	Urea Cycle/Amino Group Metabolism	1.5.1.2	[c] : 1pyr5c + 2 h + nadph --> nadp + pro-l
R_AKP1	alkalinephosphatase(Dihydroneopterin)	Folate Metabolism	3.1.3.1	[c] : ahdt + 3 h2o --> dhnpt + 2 h + 3 pi
R_AGPATr_CT	R_1_acylglycerol_3_phosphate_O_acyltransferase	Cell Envelope Biosynthesis	2.3.1.51	[c] : 1ag3p_ct + 0.051 dcacoa + 0.022 trdacoa + 0.023 fa1coa + 0.0092 tdcoa + 0.022 tdecoa + 0.323 fa6coa + 0.2293 pmtcoa + 0.1334 fa12coa + 0.1255 hdcoa + 0.0479 stcoaiso + 0.0137 stcoa --> 12dag3p_ct + coa
R_KARA1	ketol-acidreductoisomerase(2,3-dihydroxy-3-methylbutanoate)	Valine, Leucine, and Isoleucine Metabolism	1.1.1.86	[c] : 23dhmb + nadp <==> alac-s + h + nadph
R_KARA2	ketol-acidreductoisomerase(2-Acetolactate)	Valine, Leucine, and Isoleucine Metabolism	1.1.1.86	[c] : 2ahbut + h + nadph <==> 23dhmp + nadp
R_G3PD1	glycerol-3-phosphatedehydrogenase(NAD)	Glycolysis/Gluconeogenesis	1.1.1.94	[c] : glyc3p + nad <==> dhap + h + nadh
R_NMNTP	nmntp	Transport, Extracellular	.	h[e] + nmn[e] --> h[c] + nmn[c]
R_ARGt2r	L-arganinereversibletransportviaprotonsymport	Transport, Extracellular	.	arg-l[e] + h[e] <==> arg-l[c] + h[c]

www.manaraa.com

R_DXPS	_1-deoxy-D-xylulose5-phosphatesynthase	Sterol Metabolism	2.2.1.7	[c] : g3p + h + pyr --> co2 + dxyl5p
R_GALU	UTP-glucose-1-phosphateuridylyltransferase	Nucleotide Sugar Metabolism	2.7.7.9	[c] : g1p + h + utp <==> ppi + udpg
R_IGPDH	imidazoleglycerol-phosphatedehydratase	Histidine Metabolism	4.2.1.19	[c] : eig3p --> h2o + imacp
R_MGSA	methylglyoxalsynthase	Pyruvate Metabolism	4.2.3.3	[c] : dhap --> mthgxl + pi
R_DAPDH	diaminopimelatedehydrogenase	Threonine and Lysine Metabolism	1.4.1.16	[c] : 2 h + nadph + nh4 + thdp --> 26dap-m + nadp
R_TDPDRE	dTDP-4-dehydrorhamnose3,5-epimerase	Nucleotide Sugar Metabolism	5.1.3.13	[c] : dtdp4d6dg --> dtdp4d6dm
R_ILETRS	Isoleucyl-tRNAsynthetase	Valine, Leucine, and Isoleucine Metabolism	6.1.1.5	[c] : atp + ile-l + trnaile --> amp + iletrna + ppi
R_PIabc	phosphatetransportviaABCsystem	Transport, Extracellular	.	atp[c] + h2o[c] + pi[e] --> adp[c] + h[c] + 2 pi[c]
R_LEUTRS	Leucyl-tRNAsynthetase	Valine, Leucine, and Isoleucine Metabolism	6.1.1.4	[c] : atp + leu-l + trnaleu --> amp + leutrna + ppi
R_ACNAM9PL	N-Acetylneuraminate9-phosphatepyruvate-lyase(pyruvate-phosphorylating)	Aminosugar Metabolism	2.5.1.56 2.5.1.57	[c] : acmanap + h2o + pep --> acnamp + pi
R_METabc	L-methioninetransportviaABCsystem	Transport, Extracellular	.	atp[c] + h2o[c] + met-l[e] --> adp[c] + h[c] + met-l[c] + pi[c]
R_ETOHt2r	ethanolreversibletransportviaprotonsymport	Transport, Extracellular	.	etoh[e] + h[e] <==> etoh[c] + h[c]
R_GK1	guanylatekinase(GMP:ATP)	Nucleobase Metabolism	2.7.4.8	[c] : atp + gmp <==> adp + gdp
R_GLYTRS	Glycyl-tRNAsynthetase	Glycine, Serine, and Threonine Metabolism	6.1.1.14	[c] : atp + gly + trnagly --> amp + glytrna + ppi
R_PMDPHT	pyrimidinephosphatase	Vitamins and Cofactors - Riboflavin Metabolism	3.1.3.-	[c] : 5aprbu + h2o --> 4r5au + pi
R_AGDC_r	N-acetylglucosamine-6-phosphatedeacetylase(reversible)	Aminosugar Metabolism	3.5.1.25	[c] : acgam6p + h2o <==> ac + gam6p
R_AIRCr	phosphoribosylaminoimidazolecarboxylase	Purine Metabolism	4.1.1.21	[c] : air + co2 <==> 5aizc + h
R_BIFHASE	Bifurcating hydrogenase	Miscellaneous	.	[c] : 3 h + 2 fdred + nadh <==> fdox + nad + 2 h2
R_GLNt2r	L-glutaminereversibletransportviaprotonsymport	Transport, Extracellular	.	gln-l[e] + h[e] <==> gln-l[c] + h[c]
R_XYLabc	D-xylosetransportviaABCsystem	Transport, Extracellular	3.6.3.17	atp[c] + h2o[c] + xyl-d[e] --> adp[c] + h[c] + pi[c] + xyl-d[c]
R_PGPPH_CT	R_Phosphatidylglycerophosphate_phosphohydrolase	Cell Envelope Biosynthesis	3.1.3.27	[c] : h2o + pglyp_ct --> pgly_ct + pi
R_OHPBAT	O-Phospho-4-hydroxy-L-threonine:2-oxoglutarateaminotransferase	Vitamins and Cofactors - B6 Metabolism	2.6.1.52	[c] : glu-l + ohpb <==> akg + phthr
R_NO3R1	Nitratereductase(Ubiquinol-8)	Miscellaneous	1.7.99.4	2 h[c] + no3[c] + q8h2[c] --> 2 h[e] + h2o[c] + no2[c] + q8[c]
R_AICART	phosphoribosylaminoimidazolecarboxamideformyltransferase	Purine Metabolism	2.1.2.3	[c] : 10fthf + aicar <==> fprica + thf

www.manaraa.com

R_GLYt2r	glycinereversibletransportviaprotonsymport	Transport, Extracellular	.	gly[e] + h[e] <==> gly[c] + h[c]
R_PHCYT_CT	R_phosphatidate_cytidylyltransferase	Cell Envelope Biosynthesis	2.7.7.41	[c] : 12dag3p_ct + ctp + h --> cdpdag_ct + ppi
R_UGT1_CT	R_UDP_glucosyltransferase__monoglucosyl_	Cell Envelope Biosynthesis	.	[c] : 12dgr_ct + udpg --> h + m12dg_ct + udp
R_FDH2	formatedehydrogenase(quinone-8:2protons)	Miscellaneous	1.2.2.1	for[c] + 3 h[c] + q8[c] --> co2[c] + 2 h[e] + q8h2[c]
R_URIK3	uridinekinase(ITP:Uridine)	Pyrimidine Metabolism	2.7.1.48	[c] : itp + uri --> h + idp + ump
R_PEPCKr	Phosphoenolpyruvatecarboxykinase	Glycolysis/Gluconeogenesis	4.1.1.32	[c] : atp + oaa <==> co2 + adp + pep
R_AMID3	amidase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	3.5.1.4	[c] : h2o + iad --> ind3ac + nh4
R_AMID2	amidase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	3.5.1.4	[c] : h2o + pad --> nh4 + pac
R_QULNS	quinolinatesynthase	Vitamins and Cofactors - NAD Metabolism	4.1.99.-	[c] : dhap + iasp --> h + 2 h2o + pi + quln
R_Nat1	SodiumexportviaATPase	Transport, Extracellular	.	atp[c] + na1[c] + h2o[c] --> adp[c] + na1[e] + h[c] + pi[c]
R_DHFS	dihydrofolatesynthase	Folate Metabolism	6.3.2.17	[c] : atp + dhpt + glu-l --> adp + dhf + pi
R_CDPGDH	CDP-glucose46dehydratase	Sugar Metabolism	4.2.1.45	[c] : cdpglu --> cdpddglu + h2o
R_FACOAL161	R_fatty_acid__CoA_ligase__hexadecenoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + hdcea <==> amp + hdcoa + ppi
R_FACOAL160	R_fatty_acid__CoA_ligase__hexadecanoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + hdca <==> amp + pmtcoa + ppi
R_TDPGDH	dTDPglucose4,6-dehydratase	Nucleotide Sugar Metabolism	4.2.1.46	[c] : dtdpglu --> dtdp4d6dg + h2o
R_LEUt2r	L-leucinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + leu-l[e] <==> h[c] + leu-l[c]
R_FTHFL	formate-tetrahydrofolateligase	Folate Metabolism	3.5.4.9 6.3.4.3	[c] : atp + for + thf <==> 10fthf + adp + pi
R_CLPNS2_CT	R_cardiolipin_synthase	Cell Envelope Biosynthesis	2.7.8.-	[c] : 2 pgly_ct --> cdlp_ct + glyc
R_MECDPDH	_2C-methyl-D-erythritol2,4cyclodiphosphatedehydratase	Sterol Metabolism	1.17.4.3	[c] : 2mecdp + h --> h2mb4p + h2o
R_LYSLG_CT	R_lysylphosphatidylglycerol_synthesis	Cell Envelope Biosynthesis	.	[c] : atp + lys-l + pgly_ct --> amp + h + lysylpgly_ct + ppi
R_ALCD2x	alcoholdehydrogenase(ethanol)	Pyruvate Metabolism	1.1.1.1	[c] : etoh + nad <==> acald + h + nadh
R_NTD4	_5'-nucleotidase(CMP)	Pyrimidine Catabolism	3.1.3.5	[c] : cmp + h2o --> cytd + pi
R_VALTRS	Valyl-tRNAsynthetase	Valine, Leucine, and Isoleucine Metabolism	6.1.1.9	[c] : atp + trnaval + val-l --> amp + ppi + valtrna
R_METS	methioninesynthase	Methionine Metabolism	2.1.1.13	[c] : 5mthf + hcys-l --> h + met-l + thf
R_METTRS	methionyl-tRNAsynthetase	Methionine Metabolism	6.1.1.10	[c] : atp + met-l + trnamet --> amp + mettrna + ppi
R_URIDK2r	uridylatekinase(dUMP)	Purine Metabolism	2.7.4.9	[c] : atp + dump <==> adp + dudp
R_PGLYCP	Phosphoglycolatephosphatase	Miscellaneous	3.1.3.18	[c] : 2pglyc + h2o --> glyclt + pi
R_UAAGDS	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelatesynthetase	Cell Envelope Biosynthesis	6.3.2.13	[c] : 26dap-m + atp + uamag --> adp + h + pi + ugmd

www.manaraa.com

R_FACOAL181	R_fatty_acid__CoA_ligase__octadecenoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + ocdcea <==> amp + odecoa + ppi
R_FACOAL180	R_fatty_acid__CoA_ligase__octadecanoate_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + ocdca <==> amp + ppi + stcoa
R_METB1r	metb1(rev)	Methionine Metabolism	2.5.1.49	[c] : achms + cys-l <==> ac + cyst-l + h
R_NBAHH	Nalpha-(beta-alanyl)-L-histidinehydrolase	Alanine and Aspartate Metabolism	3.4.13.3	[c] : carn + h2o <==> ala-b + his-l
R_PDX5PS	Pyridoxine5'-phosphatesynthase	Vitamins and Cofactors - B6 Metabolism	.	[c] : dxyl5p + nad + phthr --> co2 + h + 2 h2o + nadh + pdx5p + pi
R_PFL	pyruvateformatelyase	Pyruvate Metabolism	2.3.1.54	[c] : coa + pyr --> accoa + for
R_BTS3r	biotinsynthase	Vitamins and Cofactors - Biotin Metabolism	2.8.1.6	[c] : dtbt + 2 s <==> btn + h2s
R_RZ5PP	alpha-ribazole5-phosphatephosphatase	Vitamins and Cofactors - Riboflavin Metabolism	3.1.3.73	[c] : 5prdmbz + h2o --> pi + rdmbzi
R_CDGPT_CT	R_CDPdiacylglycerolsn_glycerol_3_phosphate_3_phosphatidyltransferase	Cell Envelope Biosynthesis	2.7.8.5	[c] : cdpdag_ct + glyc3p --> cmp + h + pglyp_ct
R_UAPGR	UDP-N-acetylenolpyruvoylglucosaminereductase	Cell Envelope Biosynthesis	1.1.1.158	[c] : h + nadph + uaccg --> nadp + uamr
R_PPND	prephenatedehydrogenase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	1.3.1.12	[c] : nad + pphn --> 34hpp + co2 + nadh
R_DDPA	_3-deoxy-D-arabino-heptulosonate7-phosphatesynthetase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.5.1.54	[c] : e4p + h2o + pep --> 2dda7p + pi
R_TMPPP	thiamine-phosphatediphosphorylase	Vitamins and Cofactors - Thiamine Metabolism	2.5.1.3	[c] : 2mahmp + 4mpetz + h --> ppi + thmmp
R_cellttp	cellotetraosephosphorylase	Cellulose Metabolism	2.4.1.49	[c] : cellotetraose + pi --> g1p + cellotriose
R_ANPRT	anthranilatephosphoribosyltransferase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.4.2.18	[c] : anth + prpp --> ppi + pran
R_UAG2E	UDP-N-acetylglucosamine2-epimerase	Cell Envelope Biosynthesis	5.1.3.14	[c] : uacgam <==> uacmam
R_PSCVTi	_3-phosphoshikimate1-carboxyvinyltransferase,irreversible	Tyrosine, Tryptophan, and Phenylalanine Metabolism	2.5.1.19	[c] : pep + skm5p --> 3psme + pi
R_SULabc	sulfatetransportviaABCsystem	Transport, Extracellular	3.6.3.30	atp[c] + h2o[c] + so4[e] --> adp[c] + h[c] + pi[c] + so4[c]
R_GCALDt	Glycoaldehydyereversibletransport	Transport, Extracellular	.	gcald[e] <==> gcald[c]
R_TECA3S45	R_glycerol_teichoic_acid__n45___unlinked__glucose_substituted	Cell Envelope Biosynthesis	.	[c] : 45 cdpglyc + h2o + uacgam + uacmam + 45 udpg --> cmp + gtca3_45_ct + h + udp + ump
R_PACCOAL	phenylacetate-CoAligase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.2.1.30	[c] : atp + coa + pac --> amp + phaccoa + ppi
R_CYTD	cytidinedeaminase	Pyrimidine Metabolism	3.5.4.5	[c] : cytd + h + h2o --> nh4 + uri
R_PROt4r	prolinetransporter	Transport, Extracellular	.	na1[e] + pro-l[e] <==> na1[c] + pro-l[c]
R_bglu5_ext	betaglucosidase	Cellulose Metabolism	3.2.1.21	[e] : cellopentaose + h2o --> glc-d + cellotetraose
R_MTHFD	methylenetetrahydrofolatedehydrogenase(NADP)	Folate Metabolism	1.5.1.5	[c] : mlthf + nadp <==> methf + nadph
R_MTHFC	methenyltetrahydrofolatecyclohydrolase	Folate Metabolism	3.5.4.9 6.3.4.3	[c] : h2o + methf <==> 10fthf + h

www.manaraa.com

R_UAGPT3	R_UDP_N_acetylglucosamine_N_acetylmuramyl__pentapeptide_pyrophosphoryl_undecaprenol_N_acetylglucosamine_transferase	Aminosugar Metabolism	.	[c] : uacgam + uagmda --> h + uaagmda + udp
R_PGCD	phosphoglyceratedehydrogenase	Glycine, Serine, and Threonine Metabolism	1.1.1.95	[c] : 3pg + nad --> 3php + h + nadh
R_OMCDC	_2-Oxo-4-methyl-3-carboxypentanoatedecarboxylation	Valine, Leucine, and Isoleucine Metabolism	.	[c] : 3c4mop + h --> 4mop + co2
R_MECDPS	_2-C-methyl-D-erythritol2,4-cyclodiphosphatesynthase	Sterol Metabolism	4.6.1.12	[c] : 2p4c2me --> 2mecdp + cmp
R_HEXTT	R_trans_hexaprenyltranstransferase	Sterol Metabolism	.	[c] : hexdp + ipdp --> hepdp + ppi
R_PTRCabc	putrescinetransportviaABCsystem	Transport, Extracellular	3.6.3.31	atp[c] + h2o[c] + ptrc[e] --> adp[c] + h[c] + pi[c] + ptrc[c]
R_FACOAL180_ISO_	R_fatty_acid__CoA_ligase__Iso_C180_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + ocdcaiso <==> amp + ppi + stcoaiso
R_NAPRT	NAPRTase	Vitamins and Cofactors - NAD Metabolism	2.4.2.11	[c] : h + nac + prpp --> nicrnt + ppi
R_CHTNASE1	chitinase,ptl	Aminosugar Metabolism	3.2.1.14	[c] : 2 chtn + h2o --> ctobiose
R_PGMT	phosphoglucomutase	Glycolysis/Gluconeogenesis	5.4.2.2	[c] : g1p <==> g6p
R_DHFR	dihydrofolatereductase	Folate Metabolism	1.5.1.3	[c] : dhf + h + nadph <==> nadp + thf
R_UDPDPS	undecaprenyl-diphosphatesynthase	Sterol Metabolism	2.5.1.31	[c] : decdp + ipdp --> ppi + udcpdp
R_MEPCT	_2-C-methyl-D-erythritol4-phosphatecytidylyltransferase	Sterol Metabolism	2.7.7.60	[c] : 2me4p + ctp + h --> 4c2me + ppi
R_ADOCBIK	Adenosylcobinamidekinase	Vitamins and Cofactors - Porphyrin Metabolism	2.7.1.156	[c] : adocbi + atp --> adocbip + adp + h
R_G1SATi	glutamate-1-semialdehydeaminotransferase	Vitamins and Cofactors - Porphyrin Metabolism	5.4.3.8	[c] : glu1sa --> 5aop
R_PPAKr	Propionatekinase	Pyruvate Metabolism	2.7.2.1	[c] : adp + ppap <==> atp + ppa
R_GLNS	glutaminesynthetase	Glutamate Metabolism	6.3.1.2	[c] : atp + glu-l + nh4 --> adp + gln-l + h + pi
R_HXPRT	hypoxanthinephosphoribosyltransferase(Hypoxanthine)	Purine Metabolism	2.4.2.8	[c] : hxan + prpp --> imp + ppi
R_ACt2r	acetatereversibletransportviaprotonsymport	Transport, Extracellular	.	ac[e] + h[e] <==> ac[c] + h[c]
R_celloptx	cellopentaosetransport	Transport, Extracellular	.	cellopentaose[e] + atp[c] + h2o[c] --> cellopentaose[c] + adp[c] + pi[c] + h[c]
R_ME1	malicenzyme(NAD)	Pyruvate Metabolism	1.1.1.38 4.1.1.3	[c] : mal-l + nad --> co2 + nadh + pyr
R_G1PTT	glucose-1-phosphatethymidylyltransferase	Nucleotide Sugar Metabolism	2.7.7.33	[c] : dttp + g1p + h --> dtdpglu + ppi
R_ACNMSr	N-acetylneuraminate synthase(reversible)	Aminosugar Metabolism	2.5.1.56	[c] : acnam + pi <==> acmana + pep + h2o
R_NADK	NADkinase	Vitamins and Cofactors - NAD Metabolism	2.7.1.23	[c] : atp + nad --> adp + h + nadp
R_SUCCt2r	succinatetransportviaprotonsymport	Transport, Extracellular	.	h[e] + succ[e] <==> h[c] + succ[c]
R_IPPMIa	_3-	Valine, Leucine, and	4.2.1.33	[c] : 3c2hmp <==> 2ippm + h2o

www.manaraa.com

	isopropylmalatedehydratase	Isoleucine Metabolism		
R_IPPMIb	_2-isopropylmalatehydratase	Valine, Leucine, and Isoleucine Metabolism	4.2.1.33	[c] : 2ippm + h2o <==> 3c3hmp
R_HISt2r	L-histidinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + his-l[e] <==> h[c] + his-l[c]
R_GHMT2r	glycinehydroxymethyltransferase,reversible	Glycine, Serine, and Threonine Metabolism	2.1.2.1	[c] : ser-l + thf <==> gly + h2o + mlthf
R_bglu2_ext	betaglucosidase	Cellulose Metabolism	3.2.1.21	[e] : cellobiose + h2o --> 2 glc-d
R_DAPE	diaminopimelateepimerase	Threonine and Lysine Metabolism	5.1.1.7	[c] : 26dap-ll <==> 26dap-m
R_PYNP1r	pyrimidine-nucleosidephosphorylase(cytosine)	Pyrimidine Metabolism	2.4.2.2	[c] : csn + r1p <==> cytd + pi
R_TKT1	transketolase	Pentose Phosphate Pathway	2.2.1.1	[c] : r5p + xu5p-d <==> g3p + s7p
R_TKT2	transketolase	Pentose Phosphate Pathway	2.2.1.1	[c] : e4p + xu5p-d <==> f6p + g3p
R_cellotettx	cellotetraosetransport	Transport, Extracellular	.	cellotetraose[e] + atp[c] + h2o[c] --> cellotetraose[c] + adp[c] + pi[c] + h[c]
R_DHAD1	dihydroxy-aciddehydratase(2,3-dihydroxy-3-methylbutanoate)	Valine, Leucine, and Isoleucine Metabolism	4.2.1.9	[c] : 23dhmb --> 3mob + h2o
R_URIK2	uridinekinase(GTP:Uridine)	Pyrimidine Metabolism	2.7.1.48	[c] : gtp + uri --> gdp + h + ump
R_URIK1	uridinekinase(ATP:Uridine)	Pyrimidine Metabolism	2.7.1.48	[c] : atp + uri --> adp + h + ump
R_DHAD2	Dihydroxy-aciddehydratase(2,3-dihydroxy-3-methylpentanoate)	Valine, Leucine, and Isoleucine Metabolism	4.2.1.9	[c] : 23dhmp --> 3mop + h2o
R_FACOAL100	R_fatty_acid__CoA_ligase__Iso_C100_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + dca <==> amp + dcacoa + ppi
R_CITt2r	citratereversibletransportviasymport	Transport, Extracellular	.	cit[e] + h[e] <==> cit[c] + h[c]
R_CMPSAS	CMPsialicacidsynthase	Aminosugar Metabolism	2.7.7.43	[c] : acnam + ctp --> cmpacna + ppi
R_SADT	sulfateadenylyltransferase	Purine Metabolism	2.7.7.4	[c] : atp + h + so4 --> aps + ppi
R_ADKd	adenylatekinase(dform)	Nucleobase Metabolism	2.7.4.3	[c] : damp + datp <==> 2 dadp
R_METAT	methionineadenosyltransferase	Methionine Metabolism	2.5.1.6	[c] : atp + h2o + met-l --> amet + pi + ppi
R_OMPDC	orotidine-5'-phosphatedecarboxylase	Pyrimidine Metabolism	4.1.1.23	[c] : h + orot5p --> co2 + ump
R_DHPPDA	diaminohydroxyphosphoribosylaminopyrimidinedeaminase	Vitamins and Cofactors - Riboflavin Metabolism	3.5.4.26	[c] : 25dhpp + h + h2o --> 5apru + nh4
R_VALt2r	L-valinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + val-l[e] <==> h[c] + val-l[c]
R_CYSS	cysteinesynthase	Cysteine Metabolism	2.5.1.47	[c] : acser + h2s --> ac + cys-l + h
R_SERAT	serineO-acetyltransferase	Cysteine Metabolism	2.3.1.30	[c] : accoa + ser-l <==> acser + coa
R_CBPS	carbamoyl-phosphatesynthase(glutamine-hydrolysing)	Pyrimidine Metabolism	6.3.5.5	[c] : 2 atp + gln-l + h2o + hco3 --> 2 adp + cbp + glu-l + 2 h + pi

المنارة للاستشارات
www.manaraa.com

R_NADOX	hydrogen:NAD+ oxidoreductase	Miscellaneous	1.12.1.2	[c] : h2 + nad <==> h + nadh
R_ADK1	adenylatekinase	Nucleobase Metabolism	2.7.4.3	[c] : amp + atp <==> 2 adp
R00765	D-glucosamine-6-phosphate aminohydrolase (ketol isomerizing)	Aminosugar Metabolism	3.5.99.6	[c] : gam6p + h2o <==> f6p + nh4
R_FACOAL130	R_fatty_acid__CoA_ligase__Iso_C130_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + trdca <==> amp + trdacoa + ppi
R_RNDR3	ribonucleoside-diphosphatereductase(CDP)	Pyrimidine Metabolism	1.17.4.1	[c] : cdp + trdrd --> dcdp + h2o + trdox
R_RIBabc	D-ribosetransportviaABCsystem	Transport, Extracellular	3.6.3.17	atp[c] + h2o[c] + rib-d[e] --> adp[c] + h[c] + pi[c] + rib-d[c]
R_MDH	malatedehydrogenase	Citric Acid Cycle	1.1.1.37	[c] : mal-l + nad <==> h + nadh + oaa
R_RPE	ribulose5-phosphate3-epimerase	Pentose Phosphate Pathway	5.1.3.1	[c] : ru5p-d <==> xu5p-d
R_KAS17	R_b_ketoacyl_synthetase__n_C181_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 22 h + 8 malcoa + 15 nadph --> 8 co2 + 9 coa + 7 h2o + 15 nadp + ocdcea
R_ORPT	orotatephosphoribosyltransferase	Pyrimidine Metabolism	2.4.2.10	[c] : orot5p + ppi <==> orot + prpp
R_KAS11	R_b_ketoacyl_synthetase__Iso_C170_	Fatty Acid Metabolism	2.3.1.41	[c] : 17 h + ivcoa + 6 malcoa + 12 nadph --> 6 co2 + 7 coa + fa11 + 5 h2o + 12 nadp
R_KAS12	R_b_ketoacyl_synthetase__Anteiso_C170_	Fatty Acid Metabolism	2.3.1.41	[c] : 2mbcoa + 17 h + 6 malcoa + 12 nadph --> 6 co2 + 7 coa + fa12 + 5 h2o + 12 nadp
R_KAS13	R_b_ketoacyl_synthetase__n_C180_	Fatty Acid Metabolism	2.3.1.41	[c] : accoa + 23 h + 8 malcoa + 16 nadph --> 8 co2 + 9 coa + 7 h2o + 16 nadp + ocdca
R_RPI	ribose-5-phosphateisomerase	Pentose Phosphate Pathway	5.3.1.6	[c] : r5p <==> ru5p-d
R_OBTFL	_2-Oxobutanoateformatelyase	Pyruvate Metabolism	2.3.1.54	[c] : 2obut + coa --> for + ppcoa
R_ACS	acetyl-CoAsynthetase	Pyruvate Metabolism	6.2.1.1	[c] : ac + atp + coa --> accoa + amp + ppi
R_IPPS	_2-isopropylmalatesynthase	Valine, Leucine, and Isoleucine Metabolism	2.3.3.13	[c] : 3mob + accoa + h2o --> 3c3hmp + coa + h
R_UGT2_CT	R_UDP_glucosyltransferase__triglucosyl_	Cell Envelope Biosynthesis	.	[c] : 12dgr_ct + 3 udpg --> 3 h + t12dg_ct + 3 udp
R_DHDPS	dihydrodipicolinatesynthase	Threonine and Lysine Metabolism	4.2.1.52	[c] : aspsa + pyr --> 23dhdp + h + 2 h2o
R_NIT_n1p4	nitrogenase	Miscellaneous	1.18.6.1	[c] : 32 atp + 8 fdred + 32 h2o + n2 --> 32 adp + 8 fdox + 6 h + h2 + 2 nh4 + 32 pi
R_ASNS2	asparaginesynthetase	Alanine and Aspartate Metabolism	6.3.1.1	[c] : asp-l + atp + nh4 --> amp + asn-l + h + ppi
R_ASNS1	asparaginesynthase(glutamine-hydrolysing)	Alanine and Aspartate Metabolism	6.3.5.4	[c] : asp-l + atp + gln-l + h2o --> amp + asn-l + glu-l + h + ppi
R_UDCPDP	undecaprenyl-diphosphatase	Cell Envelope Biosynthesis	3.6.1.27	[c] : h2o + udcpdp --> h + pi + udcpp
R_ATPS2Na_i	ATPsynthase(twoionsforone ATP)	Transport, Extracellular	.	atp[c] + 2 na1[c] + h2o[c] --> adp[c] + 2 na1[e] + h[c] + pi[c]
R_ALATA_L	L-alaninetransaminase	Glutamate Metabolism	2.6.1.2	[c] : akg + ala-l <==> glu-l + pyr
R_AHC	adenosylhomocysteinase	Methionine Metabolism	3.3.1.1	[c] : ahcys + h2o <==> adn + hcys-l
R_SLDx	L-sulfolactatedehydrogenase(NAD+)	Cysteine Metabolism	1.1.1.272	[c] : nad + sl-l <==> 3spyr + h + nadh
R_NMNATr	nicotinamide-nucleotideadenylyltransferase	Vitamins and Cofactors - NAD Metabolism	2.7.7.18	[c] : atp + h + nmn <==> nad + ppi

www.manaraa.com

R_PFK_2	tagatose 6-phosphate kinase	Sugar Metabolism	2.7.1.144	[c] : atp + tag6p-d --> adp + h + tagdp-d
R_AMID4	amidase	Urea Cycle/Amino Group Metabolism	3.5.1.4	[c] : ad + h2o --> ac + nh4
R_BTSr	biotinsynthase	Vitamins and Cofactors - Biotin Metabolism	2.8.1.6	[c] : dtbt + s <==> btn + 2 h
R_CS	citratesynthase	Citric Acid Cycle	2.3.3.1	[c] : accoa + h2o + oaa --> cit + coa + h
R_UDPGDr	UDPglucose6-dehydrogenase	Nucleotide Sugar Metabolism	1.1.1.22	[c] : h2o + 2 nad + udpg <==> 3 h + 2 nadh + udpglcur
R_NNDPR	nicotinate-nucleotidediphosphorylase(carboxylating)	Vitamins and Cofactors - NAD Metabolism	2.4.2.19	[c] : 2 h + prpp + quln --> co2 + nicrnt + ppi
R_IPDPS	_1-hydroxy-2-methyl-2-(E)-butenyl4-diphosphatereductase(ipdp)	Sterol Metabolism	1.17.1.2	[c] : h + h2mb4p + nadh --> h2o + ipdp + nad
R_ACKr	acetatekinase	Pyruvate Metabolism	2.7.2.1	[c] : ac + atp <==> actp + adp
R_CYSTRS	Cysteinyl-tRNAsynthetase	Cysteine Metabolism	6.1.1.16	[c] : atp + cys-l + trnacys --> amp + cystrna + ppi
R_GLUSy	glutamatesynthase(NADPH)	Glutamate Metabolism	1.4.1.13	[c] : akg + gln-l + h + nadph --> 2 glu-l + nadp
R_NADPOX	hydrogen:NADP+ oxidoreductase	Miscellaneous	1.12.1.3	[c] : h2 + nadp <==> h + nadph
R_ADPT	adeninephosphoribosyltransferase	Purine Metabolism	2.4.2.7 2.4.2.8	[c] : ade + prpp --> amp + ppi
R_FBA	fructose-bisphosphatealdolase	Glycolysis/Gluconeogenesis	4.1.2.13	[c] : fdp <==> dhap + g3p
R_Cut1	CopperexportviaATPase	Transport, Extracellular	3.6.3.4	atp[c] + cu2[c] + h2o[c] --> adp[c] + cu2[e] + h[c] + pi[c]
R_HISTP	histidinol-phosphatase	Histidine Metabolism	3.1.3.15	[c] : h2o + hisp --> histd + pi
R_GLNTRS	Glutaminyl-tRNAsynthetase	Glutamate Metabolism	6.1.1.18	[c] : atp + gln-l + trnagln --> amp + glntrna + ppi
R_IG3PS	Imidazole-glycerol-3-phosphatesynthase	Histidine Metabolism	2.4.2.-	[c] : gln-l + prlp --> aicar + eig3p + glu-l + h
R_PNTK	pantothenatekinase	Vitamins and Cofactors - Pantothenate and CoA Biosynthesis	2.7.1.33	[c] : atp + pnto-r --> 4ppan + adp + h
R_ADNCYC	adenylatecyclase	Pyrimidine Metabolism	4.6.1.1	[c] : atp --> camp + ppi
R_TECA1S45	R_glycerol_teichoic_acid__n45___unlinked__unsubstituted	Cell Envelope Biosynthesis	.	[c] : 45 cdpglyc + uacgam + uacmam --> cmp + gtca1_45_ct + h + udp + ump
R_HISTD	histidinoldehydrogenase	Histidine Metabolism	1.1.1.23	[c] : h2o + histd + 2 nad --> 3 h + his-l + 2 nadh
R_cellhp	cellohexaosephosphorylase	Cellulose Metabolism	2.4.1.49	[c] : cellohexaose + pi --> g1p + cellopentaose
R_DHPS	dihydropteroatesynthase	Folate Metabolism	2.5.1.15	[c] : 2ahhmp + 4abz --> dhpt + h2o
R_CDPDSP_CT	R_CDPdiacylglycerol_serine_O_phosphatidyltransferase	Cell Envelope Biosynthesis	2.7.8.8	[c] : cdpdag_ct + ser-l --> cmp + h + ps_ct
R_BTS2	biotinsynthase(ala-Lproducing)	Vitamins and Cofactors - Biotin Metabolism	2.8.1.6	[c] : cys-l + dtbt <==> ala-l + btn + 2 h
R_TYRTRS	tyrosyl-tRNAsynthetase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.1.1.1	[c] : atp + trnatyr + tyr-l --> amp + ppi + tyrtrna
R_CHORM	chorismatemutase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	5.4.99.5	[c] : chor --> pphn
R_MCOATA	Malonyl-CoA-ACPtransacylase	Fatty Acid Metabolism	2.3.1.39	[c] : acp + malcoa <==> coa + malacp
R_AGPR	N-acetyl-g-glutamyl-phosphatereductase	Urea Cycle/Amino Group Metabolism	1.2.1.38	[c] : acg5sa + nadp + pi <==> acg5p + h + nadph

R_KAS10	R_b_ketoacyl_synthetase__Anteiso_C171_	Fatty Acid Metabolism	2.3.1.41	[c] : 2mbcoa + 16 h + 6 malcoa + 11 nadph --> 6 co2 + 7 coa + fa10 + 5 h2o + 11 nadp
R_FACOAL161_ISO_	R_fatty_acid__CoA_ligase__Iso_C161_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa5 <==> amp + fa5coa + ppi
R_UREA	urease	Urea Cycle/Amino Group Metabolism	3.5.1.5	[c] : 2 h + h2o + urea --> co2 + 2 nh4
R_TYRt2r	L-tyrosinereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + tyr-l[e] <==> h[c] + tyr-l[c]
R_PRAI	phosphoribosylanthranilateisomerase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	5.3.1.24	[c] : pran <==> 2cpr5p
R_FRUpts	D-fructosetransportviaPEP:PyrPTS	Transport, Extracellular	.	fru[e] + pep[c] --> f1p[c] + pyr[c]
R_FACOAL170_ISO_	R_fatty_acid__CoA_ligase__Iso_C170_	Fatty Acid Metabolism	6.2.1.3	[c] : atp + coa + fa11 <==> amp + fa11coa + ppi
R_celltp	cellotriosephosphorylase	Cellulose Metabolism	2.4.1.49	[c] : cellotriose + pi --> g1p + cellobiose
R_HSDxi	homoserinedehydrogenase(NADH),irreversible	Glycine, Serine, and Threonine Metabolism	1.1.1.3	[c] : aspsa + h + nadh --> hom-l + nad
R_PRASCS	phosphoribosylaminoimidazolesuccinocarboxamidesynthase	Purine Metabolism	6.3.2.6	[c] : 5aizc + asp-l + atp <==> 25aics + adp + h + pi
R_UAG2EMA	UDP-N-acetyl-D-glucosamine2-epimerase(Hydrolysis)	Cell Envelope Biosynthesis	5.1.3.14	[c] : h2o + uacgam <==> acmana + h + udp
R_YUMPS	yUMPsynthetase	Pyrimidine Metabolism	4.2.1.70	[c] : r5p + ura <==> h2o + psd5p
R_LIPO3S24_CT	R_lipoteichoic_acid_synthesis__n24___unliked__D_alanine_substituted	Cell Envelope Biosynthesis	.	[c] : 24 ala-d + 24 atp + 24 cdpglyc + d12dg_ct + 24 h2o --> 24 amp + 24 cmp + 48 h + lipo3_24_ct + 24 ppi
R_AOXSr	_8-amino-7-oxononanoatesynthase	Vitamins and Cofactors - Biotin Metabolism	2.3.1.47	[c] : ala-l + h + pmcoa <==> 8aonn + co2 + coa
R_PRFGS	phosphoribosylformylglycinamidinesynthase	Purine Metabolism	6.3.5.3	[c] : atp + fgam + gln-l + h2o --> adp + fpram + glu-l + h + pi
R_G1PACT	glucosamine-1-phosphateN-acetyltransferase	Cell Envelope Biosynthesis	2.3.1.157	[c] : accoa + gam1p --> acgam1p + coa + h
R_ALAALAr	D-alanine-D-alanineligase(reversible)	Cell Envelope Biosynthesis	6.3.2.4	[c] : 2 ala-d + atp <==> adp + alaala + h + pi
R_GLUPRT	glutaminephosphoribosyldiphosphateamidotransferase	Purine Metabolism	2.4.2.14	[c] : gln-l + h2o + prpp --> glu-l + ppi + pram
R_CHTNASE	chitinase	Aminosugar Metabolism	3.2.1.14	[c] : chtn + 2 h2o --> 3 acgam
R_PPDK	Pyruvate,phosphatedikinase.	Pyruvate Metabolism	2.7.9.1	[c] : atp + pi + pyr --> amp + h + pep + ppi
R_UAGDP	UDP-N-acetylglucosaminediphosphorylase	Aminosugar Metabolism	2.7.7.23	[c] : acgam1p + h + utp --> ppi + uacgam
R_PNP	purine-nucleosidephosphorylase	Vitamins and Cofactors - NAD Metabolism	2.4.2.1	[c] : pi + rnam <==> h + ncam + r1p
R_ADMDCr	Adenosylmethioninedecarboxylase	Urea Cycle/Amino Group Metabolism	4.1.1.50	[c] : amet + h <==> ametam + co2
R_RBFK	riboflavinkinase	Vitamins and Cofactors - Riboflavin Metabolism	2.7.1.26	[c] : atp + ribflv --> adp + fmn + h
R_PSP_L	phosphoserinephosphatase(L-serine)	Glycine, Serine, and Threonine Metabolism	3.1.3.3	[c] : h2o + pser-l --> pi + ser-l
R_DBTSr	dethiobiotinsynthase	Vitamins and Cofactors - Biotin Metabolism	6.3.3.3	[c] : atp + co2 + dann <==> adp + dtbt + 3 h + pi

المنارة للاستشارات

www.manaraa.com

R_H2CO3D2	carboxylicaciddissociation	Miscellaneous	.	[c] : h + hco3 <==> h2co3
R_GTPCII	GTPcyclohydrolaseII	Vitamins and Cofactors - Riboflavin Metabolism	3.5.4.25	[c] : gtp + 3 h2o --> 25dhpp + for + 2 h + ppi
R_SUCOAS	succinyl-CoAsynthetase(ADP-forming)	Citric Acid Cycle	6.2.1.5	[c] : atp + coa + succ <==> adp + pi + succoa
R_bglu6_ext	betaglucosidase	Cellulose Metabolism	3.2.1.21	[e] : cellohexaose + h2o --> glc-d + cellopentaose
R_cellbp	cellobiosephosphorylase	Cellulose Metabolism	2.4.1.20	[c] : cellobiose + pi --> g1p + glc-d
R_CYSTS	cystathioninebeta-synthase	Cysteine Metabolism	4.2.1.22	[c] : hcys-l + ser-l --> cyst-l + h2o
R_Nat_4	Na+ exchanging ATPase_4	Transport, Extracellular	3.6.3.15	atp[c] + h2o[c] + 4 na1[c] --> adp[c] + h[c] + 4 na1[e] + pi[c]
R_NNAM	nicotinamidase	Vitamins and Cofactors - NAD Metabolism	3.5.1.19	[c] : h2o + ncam --> nac + nh4
R_PNTOt2	Pantothenatereversibletransportviaprotonsymport	Transport, Extracellular	.	h[e] + pnto-r[e] <==> h[c] + pnto-r[c]
R_FE3abc	iron(III)transportviaABCsystem	Transport, Extracellular	3.6.3.30	atp[c] + fe3[e] + h2o[c] --> adp[c] + fe3[c] + h[c] + pi[c]
R_HISTRS	Histidyl-tRNAsynthetase	Histidine Metabolism	6.1.1.21	[c] : atp + his-l + trnahis --> amp + histrna + ppi
R_CYSTL	cystathionineb-lyase	Methionine Metabolism	4.4.1.8	[c] : cyst-l + h2o --> hcys-l + nh4 + pyr
R_GLU5K	glutamate5-kinase	Urea Cycle/Amino Group Metabolism	2.7.2.11	[c] : atp + glu-l --> adp + glu5p
R_MTAP	_5'-methylthioadenosinephosphorylase	Methionine Metabolism	2.4.2.28	[c] : 5mta + pi --> 5mdr1p + ade
R_bglu4_ext	betaglucosidase	Cellulose Metabolism	3.2.1.21	[e] : cellotetraose + h2o --> glc-d + cellotriose
R_ICDHyr	isocitratedehydrogenase(NADP)	Citric Acid Cycle	1.1.1.42	[c] : icit + nadp <==> akg + co2 + nadph
R_GTPCI	GTPcyclohydrolaseI	Folate Metabolism	3.5.4.16	[c] : gtp + h2o --> ahdt + for + h
R_SHSL2r	O-succinylhomoserinelyase(H2S)	Methionine Metabolism	2.5.1.48	[c] : h2s + suchms <==> h + hcys-l + succ
R_FDXHASE	ferredoxinhydrogenase	Miscellaneous	1.12.7.2	[c] : 2 fdred + 2 h <==> h2 + 2 fdox
R_DCMPDA	dCMPdeaminase	Pyrimidine Metabolism	3.5.4.12	[c] : dcmp + h + h2o <==> dump + nh4
R_ACALD	acetaldehydedehydrogenase(acetylating)	Pyruvate Metabolism	1.2.1.10	[c] : acald + coa + nad <==> accoa + h + nadh
R_CBIAT	Cobinamideadenyltransferase	Vitamins and Cofactors - Porphyrin Metabolism	2.5.1.17	[c] : atp + cbi + h <==> adocbi + pppi
R_PYNP2r	pyrimidine-nucleosidephosphorylase(uracil)	Pyrimidine Metabolism	2.4.2.2	[c] : pi + uri <==> r1p + ura
R_MMCD	Methylmalonyl-CoAdecarboxylase	Valine, Leucine, and Isoleucine Metabolism	4.1.1.41	[c] : h + mmcoa-s --> co2 + ppcoa
R_bglu3_ext	betaglucosidase	Cellulose Metabolism	3.2.1.21	[e] : cellotriose + h2o --> glc-d + cellobiose
R_DXPRIi	_1-deoxy-D-xylulosereductoisomerase	Sterol Metabolism	1.1.1.267	[c] : dxyl5p + h + nadph --> 2me4p + nadp
R_PHETRS	Phenylalanyl-tRNAsynthetase	Tyrosine, Tryptophan, and Phenylalanine Metabolism	6.1.1.20	[c] : atp + phe-l + trnaphe --> amp + phetrna + ppi
R_PRATPP	phosphoribosyl-ATPpyrophosphatase	Histidine Metabolism	3.6.1.31	[c] : h2o + prbatp --> h + ppi + prbamp
R_PRAIS	phosphoribosylaminoimidazolesynthase	Purine Metabolism	6.3.3.1	[c] : atp + fpram --> adp + air + 2 h + pi
R_IOR2	indolepyruvateferredoxinoxidoreductase(phenylpyvurate	Tyrosine, Tryptophan, and Phenylalanine	1.2.7.8	[c] : coa + 2 fdox + phpyr <==> co2 + 2 fdred + h + phaccoa

www.manaraa.com

	)	Metabolism		
R_GUAPRT	R_guanine_phosphoribosyltransferase	S_Nucleotide_Salvage_Pathway	.	[c] : gua + prpp --> gmp + ppi
R_GSNK	R_guanosine_kinase	S_Nucleotide_Salvage_Pathway	.	[c] : atp + gsn --> adp + gmp + h
R_PUNP2	R_purine_nucleoside_phosphorylase__Deoxyadenosine_	S_Nucleotide_Salvage_Pathway	.	[c] : dad-2 + pi <==> 2dr1p + ade

1. Biomass reactions

R_BIOMASS Total biomass flux including extracellular protein cellmass[c] + 1.3 cellome[e] --> biomass[c]

R_BIOMASScell Biomass flux describing cell mass alone, excluding extracellular protein [c] : 0 glycogen + 0.257980584 ala-l + 0.176597196 arg-l + 0.218152121 asn-l + 0.230278549 asp-l + 0.047389692 cys-l + 0.308171564 glu-l + 0.103857352 gln-l + 0.270219364 gly + 0.057785913 his-l + 0.343142715 ile-l + 0.350014856 leu-l + 0.327413561 lys-l + 0.101531685 met-l + 0.171391596 phe-l + 0.137139492 pro-l + 0.244438532 ser-l + 0.203965923 thr-l + 0.035839999 trp-l + 0.170934701 tyr-l + 0.287641274 val-l + 0.017366666 datp + 0.009566749 dctp + 0.009566749 dgtp + 0.017366666 dttp + 0.008562798 m12dg_ct + 0.011044141 d12dg_ct + 0.006631063 t12dg_ct + 4.58142E-4 cdlp_ct + 0.017672003 pgly_ct + 0.002174587 lysylpgly_ct + 0.05622271 psetha_ct + 6.83646E-4 lipo1_24_ct + 6.12248E-4 lipo2_24_ct + 0.001816095 lipo3_24_ct + 0.001467494 lipo4_24_ct + 0.101806256 peptido_ct + 0.003618287 gtca1_45_ct + 0.00234387 gtca2_45_ct + 0.001816435 gtca3_45_ct + 0.003119816 tcam_ct + 0.7063 k + 0.1017 mg2 + 0.0034 fe3 + 0.0032 ca2 + 9.0E-4 ppi + 0.0162 nad + 0.0047 amp + 0.0010 cmp + 9.0E-4 nadp + 9.0E-4 fad + 5.0E-4 gmp + 3.0E-4 cdp + 2.0E-4 nadph + 2.0E-4 gdp + 0.0415 utp + 150 atp + 0.0394 ctp + 0.0674 gtp + 0 coa --> cellmass + 150 pi + 150 h + 150 adp

المنارة للاستشارات

www.manaraa.com

R_CELLOME Cellulosome biosynthesis and export FALSE Cellulosome . 0.073961 ala-l[c] + 0.031185 arg-l[c] + 0.056501 asn-l[c] + 0.072215 asp-l[c] + 0.008415 cys-l[c] + 0.054069 glu-l[c] + 0.023909 gln-l[c] + 0.082939 gly[c] + 0.013743 his-l[c] + 0.060879 ile-l[c] + 0.065801 leu-l[c] + 0.056547 lys-l[c] + 0.019202 met-l[c] + 0.036677 phe-l[c] + 0.062856 pro-l[c] + 0.061534 ser-l[c] + 0.073364 thr-l[c] + 0.021761 trp-l[c] + 0.054915 tyr-l[c] + 0.069528 val-l[c] + atp[c] --> cellome[e] + adp[c] + pi[c]

B. Metabolite list

metabolite abbrev	metabolite name	compartment
10fthf[c]	10-Formyltetrahydrofolate	Cytosol
12dag3p_ct[c]	1_2_diacyl_sn_glycerol_3_phosphate	Cytosol
12dgr_ct[c]	1,2-Diacylglycerol	Cytosol
13dpg[c]	3-Phospho-D-glyceroyl phosphate	Cytosol
14glun[c]	(1,4-alpha-D-Glucosyl)n	Cytosol
1ag3p_ct[c]	1-Acyl-sn-glycerol 3-phosphate	Cytosol
1pyr5c[c]	1-Pyrroline-5-carboxylate	Cytosol
23dhdp[c]	2,3-Dihydrodipicolinate	Cytosol
23dhmb[c]	(R)-2,3-Dihydroxy-3-methylbutanoate	Cytosol
23dhmp[c]	(R)-2,3-Dihydroxy-3-methylpentanoate	Cytosol
25aics[c]	(S)-2-[5-Amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamido]succinate	Cytosol
25dhpp[c]	2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine	Cytosol
26dap-ll[c]	LL-2,6-Diaminoheptanedioate	Cytosol
26dap-m[c]	meso-2,6-Diaminoheptanedioate	Cytosol
26dap-m[e]	meso-2,6-Diaminoheptanedioate	Extraorganism
2ahbut[c]	(S)-2-Aceto-2-hydroxybutanoate	Cytosol
2ahhmd[c]	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine diphosphate	Cytosol
2ahhmp[c]	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	Cytosol
2aobut[c]	L-2-Amino-3-oxobutanoate	Cytosol
2cpr5p[c]	1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate	Cytosol
2dda7p[c]	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate	Cytosol
2dhp[c]	2-Dehydropantoate	Cytosol
2dr1p[c]	2-Deoxy-D-ribose 1-phosphate	Cytosol
2dr5p[c]	2-Deoxy-D-ribose 5-phosphate	Cytosol
2hb[c]	2-Hydroxybutyrate	Cytosol
2ippm[c]	2-Isopropylmaleate	Cytosol
2mahmp[c]	2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate	Cytosol
2mbcoa[c]	2-Methylbutanoyl-CoA	Cytosol
2mcit[c]	2-Methylcitrate	Cytosol
2me4p[c]	2-C-methyl-D-erythritol 4-phosphate	Cytosol
2mecdp[c]	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	Cytosol
2obut[c]	2-Oxobutanoate	Cytosol
2p4c2me[c]	2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol	Cytosol
2pg[c]	D-Glycerate 2-phosphate	Cytosol
2pglyc[c]	2-Phosphoglycolate	Cytosol
34hpp[c]	3-(4-Hydroxyphenyl)pyruvate	Cytosol

المنارة للاستشارات
www.manaraa.com

3c2hmp[c]	3-Carboxy-2-hydroxy-4-methylpentanoate	Cytosol
3c3hmp[c]	3-Carboxy-3-hydroxy-4-methylpentanoate	Cytosol
3c4mop[c]	3-Carboxy-4-methyl-2-oxopentanoate	Cytosol
3dhq[c]	3-Dehydroquinate	Cytosol
3dhsk[c]	3-Dehydroshikimate	Cytosol
3ig3p[c]	C'-(3-Indolyl)-glycerol 3-phosphate	Cytosol
3mob[c]	3-Methyl-2-oxobutanoate	Cytosol
3mop[c]	(S)-3-Methyl-2-oxopentanoate	Cytosol
3pg[c]	3-Phospho-D-glycerate	Cytosol
3php[c]	3-Phosphohydroxypyruvate	Cytosol
3psme[c]	5-O-(1-Carboxyvinyl)-3-phosphoshikimate	Cytosol
3spyr[c]	3-Sulfopyruvate	Cytosol
4abz[c]	4-Aminobenzoate	Cytosol
4adcho[c]	4-amino-4-deoxychorismate	Cytosol
4c2me[c]	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol	Cytosol
4h2oglt[c]	4-Hydroxy-2-oxoglutarate	Cytosol
4hphac[c]	4-Hydroxyphenylacetate	Cytosol
4mop[c]	4-Methyl-2-oxopentanoate	Cytosol
4mpetz[c]	4-Methyl-5-(2-phosphoethyl)-thiazole	Cytosol
4pasp[c]	4-Phospho-L-aspartate	Cytosol
4ppan[c]	D-4'-Phosphopantothenate	Cytosol
4ppcys[c]	N-((R)-4-Phosphopantothenoyl)-L-cysteine	Cytosol
4r5au[c]	4-(1-D-Ribitylamino)-5-aminouracil	Cytosol
5aizc[c]	5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxylate	Cytosol
5aop[c]	5-Amino-4-oxopentanoate	Cytosol
5aprbu[c]	5-Amino-6-(5'-phosphoribitylamino)uracil	Cytosol
5apru[c]	5-Amino-6-(5'-phosphoribosylamino)uracil	Cytosol
5fthf[c]	5-Formyltetrahydrofolate	Cytosol
5mdr1p[c]	5-Methylthio-5-deoxy-D-ribose 1-phosphate	Cytosol
5mdru1p[c]	5-Methylthio-5-deoxy-D-ribulose 1-phosphate	Cytosol
5mta[c]	5-Methylthioadenosine	Cytosol
5mthf[c]	5-Methyltetrahydrofolate	Cytosol
5prdmbz[c]	N1-(5-Phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole	Cytosol
6pthp[c]	6-Pyruvoyl-5,6,7,8-tetrahydropterin	Cytosol
8aonn[c]	8-Amino-7-oxononanoate	Cytosol
ac[c]	Acetate	Cytosol
ac[e]	Acetate	Extraorganism
acald[c]	Acetaldehyde	Cytosol
accoa[c]	Acetyl-CoA	Cytosol
acg5p[c]	N-Acetyl-L-glutamyl 5-phosphate	Cytosol
acg5sa[c]	N-Acetyl-L-glutamate 5-semialdehyde	Cytosol
acgam[c]	N-Acetyl-D-glucosamine	Cytosol
acgam1p[c]	N-Acetyl-D-glucosamine 1-phosphate	Cytosol
acgam6p[c]	N-Acetyl-D-glucosamine 6-phosphate	Cytosol
acglu[c]	N-Acetyl-L-glutamate	Cytosol
achms[c]	O-Acetyl-L-homoserine	Cytosol
acmam[c]	N-Acetyl-D-muramoate	Cytosol
acmama[c]	N-Acetyl-D-muramoyl-L-alanine	Cytosol
acmana[c]	N-Acetyl-D-mannosamine	Cytosol
acmanap[c]	N-Acetyl-D-mannosamine 6-phosphate	Cytosol
acnam[c]	N-Acetylneuraminate	Cytosol
acnamp[c]	N-Acetylneuraminate 9-phosphate	Cytosol
acorn[c]	N2-Acetyl-L-ornithine	Cytosol
acp[c]	acyl carrier protein	Cytosol
acser[c]	O-Acetyl-L-serine	Cytosol
actp[c]	Acetyl phosphate	Cytosol
ad[c]	acetamide	Cytosol
ade[c]	Adenine	Cytosol
adn[c]	Adenosine	Cytosol

المنارة للاستشارات

www.manaraa.com

adocbi[c]	Adenosyl cobinamide	Cytosol
adocbip[c]	Adenosyl cobinamide phosphate	Cytosol
adocbl[c]	Adenosylcobalamin	Cytosol
adp[c]	ADP	Cytosol
adpglc[c]	ADPglucose	Cytosol
adprib[c]	ADPribose	Cytosol
agdpcbi[c]	Adenosine-GDP-cobinamide	Cytosol
agm[c]	Agmatine	Cytosol
ahcys[c]	S-Adenosyl-L-homocysteine	Cytosol
ahdt[c]	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl)dihydropteridine triphosphate	Cytosol
aicar[c]	5-Amino-1-(5-Phospho-D-ribosyl)imidazole-4-carboxamide	Cytosol
air[c]	5-amino-1-(5-phospho-D-ribosyl)imidazole	Cytosol
akg[c]	2-Oxoglutarate	Cytosol
ala-b[c]	beta-Alanine	Cytosol
ala-d[c]	D-Alanine	Cytosol
ala-l[c]	L-Alanine	Cytosol
alaala[c]	D-Alanyl-D-alanine	Cytosol
alac-s[c]	(S)-2-Acetolactate	Cytosol
alatrna[c]	L-Alanyl-tRNA(Ala)	Cytosol
amet[c]	S-Adenosyl-L-methionine	Cytosol
ametam[c]	S-Adenosylmethioninamine	Cytosol
amob[c]	S-Adenosyl-4-methylthio-2-oxobutanoate	Cytosol
amp[c]	AMP	Cytosol
anth[c]	Anthranilate	Cytosol
aps[c]	Adenosine 5'-phosphosulfate	Cytosol
arab-l[c]	L-Arabinose	Cytosol
arab-l[e]	L-Arabinose	Extraorganism
arg-l[c]	L-Arginine	Cytosol
argsuc[c]	N(omega)-(L-Arginino)succinate	Cytosol
argtrna[c]	L-Arginyl-tRNA(Arg)	Cytosol
asn-l[c]	L-Asparagine	Cytosol
asntrna[c]	L-Asparaginyl-tRNA(Asn)	Cytosol
asp-l[c]	L-Aspartate	Cytosol
aspsa[c]	L-Aspartate 4-semialdehyde	Cytosol
asptrna[c]	L-Aspartyl-tRNA(Asp)	Cytosol
atp[c]	ATP	Cytosol
biomass[c]	Biomass	Cytosol
btamp[c]	Biotinyl-5'-AMP	Cytosol
btn[c]	Biotin	Cytosol
ca2[c]	Calcium	Cytosol
ca2[e]	Calcium	Extraorganism
camp[c]	cAMP	Cytosol
carn[c]	L-Carnosine	Cytosol
cbasp[c]	N-Carbamoyl-L-aspartate	Cytosol
cbi[c]	Cobinamide	Cytosol
cbp[c]	Carbamoyl phosphate	Cytosol
cdlp_ct[c]	Cardiolipin	Cytosol
cdp[c]	CDP	Cytosol
cdpdag_ct[c]	CDPdiacylglycerol	Cytosol
cdpddglu[c]	CDP-4-dehydro-6-deoxy-D-glucose	Cytosol
cdpglu[c]	CDP-glucose	Cytosol
cdpglyc[c]	CDPglycerol	Cytosol
cellmass[c]	Cellmass	Cytosol
cellobiose[c]	Cellobiose	Cytosol
cellobiose[e]	Cellobiose	Extraorganism
cellohexaose[c]	Cellohexaose	Cytosol
cellohexaose[e]	Cellohexaose	Extraorganism
cellome[e]	Cellulosome	Extraorganism

المنارة للاستشارات

www.manaraa.com

cellopentaose[c]	Cellopentaose	Cytosol
cellopentaose[e]	Cellopentaose	Extraorganism
cellotetraose[c]	Cellotetraose	Cytosol
cellotetraose[e]	Cellotetraose	Extraorganism
cellotriose[c]	Cellotriose	Cytosol
cellotriose[e]	Cellotriose	Extraorganism
cellulose[e]	Cellulose	Extraorganism
ch4s[c]	Methanethiol	Cytosol
chor[c]	chorismate	Cytosol
chtn[c]	chitin	Cytosol
cit[c]	Citrate	Cytosol
cit[e]	Citrate	Extraorganism
citr-l[c]	L-Citrulline	Cytosol
cmp[c]	CMP	Cytosol
cmpacna[c]	CMP-N-acetylneuraminate	Cytosol
co[c]	Carbon monoxide	Cytosol
co2[c]	CO2	Cytosol
co2[e]	CO2	Extraorganism
coa[c]	Coenzyme A	Cytosol
csn[c]	Cytosine	Cytosol
ctobiose[c]	Chitobiose	Cytosol
ctp[c]	CTP	Cytosol
cu2[c]	Cu2+	Cytosol
cu2[e]	Cu2+	Extraorganism
cys-l[c]	L-Cysteine	Cytosol
cyst-l[c]	L-Cystathionine	Cytosol
cystrna[c]	L-Cysteinyl-tRNA(Cys)	Cytosol
cytd[c]	Cytidine	Cytosol
d12dg_ct[c]	diglucosyl_1_2_diacylglycerol	Cytosol
dadp[c]	dADP	Cytosol
damp[c]	dAMP	Cytosol
dann[c]	7,8-Diaminononanoate	Cytosol
datp[c]	dATP	Cytosol
db4p[c]	3,4-dihydroxy-2-butanone 4-phosphate	Cytosol
dca[c]	Decanoate (n-C10:0)	Cytosol
dcacoa[c]	Decanoyl-CoA (n-C10:0CoA)	Cytosol
dcamp[c]	N6-(1,2-Dicarboxyethyl)-AMP	Cytosol
dcdp[c]	dCDP	Cytosol
dcmp[c]	dCMP	Cytosol
dctp[c]	dCTP	Cytosol
decdp[c]	all trans Decaprenyl diphosphate	Cytosol
dgdp[c]	dGDP	Cytosol
dgmp[c]	dGMP	Cytosol
dgsn[c]	Deoxyguanosine	Cytosol
dgtp[c]	dGTP	Cytosol
dhap[c]	Dihydroxyacetone phosphate	Cytosol
dhf[c]	7,8-Dihydrofolate	Cytosol
dhnpt[c]	Dihydroneopterin	Cytosol
dhor-s[c]	(S)-Dihydroorotate	Cytosol
dhpt[c]	Dihydropteroate	Cytosol
din[c]	Deoxyinosine	Cytosol
dmbzid[c]	5,6-Dimethylbenzimidazole	Cytosol
dmlz[c]	6,7-Dimethyl-8-(1-D-ribityl)lumazine	Cytosol
dmpp[c]	Dimethylallyl diphosphate	Cytosol
dnad[c]	Deamino-NAD+	Cytosol
dpcoa[c]	Dephospho-CoA	Cytosol
dscl[c]	dihydrosirohydrochlorin	Cytosol
dtbt[c]	Dethiobiotin	Cytosol
dtdp[c]	dTDP	Cytosol

www.manaraa.com

dtdp4d6dg[c]	dTDP-4-dehydro-6-deoxy-D-glucose	Cytosol
dtdp4d6dm[c]	dTDP-4-dehydro-6-deoxy-L-mannose	Cytosol
dtdpglu[c]	dTDPglucose	Cytosol
dtmp[c]	dTMP	Cytosol
dttp[c]	dTTP	Cytosol
dudp[c]	dUDP	Cytosol
dump[c]	dUMP	Cytosol
dutp[c]	dUTP	Cytosol
dxyl5p[c]	1-deoxy-D-xylulose 5-phosphate	Cytosol
e4hglu[c]	L-erythro-4-Hydroxyglutamate	Cytosol
e4p[c]	D-Erythrose 4-phosphate	Cytosol
eig3p[c]	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate	Cytosol
etoh[c]	Ethanol	Cytosol
etoh[e]	Ethanol	Extraorganism
f1p[c]	D-Fructose 1-phosphate	Cytosol
f6p[c]	D-Fructose 6-phosphate	Cytosol
fa1[c]	Fatty acid (Iso-C14:0)	Cytosol
fa10[c]	Fatty acid (Anteiso-C17:1)	Cytosol
fa10coa[c]	Anteiso_C171_CoA	Cytosol
fa11[c]	Fatty acid (Iso-C17:0)	Cytosol
fa11coa[c]	Iso_C170_CoA	Cytosol
fa12[c]	Fatty acid (Anteiso-C17:0)	Cytosol
fa12coa[c]	Anteiso_C170_CoA	Cytosol
fa1coa[c]	Iso_C140_CoA	Cytosol
fa3[c]	Fatty acid (Iso-C15:0)	Cytosol
fa3coa[c]	Iso_C150_CoA	Cytosol
fa4[c]	Fatty acid (Anteiso-C15:0)	Cytosol
fa4coa[c]	Anteiso_C150_CoA	Cytosol
fa5[c]	Fatty acid (Iso-C16:1)	Cytosol
fa5coa[c]	Iso_C161_CoA	Cytosol
fa6[c]	Fatty acid (iso-C16:0)	Cytosol
fa6coa[c]	Iso_C160_CoA	Cytosol
fa9[c]	Fatty acid (Iso-C17:1)	Cytosol
fa9coa[c]	Iso_C171_CoA	Cytosol
fad[c]	Flavin adenine dinucleotide oxidized	Cytosol
fdox[c]	ferredoxin (oxidized) 2[4Fe-4S]	Cytosol
fdp[c]	D-Fructose 1,6-bisphosphate	Cytosol
fdred[c]	ferredoxin (reduced) 2[4Fe-4S]	Cytosol
fe3[c]	Fe3+	Cytosol
fe3[e]	Fe3+	Extraorganism
fgam[c]	N2-Formyl-N1-(5-phospho-D-ribosyl)glycinamide	Cytosol
fmettrna[c]	N-Formylmethionyl-tRNA	Cytosol
fmn[c]	FMN	Cytosol
for[c]	Formate	Cytosol
for[e]	Formate	Extraorganism
fpram[c]	2-(Formamido)-N1-(5-phospho-D-ribosyl)acetamidine	Cytosol
fprica[c]	5-Formamido-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide	Cytosol
frdp[c]	Farnesyl diphosphate	Cytosol
fru[c]	D-Fructose	Cytosol
fru[e]	D-Fructose	Extraorganism
fum[c]	Fumarate	Cytosol
g1p[c]	D-Glucose 1-phosphate	Cytosol
g3p[c]	Glyceraldehyde 3-phosphate	Cytosol
g6p[c]	D-Glucose 6-phosphate	Cytosol
gal[c]	D-Galactose	Cytosol
gal[e]	D-Galactose	Extraorganism
gam1p[c]	D-Glucosamine 1-phosphate	Cytosol
gam6p[c]	D-Glucosamine 6-phosphate	Cytosol
gar[c]	N1-(5-Phospho-D-ribosyl)glycinamide	Cytosol

المنارة للاستشارات

www.manaraa.com

gcald[c]	Glycolaldehyde	Cytosol
gcald[e]	Glycolaldehyde	Extraorganism
gdp[c]	GDP	Cytosol
gdp[e]	GDP	Extraorganism
gdptp[c]	Guanosine 3'-diphosphate 5'-triphosphate	Cytosol
ggdp[c]	Geranylgeranyl diphosphate	Cytosol
glc-d[c]	D-Glucose	Cytosol
glc-d[e]	D-Glucose	Extraorganism
gln-l[c]	L-Glutamine	Cytosol
glntrna[c]	L-Glutaminyl-tRNA(Gln)	Cytosol
glu-d[c]	D-Glutamate	Cytosol
glu-l[c]	L-Glutamate	Cytosol
glu1sa[c]	L-Glutamate 1-semialdehyde	Cytosol
glu5p[c]	L-Glutamate 5-phosphate	Cytosol
glu5sa[c]	L-Glutamate 5-semialdehyde	Cytosol
glutrna[c]	L-Glutamyl-tRNA(Glu)	Cytosol
glx[c]	Glyoxylate	Cytosol
gly[c]	Glycine	Cytosol
glyc-r[c]	(R)-Glycerate	Cytosol
glyc[c]	Glycerol	Cytosol
glyc3p[c]	Glycerol 3-phosphate	Cytosol
glyclt[c]	Glycolate	Cytosol
glycogen[c]	glycogen	Cytosol
glytrna[c]	Glycyl-tRNA(Gly)	Cytosol
gmp[c]	GMP	Cytosol
grdp[c]	Geranyl diphosphate	Cytosol
gsn[c]	Guanosine	Cytosol
gtca1_45_ct[c]	glycerol_teichoic_acid__n45___unlinked__unsubstituted	Cytosol
gtca2_45_ct[c]	glycerol_teichoic_acid__n45___unlinked__D_ala_substituted	Cytosol
gtca3_45_ct[c]	glycerol_teichoic_acid__n45___unlinked__glucose_substituted	Cytosol
gtp[c]	GTP	Cytosol
gtp[e]	GTP	Extraorganism
gua[c]	Guanine	Cytosol
h[c]	H+	Cytosol
h[e]	H+	Extraorganism
h2[c]	H2	Cytosol
h2[e]	H2	Extraorganism
h2co3[c]	carbonic acid	Cytosol
h2mb4p[c]	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate	Cytosol
h2o[c]	H2O	Cytosol
h2o[e]	H2O	Extraorganism
h2o2[c]	Hydrogen peroxide	Cytosol
h2s[c]	Hydrogen sulfide	Cytosol
h2s[e]	Hydrogen sulfide	Extraorganism
h4spt[c]	tetrahydrosarcinapterin	Cytosol
hco3[c]	Bicarbonate	Cytosol
hcys-l[c]	L-Homocysteine	Cytosol
hdca[c]	Hexadecanoate (n-C16:0)	Cytosol
hdcea[c]	Hexadecenoate (n-C16:1)	Cytosol
hdcoa[c]	Hexadecenoyl-CoA (n-C16:1CoA)	Cytosol
hepdp[c]	all-trans-Heptaprenyl diphosphate	Cytosol
hexdp[c]	all-trans-Hexaprenyl diphosphate	Cytosol
his-l[c]	L-Histidine	Cytosol
hisp[c]	L-Histidinol phosphate	Cytosol
hista[c]	Histamine	Cytosol
histd[c]	L-Histidinol	Cytosol
histrna[c]	L-Histidyl-tRNA(His)	Cytosol
hmbil[c]	Hydroxymethylbilane	Cytosol
hom-l[c]	L-Homoserine	Cytosol

www.manaraa.com

hphaccoa[c]	4-Hydroxyphenylacetyl-CoA	Cytosol
hpyr[c]	Hydroxypyruvate	Cytosol
hxan[c]	Hypoxanthine	Cytosol
iad[c]	Indole-3-acetamide	Cytosol
iasp[c]	Iminoaspartate	Cytosol
ibcoa[c]	Isobutyryl-CoA	Cytosol
icit[c]	Isocitrate	Cytosol
idp[c]	IDP	Cytosol
ile-l[c]	L-Isoleucine	Cytosol
ile-l[e]	L-Isoleucine	Extraorganism
iletrna[c]	L-Isoleucyl-tRNA(Ile)	Cytosol
imacp[c]	3-(Imidazol-4-yl)-2-oxopropyl phosphate	Cytosol
imp[c]	IMP	Cytosol
ind3ac[c]	Indole-3-acetate	Cytosol
indaccoa[c]	S-2-(indol-3-yl)acetyl-CoA	Cytosol
indole[c]	Indole	Cytosol
indpyr[c]	Indolepyruvate	Cytosol
ins[c]	Inosine	Cytosol
ipdp[c]	Isopentenyl diphosphate	Cytosol
itp[c]	ITP	Cytosol
ivcoa[c]	Isovaleryl-CoA	Cytosol
k[c]	potassium	Cytosol
k[e]	potassium	Extraorganism
lac-l[c]	L-Lactate	Cytosol
lac-l[e]	L-Lactate	Extraorganism
lcts[c]	Lactose	Cytosol
leu-l[c]	L-Leucine	Cytosol
leu-l[e]	L-Leucine	Extraorganism
leutrna[c]	L-Leucyl-tRNA(Leu)	Cytosol
lipo1_24_ct[c]	lipoteichoic_acid__n24___linked__glucose_substituted	Cytosol
lipo2_24_ct[c]	lipoteichoic_acid__n24___linked__N_acetyl_D_glucosamine	Cytosol
lipo3_24_ct[c]	lipoteichoic_acid__n24___linked__D_alanine_substituted	Cytosol
lipo4_24_ct[c]	lipoteichoic_acid__n24___linked__unsubstituted	Cytosol
lys-l[c]	L-Lysine	Cytosol
lystrna[c]	L-Lysine-tRNA (Lys)	Cytosol
lysylpgly_ct[c]	lysylphophatidylglycerol	Cytosol
m12dg_ct[c]	monoglucosyl_1_2_diacylglycerol	Cytosol
mal-l[c]	L-Malate	Cytosol
mal-l[e]	L-Malate	Extraorganism
malacp[c]	Malonyl-[acyl-carrier protein]	Cytosol
malcoa[c]	Malonyl-CoA	Cytosol
man6p[c]	D-Mannose 6-phosphate	Cytosol
met-l[c]	L-Methionine	Cytosol
met-l[e]	L-Methionine	Extraorganism
methf[c]	5,10-Methenyltetrahydrofolate	Cytosol
mettrna[c]	L-Methionyl-tRNA (Met)	Cytosol
mg2[c]	magnesium	Cytosol
mg2[e]	magnesium	Extraorganism
mh4spt[c]	N5-methyl-tetrahydrosarcinapterin	Cytosol
mlthf[c]	5,10-Methylenetetrahydrofolate	Cytosol
mmcoa-r[c]	(R)-Methylmalonyl-CoA	Cytosol
mmcoa-s[c]	(S)-Methylmalonyl-CoA	Cytosol
mthgxl[c]	Methylglyoxal	Cytosol
n2[c]	nitrogen	Cytosol
na1[c]	Sodium	Cytosol
na1[e]	Sodium	Extraorganism
nac[c]	Nicotinate	Cytosol
nad[c]	Nicotinamide adenine dinucleotide	Cytosol
nadh[c]	Nicotinamide adenine dinucleotide - reduced	Cytosol

المنارة للاستشارات

www.manaraa.com

nadp[c]	Nicotinamide adenine dinucleotide phosphate	Cytosol
nadph[c]	Nicotinamide adenine dinucleotide phosphate - reduced	Cytosol
ncam[c]	Nicotinamide	Cytosol
nh4[c]	Ammonium	Cytosol
nh4[e]	Ammonium	Extraorganism
nicrnt[c]	Nicotinate D-ribonucleotide	Cytosol
nmn[c]	NMN	Cytosol
nmn[e]	NMN	Extraorganism
no2[c]	Nitrite	Cytosol
no2[e]	Nitrite	Extraorganism
no3[c]	Nitrate	Cytosol
no3[e]	Nitrate	Extraorganism
o2[c]	O2	Cytosol
o2s[c]	Superoxide anion	Cytosol
oaa[c]	Oxaloacetate	Cytosol
ocdca[c]	octadecanoate (n-C18:0)	Cytosol
ocdcaiso[c]	Iso_C180	Cytosol
ocdcea[c]	octadecenoate (n-C18:1)	Cytosol
odecoa[c]	Octadecenoyl-CoA (n-C18:1CoA)	Cytosol
ohpb[c]	2-Oxo-3-hydroxy-4-phosphobutanoate	Cytosol
orn[c]	Ornithine	Cytosol
orot[c]	Orotate	Cytosol
orot5p[c]	Orotidine 5'-phosphate	Cytosol
pac[c]	Phenylacetic acid	Cytosol
pad[c]	2-Phenylacetamide	Cytosol
pan4p[c]	Pantetheine 4'-phosphate	Cytosol
pant-r[c]	(R)-Pantoate	Cytosol
pap[c]	Adenosine 3',5'-bisphosphate	Cytosol
paps[c]	3'-Phosphoadenylyl sulfate	Cytosol
pdx5p[c]	Pyridoxine 5'-phosphate	Cytosol
pendp[c]	all_trans_Pentaprenyl_diphosphate	Cytosol
pep[c]	Phosphoenolpyruvate	Cytosol
peptido_ct[c]	Peptidoglycan subunit	Cytosol
pgly_ct[c]	phosphatidylglycerol	Cytosol
pglyp_ct[c]	Phosphatidylglycerophosphate	Cytosol
phaccoa[c]	Phenylacetyl-CoA	Cytosol
phe-l[c]	L-Phenylalanine	Cytosol
phetrna[c]	L-Phenylalanyl-tRNA(Phe)	Cytosol
phom[c]	O-Phospho-L-homoserine	Cytosol
phpyr[c]	Phenylpyruvate	Cytosol
phthr[c]	O-Phospho-4-hydroxy-L-threonine	Cytosol
pi[c]	Phosphate	Cytosol
pi[e]	Phosphate	Extraorganism
pmcoa[c]	Pimeloyl-CoA	Cytosol
pmtcoa[c]	Palmitoyl-CoA (n-C16:0CoA)	Cytosol
pnto-r[c]	(R)-Pantothenate	Cytosol
pnto-r[e]	(R)-Pantothenate	Extraorganism
ppa[c]	Propionate (n-C3:0)	Cytosol
ppap[c]	Propanoyl phosphate	Cytosol
ppbng[c]	Porphobilinogen	Cytosol
ppcoa[c]	Propanoyl-CoA	Cytosol
pphn[c]	Prephenate	Cytosol
ppi[c]	Diphosphate	Cytosol
pppi[c]	Inorganic triphosphate	Cytosol
pram[c]	5-Phospho-beta-D-ribosylamine	Cytosol
pran[c]	N-(5-Phospho-D-ribosyl)anthranilate	Cytosol
prbamp[c]	1-(5-Phosphoribosyl)-AMP	Cytosol
prbatp[c]	1-(5-Phosphoribosyl)-ATP	Cytosol
prfp[c]	1-(5-Phosphoribosyl)-5-[(5-	Cytosol

المنارة للاستشارات

www.manaraa.com

	phosphoribosylamino)methylideneamino]imidazole-4-carboxamide	
prlp[c]	5-[(5-phospho-1-deoxyribulos-1-ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-carboxamide	Cytosol
pro-l[c]	L-Proline	Cytosol
pro-l[e]	L-Proline	Extraorganism
protrna[c]	L-Prolyl-tRNA(Pro)	Cytosol
prpp[c]	5-Phospho-alpha-D-ribose 1-diphosphate	Cytosol
ps_ct[c]	phosphatidylserine	Cytosol
psd5p[c]	Pseudouridine 5'-phosphate	Cytosol
pser-l[c]	O-Phospho-L-serine	Cytosol
psetha_ct[c]	phosphatidylethanolamine	Cytosol
ptrc[c]	Putrescine	Cytosol
ptrc[e]	Putrescine	Extraorganism
pyr[c]	Pyruvate	Cytosol
q8[c]	Ubiquinone-8	Cytosol
q8h2[c]	Ubiquinol-8	Cytosol
quln[c]	Quinolinate	Cytosol
r1p[c]	alpha-D-Ribose 1-phosphate	Cytosol
r5p[c]	alpha-D-Ribose 5-phosphate	Cytosol
rdmbzi[c]	N1-(alpha-D-ribosyl)-5,6-dimethylbenzimidazole	Cytosol
rib-d[c]	D-Ribose	Cytosol
rib-d[e]	D-Ribose	Extraorganism
ribflv[c]	Riboflavin	Cytosol
rnam[c]	N-Ribosylnicotinamide	Cytosol
ru5p-d[c]	D-Ribulose 5-phosphate	Cytosol
s[c]	Sulfur	Cytosol
s7p[c]	Sedoheptulose 7-phosphate	Cytosol
sbt-d[c]	D-Sorbitol	Cytosol
ser-l[c]	L-Serine	Cytosol
sertrna[c]	L-Seryl-tRNA(Ser)	Cytosol
skm[c]	Shikimate	Cytosol
skm5p[c]	Shikimate 5-phosphate	Cytosol
sl-l[c]	L-sulfolactate	Cytosol
so3[c]	Sulfite	Cytosol
so4[c]	Sulfate	Cytosol
so4[e]	Sulfate	Extraorganism
spmd[c]	Spermidine	Cytosol
spmd[e]	Spermidine	Extraorganism
stcoa[c]	Stearoyl-CoA (n-C18:0CoA)	Cytosol
stcoaiso[c]	Iso_C180_CoA	Cytosol
succ[c]	Succinate	Cytosol
succ[e]	Succinate	Extraorganism
succoa[c]	Succinyl-CoA	Cytosol
suchms[c]	O-Succinyl-L-homoserine	Cytosol
t12dg_ct[c]	triglucosyl_1_2_diacylglycerol	Cytosol
tag6p-d[c]	D-Tagatose 6-phosphate	Cytosol
tagdp-d[c]	D-Tagatose 1,6-biphosphate	Cytosol
tcam_ct[c]	minor_teichoic_acid__acetylgalactosamine_glucose_phosphate__n30	Cytosol
tdcoa[c]	Tetradecanoyl-CoA (n-C14:0CoA)	Cytosol
tdecoa[c]	Tetradecenoyl-CoA (n-C14:1CoA)	Cytosol
teich_45_ct[c]	teichuronic_acid__GlcA__GalNac__45_repeating_unit	Cytosol
thdp[c]	2,3,4,5-Tetrahydrodipicolinate	Cytosol
thf[c]	5,6,7,8-Tetrahydrofolate	Cytosol
thm[c]	Thiamin	Cytosol
thmmp[c]	Thiamin monophosphate	Cytosol
thmpp[c]	Thiamine diphosphate	Cytosol
thr-l[c]	L-Threonine	Cytosol
thrtrna[c]	L-Threonyl-tRNA(Thr)	Cytosol
thym[c]	Thymine	Cytosol

www.manaraa.com

thymd[c]	Thymidine	Cytosol
trdacoa[c]	Iso_C130_CoA	Cytosol
trdca[c]	Tridecanoate	Cytosol
trdox[c]	Oxidized thioredoxin	Cytosol
trdrd[c]	Reduced thioredoxin	Cytosol
trnaala[c]	tRNA(Ala)	Cytosol
trnaarg[c]	tRNA(Arg)	Cytosol
trnaasn[c]	tRNA(Asn)	Cytosol
trnaasp[c]	tRNA(Asp)	Cytosol
trnacys[c]	tRNA(Cys)	Cytosol
trnagln[c]	tRNA(Gln)	Cytosol
trnaglu[c]	tRNA (Glu)	Cytosol
trnagly[c]	tRNA(Gly)	Cytosol
trnahis[c]	tRNA(His)	Cytosol
trnaile[c]	tRNA(Ile)	Cytosol
trnaleu[c]	tRNA(Leu)	Cytosol
trnalys[c]	tRNA(Lys)	Cytosol
trnamet[c]	tRNA(Met)	Cytosol
trnaphe[c]	tRNA(Phe)	Cytosol
trnapro[c]	tRNA(Pro)	Cytosol
trnaser[c]	tRNA(Ser)	Cytosol
trnathr[c]	tRNA(Thr)	Cytosol
trnatrp[c]	tRNA(Trp)	Cytosol
trnatyr[c]	tRNA(Tyr)	Cytosol
trnaval[c]	tRNA(Val)	Cytosol
trp-l[c]	L-Tryptophan	Cytosol
trptrna[c]	L-Tryptophanyl-tRNA(Trp)	Cytosol
ttdca[c]	tetradecanoate (n-C14:0)	Cytosol
ttdcea[c]	tetradecenoate (n-C14:1)	Cytosol
tyr-l[c]	L-Tyrosine	Cytosol
tyrtrna[c]	L-Tyrosyl-tRNA(Tyr)	Cytosol
uaagmda[c]	Undecaprenyl-diphospho-N-acetylmuramoyl-(N-acetylglucosamine)-L-ala-D-glu-meso-2,6-diaminopimeloyl-D-ala-D-ala	Cytosol
uaccg[c]	UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine	Cytosol
uacgam[c]	UDP-N-acetyl-D-glucosamine	Cytosol
uacmam[c]	UDP-N-acetyl-D-mannosamine	Cytosol
uagmda[c]	Undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	Cytosol
uama[c]	UDP-N-acetylmuramoyl-L-alanine	Cytosol
uamag[c]	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate	Cytosol
uamr[c]	UDP-N-acetylmuramate	Cytosol
udcp[c]	Undecaprenol	Cytosol
udcpdp[c]	Undecaprenyl diphosphate	Cytosol
udcpp[c]	Undecaprenyl phosphate	Cytosol
udp[c]	UDP	Cytosol
udpacgal[c]	UDP-N-acetyl-D-galactosamine	Cytosol
udpg[c]	UDPglucose	Cytosol
udpgal[c]	UDPgalactose	Cytosol
udpglcur[c]	UDP-D-glucuronate	Cytosol
ugmd[c]	UDP-N-acetylmuramoyl-L-alanyl-D-gamma-glutamyl-meso-2,6-diaminopimelate	Cytosol
ugmda[c]	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine	Cytosol
ump[c]	UMP	Cytosol
uppg3[c]	Uroporphyrinogen III	Cytosol
ura[c]	Uracil	Cytosol
ura[e]	Uracil	Extraorganism
urea[c]	Urea	Cytosol
urea[e]	Urea	Extraorganism

www.manaraa.com

uri[c]	Uridine	Cytosol
utp[c]	UTP	Cytosol
val-l[c]	L-Valine	Cytosol
val-l[e]	L-Valine	Extraorganism
valtrna[c]	L-Valyl-tRNA(Val)	Cytosol
xan[c]	Xanthine	Cytosol
xan[e]	Xanthine	Extraorganism
xmp[c]	Xanthosine 5'-phosphate	Cytosol
xtsn[c]	Xanthosine	Cytosol
xu5p-d[c]	D-Xylulose 5-phosphate	Cytosol
xyl-d[c]	D-Xylose	Cytosol
xyl-d[e]	D-Xylose	Extraorganism

C. GPR list

R_OIVD2	.
R_OIVD3	.
R_OIVD1	.
R_ALDD2x	Cthe_2238
R_MME	Cthe_2225
R_PRPPS	Cthe_2630
R_TECA2S45	.
R_IPMD	Cthe_2209
R_PFK	Cthe_1261 or Cthe_0347
R_PANTS	Cthe_0901
R_ALATRS	Cthe_3200
R_H2CO3D	.
R_GLYCK	Cthe_1592
R_ADCL	.
R_CTPS1	Cthe_1923
R_NH4t	.
R_FACOAL150_ISO_	Cthe_1232
R_SPMS	Cthe_0694
R_PTHPS	Cthe_3106
R_SERt2r	.
R_TRPt2r	.
R_HPPK	Cthe_2583
R_bglu2	Cthe_0212 or Cthe_1256
R_G1PCTR	Cthe_2562
R_SERTRS	Cthe_2381
R_NNATr	Cthe_1241
R_DHORD10	Cthe_0948 and Cthe_0947
R_LIPO4S24_CT	.
R_ASNTRS	Cthe_0070
R_ACS2	Cthe_0551
R_CA2abc	.
R_DGK1	Cthe_1315
R_cellobtx	Cthe_2446 and Cthe_0393
R_PRAGSr	Cthe_1245
R_CHTBNASE	.
R_PPM2	Cthe_0677
R_ACGAMK	.
R_ARBabc	(Cthe_1579 and Cthe_1580 and Cthe_1581) or (Cthe_0391 and Cthe_0392 and Cthe_0393)
R_FACOAL141	Cthe_1232
R_FACOAL140	Cthe_1232

www.manaraa.com

R_CYSt2r	.
R_TRPTRS	Cthe_0686
R_PAPSR	Cthe_2535
R_FACOAL171_anteiso_	Cthe_1232
R_NTP7	Cthe_0189
R_NTD9	Cthe_2359
R_NTD8	Cthe_2359
R_ACODA	Cthe_2073
R_cellpp	Cthe_0275
R_NTD7	.
R_NTD6	.
R_ALAR	Cthe_2697
R_GUAD	Cthe_2515
R_GLUDy	Cthe_0374
R_PYRt2r	.
R_PRAMPC	Cthe_2889
R_ALAt2r	.
R_G5SADr	.
R_cellulase3	(Cthe_0269 or Cthe_0536 or Cthe_2807 or Cthe_0825 or Cthe_0543 or Cthe_2872 or Cthe_1472 or Cthe_0040 or Cthe_2089) and (Cthe_0412 or Cthe_0071 or Cthe_1235 or Cthe_2089)
R_cellulase2	(Cthe_0269 or Cthe_0536 or Cthe_2807 or Cthe_0825 or Cthe_0543 or Cthe_2872 or Cthe_1472 or Cthe_0040 or Cthe_2089) and (Cthe_0412 or Cthe_0071 or Cthe_1235 or Cthe_2089)
R_cellulase5	(Cthe_0269 or Cthe_0536 or Cthe_2807 or Cthe_0825 or Cthe_0543 or Cthe_2872 or Cthe_1472 or Cthe_0040 or Cthe_2089) and (Cthe_0412 or Cthe_0071 or Cthe_1235 or Cthe_2089)
R_cellulase4	(Cthe_0269 or Cthe_0536 or Cthe_2807 or Cthe_0825 or Cthe_0543 or Cthe_2872 or Cthe_1472 or Cthe_0040 or Cthe_2089) and (Cthe_0412 or Cthe_0071 or Cthe_1235 or Cthe_2089)
R_cellulase6	(Cthe_0269 or Cthe_0536 or Cthe_2807 or Cthe_0825 or Cthe_0543 or Cthe_2872 or Cthe_1472 or Cthe_0040 or Cthe_2089) and (Cthe_0412 or Cthe_0071 or Cthe_1235 or Cthe_2089)
R_TEICH45	.
R_GLUt2r	.
R_ASP1DC	Cthe_0900
R_GALabc	(Cthe_1579 and Cthe_1580 and Cthe_1581) or (Cthe_0391 and Cthe_0392 and Cthe_0393)
R_ACHBS	Cthe_2517 and (Cthe_2516 or Cthe_2714)
R_UDPG4E	Cthe_0186
R_FMETTRS	Cthe_0568
R_GLUTRS	Cthe_0648
R_GLUTRR	Cthe_2525
R_ASPCT	Cthe_0953
R_G5SD	Cthe_0262
R_GLYK	Cthe_2442
R_ANS	Cthe_0874 and Cthe_0875
R_HSTPT	Cthe_2883 or Cthe_0610
R_TRDR	Cthe_1945
R_cellottx	Cthe_2446 and Cthe_0393
R_ADSL2r	Cthe_0741
R_THRTRS	Cthe_1228
R_MALt2r	.
R_FACOAL150_anteiso_	Cthe_1232
R_GLCP	Cthe_0357
R_TMDS	Cthe_1227
R_DPCOAK	Cthe_0885

المنارة للاستشارات
www.manaraa.com

R_DRPA	Cthe_1583 or Cthe_1943
R_FRUK	.
R_UDCPKr	Cthe_2305
R_DMATT	.
R_POR2_i	(Cthe_2390 or Cthe_2794) and (Cthe_2391 or Cthe_2795) and (Cthe_2392 or Cthe_2796) and (Cthe_2393 or Cthe_2797)
R_UAGCVT	(Cthe_2615 or Cthe_0973) or (Cthe_0441 or Cthe_2328)
R_cellohtx	Cthe_2446 and Cthe_0393
R_DHDPRy	Cthe_0963 or Cthe_1170
R_GF6PTA	Cthe_1162
R_NO3t7	.
R_PROTRS	Cthe_2947
R_HSK	.
R_PHEt2r	.
R_ACCOAC	.
R_TPI	Cthe_0139
R_GRTT	.
R_DDPGA	Cthe_0699 and Cthe_0700 and Cthe_0701
R_ASPTA	Cthe_0755 or Cthe_0363
R_NDPK8	Cthe_0716
R_NDPK7	Cthe_0716
R_NDPK5	Cthe_0716
R_NDPK4	Cthe_0716
R_NDPK3	Cthe_0716
R_NDPK2	Cthe_0716
R_NDPK1	Cthe_0716
R_DUTPDP	Cthe_2085
R_GGTT	.
R_FORt3	.
R_DAGK_CT	.
R_LEUTAi	Cthe_0856
R_LACZ	.
R_ENO	Cthe_0143
R_DAPabc	.
R_G3POA_CT	.
R_TYRTA	Cthe_0755 or Cthe_2883 or Cthe_0610 or Cthe_0363
R_ASAD	Cthe_0961
R_GLUR	Cthe_1937
R_DCTPD2	Cthe_1477
R_AMAOTr	Cthe_0025
R_RNTR2	Cthe_0646
R_RNTR3	Cthe_0646
R_UGT_CT	.
R_KAS180iso	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_ADSL1r	Cthe_0741
R_PTAr	Cthe_1029
R_OCBT	Cthe_1869
R_CYTDK1	Cthe_0332
R_CYTDK2	Cthe_0332
R_BIOMASScell	.
R_IPDDI	.
R_ACBIPGT	Cthe_3151
R_GMPS	.
R_CO2t	.
R_GBEZ	Cthe_2191

www.manaraa.com

R_ASPTRS	Cthe_1543 or Cthe_1331
R_OORr	Cthe_0864 and Cthe_0865 and Cthe_0866 and Cthe_0867
R_KAS100iso	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_H2td	.
R_HSST	Cthe_1845
R_GLGC	Cthe_3167 and Cthe_3166
R_FOLD3	Cthe_2581
R_PTA2	Cthe_1029
R_FDH	.
R_PRMICIi	Cthe_2887
R_H2Ot	.
R_ASPO5	Cthe_2355
R_MGt5	.
R_UGMDDS	Cthe_0977
R_VALTA	Cthe_0856
R_DAPDC	Cthe_0683
R_PHETA1	Cthe_0755 or Cthe_2883 or Cthe_0610 or Cthe_0363
R_UPP3S	Cthe_2528
R_IGPS	Cthe_0872
R_PPTGS_CT	.
R_ARGDC	Cthe_2108 or Cthe_1918
R_FACOAL160_ISO_	Cthe_1232
R_ILETA	Cthe_0856
R_PGM	Cthe_0140 or Cthe_2449 or Cthe_0707 or Cthe_0946 or Cthe_1292 or Cthe_1435
R_PRO1x_r	Cthe_0672
R_PGK	Cthe_0138
R_PGI	Cthe_0217
R_FMNAT	Cthe_0987
R_TRPS3	Cthe_1411 and (Cthe_1211 or Cthe_1412)
R_TRPS2	Cthe_1411 and (Cthe_1211 or Cthe_1412)
R_TRPS1	Cthe_1411 and (Cthe_1211 or Cthe_1412)
R_Kabc	.
R_FTHFCL	Cthe_1295
R_HISDC	Cthe_3028
R_KAS6	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS7	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS4	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS5	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS2	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS3	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS1	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_ARGSL	Cthe_0178
R_SHKK	Cthe_2437 or Cthe_0731
R_KAS8	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS9	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_FRUpts2	.

www.manaraa.com

R_AHSERL	Cthe_1569 or Cthe_1842
R_THRt2r	.
R_KAS130iso	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_OAAdecarb	Cthe_0699 and Cthe_0700 and Cthe_0701
R_MAN6PI	Cthe_3116
R_PPNCL3	Cthe_1313
R_UAG4E	Cthe_0229
R_LDH_L	Cthe_1053 or Cthe_0345
R_ASPKi	Cthe_1375
R_FACOAL171_ISO_	Cthe_1232
R_APRAUR	Cthe_0104
R_FRTT	.
R_PPTT	.
R_EHGLAT	Cthe_0755 or Cthe_0363
R_ASPt2r	.
R_NADS1	Cthe_0325
R_AGMT	Cthe_0695
R_PSERT	Cthe_0295
R_PSDC_CT	.
R_DB4PS	.
R_PPCOAC	Cthe_0699 and Cthe_0700 and Cthe_0701
R_ADPRDP	Cthe_0674
R_TALA	.
R_IMPD	Cthe_0681
R_TMDPP	Cthe_0678
R_H2St	.
R_IMPC	Cthe_1246
R_UPPRT	Cthe_2598 or Cthe_0954
R_DHORTS	Cthe_0952
R_MTRI	Cthe_0621
R_ACONT	Cthe_3158
R_HEX1	Cthe_2938
R_HEX7	Cthe_0389
R_PPNDH	Cthe_2260
R_TMDPK	Cthe_0577
R_UAMAS	Cthe_2626
R_PUNP5	Cthe_1201 or Cthe_1575
R_PUNP4	Cthe_1201 or Cthe_1575 or Cthe_0678
R_PUNP7	Cthe_1201 or Cthe_1575
R_PUNP6	Cthe_1201 or Cthe_1575 or Cthe_0678
R_PUNP1	Cthe_1201 or Cthe_1575
R_PUNP3	Cthe_1201 or Cthe_1575
R_NNDMBRT	Cthe_1297
R_XANt2r	.
R_LIPO2S24_CT	.
R_KAS141	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_L_DASH_LACt4r	.
R_BIOMASS	.
R_PACCOAL2	Cthe_0615 or Cthe_2792
R_PACCOAL3	Cthe_0615 or Cthe_2792
R_UREAabc	.
R_AMAA	Cthe_0185 or Cthe_0605 or Cthe_1804 or Cthe_1953 or Cthe_2498
R_CDPMEK	Cthe_2403
R_UAMAGS	Cthe_1041

المنارة للاستشارات
www.manaraa.com

R_MTHFR3	Cthe_0099
R_RNTR1	Cthe_0646
R_ARGSSr	Cthe_0179
R_CODHr	Cthe_2801
R_SPTi	.
R_GLCt5	Cthe_2446 and Cthe_0393
R_GAPD	Cthe_0137
R_CYSTGL	Cthe_1559 or Cthe_2799
R_LIPO1S24_CT	.
R_RNDR4	Cthe_0053
R_RNDR2	Cthe_0053
R_FACOAL140_ISO_	Cthe_1232
R_RNDR1	Cthe_0053
R_PABB	Cthe_0874 or Cthe_0875
R_SHSL4r	Cthe_1559 or Cthe_2799
R_GARFTi	Cthe_1247
R_CHORS	Cthe_0732
R_ACGK	Cthe_1864
R_PGAMT	Cthe_1163
R_ACGS	Cthe_0081
R_PIt2r	.
R_ATPPRT	Cthe_2881 and Cthe_2880
R_RBFSb	Cthe_0105 and Cthe_0107
R_RBFSa	Cthe_0105 and Cthe_0107
R_IOR	Cthe_0614
R_GLYAT	Cthe_0022
R_CODH2	Cthe_2801
R_PPS	Cthe_1253
R_G3PCT	.
R_FUM	Cthe_3090 and Cthe_3091
R_ORNTAC	Cthe_0081
R_PTPATi	Cthe_1276
R_PPA	Cthe_1425
R_PPCDC	Cthe_1313
R_PPM	Cthe_1265
R_IOR3	Cthe_0614
R_AHSERL2	Cthe_1569 or Cthe_1842
R_NTP3e	Cthe_0189
R_LYSTRS	Cthe_2815
R_ADSK	Cthe_2537
R_bglu1	Cthe_0212 or Cthe_1256
R_ADSS	Cthe_3093
R_bglu3	Cthe_0212 or Cthe_1256
R_bglu4	Cthe_0212 or Cthe_1256
R_bglu5	Cthe_0212 or Cthe_1256
R_bglu6	Cthe_0212 or Cthe_1256
R_BACCL	Cthe_2585
R_DTMPK	Cthe_2107
R_DHNPA	Cthe_2582
R_NTP5	Cthe_0189
R_NTP4	Cthe_0189
R_ACOTA	Cthe_1866
R_NTP6	Cthe_0189
R_NTP1	Cthe_0189
R_SHK3D	Cthe_0854
R_NTP3	Cthe_0189

المنارة للاستشارات

www.manaraa.com

R_NTP2	Cthe_0189
R_DHQD	Cthe_0849
R_LYSt2r	.
R_NTP9	Cthe_0189
R_NTP8	Cthe_0189
R_MOHMT	Cthe_0902
R_ACLS	Cthe_2517 and (Cthe_2516 or Cthe_2714)
R_PAP_CT	.
R_ASNt2r	.
R_2HBO	Cthe_1053 or Cthe_0345
R_FACOAL170_anteiso_	Cthe_1232
R_DHQS	Cthe_0786
R_GLCS1	Cthe_1284 or Cthe_0282
R_ARGTRS	Cthe_1935
R_GTPDPK	Cthe_1344
R_ATPS2_i	(Cthe_2606 and Cthe_2608 and Cthe_2607 and Cthe_2605 and Cthe_2609) and (Cthe_2602 and Cthe_2604 and Cthe_2603)
R_ILEt2r	.
R_TECA4S_CT	.
R_PAPPT3	Cthe_0976
R_SPMDabc	Cthe_0747 and Cthe_0748 and Cthe_0749 and Cthe_0750
R_PPBNGS	Cthe_2529
R_CYTK1	Cthe_0712
R_HMBS	Cthe_2527
R_URAt2	.
R_ADOCBLS	Cthe_3152
R_URIDK1r	Cthe_1004
R_UPP3MT	Cthe_2528
R_THRS	Cthe_1381
R_P5CR	Cthe_0672
R_AKP1	.
R_AGPATr_CT	Cthe_0713
R_KARA1	Cthe_2518
R_KARA2	Cthe_2518
R_G3PD1	Cthe_1022
R_NMNTP	.
R_ARGt2r	.
R_DXPS	Cthe_0828
R_GALU	Cthe_2183
R_IGPDH	Cthe_2884
R_MGSA	Cthe_0095
R_DAPDH	Cthe_0922
R_TDPDRE	Cthe_2560 or Cthe_1366
R_ILETRS	Cthe_0787
R_PIabc	.
R_LEUTRS	Cthe_1237
R_ACNAM9PL	Cthe_2641 or Cthe_2229
R_METabc	.
R_ETOHt2r	.
R_GK1	Cthe_1315
R_GLYTRS	Cthe_1312
R_PMDPHT	.
R_AGDC_r	Cthe_2190
R_AIRCr	Cthe_1250
R_BIFHASE	Cthe_0426
R_GLNt2r	.

www.manaraa.com

R_XYLabc	(Cthe_1579 and Cthe_1580 and Cthe_1581) or (Cthe_0391 and Cthe_0392 and Cthe_0393)
R_PGPPH_CT	.
R_OHPBAT	Cthe_0295
R_NO3R1	.
R_AICART	Cthe_1246
R_GLYt2r	.
R_PHCYT_CT	Cthe_1000 or Cthe_0030
R_UGT1_CT	.
R_FDH2	.
R_URIK3	Cthe_0332
R_PEPCKr	Cthe_2874
R_AMID3	Cthe_1032
R_AMID2	Cthe_1032
R_QULNS	Cthe_2356
R_Nat1	.
R_DHFS	Cthe_1384
R_CDPGDH	Cthe_2561
R_FACOAL161	Cthe_1232
R_FACOAL160	Cthe_1232
R_TDPGDH	Cthe_1086
R_LEUt2r	.
R_FTHFL	Cthe_2399 or Cthe_1093
R_CLPNS2_CT	Cthe_1396
R_MECDPDH	Cthe_0997
R_LYSLG_CT	.
R_ALCD2x	Cthe_0101 or Cthe_0388 or Cthe_0394 or Cthe_0423 or Cthe_2445 or Cthe_2579
R_NTD4	.
R_VALTRS	Cthe_0324
R_METS	Cthe_0645
R_METTRS	Cthe_2096
R_URIDK2r	Cthe_2107
R_PGLYCP	Cthe_0261 or Cthe_1191
R_UAAGDS	Cthe_0978
R_FACOAL181	Cthe_1232
R_FACOAL180	Cthe_1232
R_METB1r	Cthe_1559 or Cthe_2799
R_NBAHH	Cthe_3149
R_PDX5PS	.
R_PFL	Cthe_0505
R_BTS3r	Cthe_0020 or Cthe_1839
R_RZ5PP	Cthe_3153
R_CDGPT_CT	Cthe_0941
R_UAPGR	Cthe_0112
R_PPND	Cthe_1796
R_DDPA	Cthe_0880 or Cthe_1795
R_TMPPP	Cthe_0601
R_cellttp	Cthe_0275
R_ANPRT	Cthe_0873
R_UAG2E	Cthe_2640 or Cthe_2601
R_PSCVTi	Cthe_3086 or Cthe_1797
R_SULabc	Cthe_2531 and Cthe_2532 and Cthe_2533 and Cthe_2534
R_GCALDt	.
R_TECA3S45	.
R_PACCOAL	Cthe_0615 or Cthe_2792
R_CYTD	Cthe_1582 or Cthe_1068
R_PROt4r	.

المنارة للاستشارات
www.manaraa.com

R_bglu5_ext	Cthe_0212 or Cthe_1256
R_MTHFD	Cthe_1093
R_MTHFC	Cthe_1093 or Cthe_2399
R_UAGPT3	.
R_PGCD	Cthe_3035
R_OMCDC	.
R_MECDPS	Cthe_2946
R_HEXTT	.
R_PTRCabc	Cthe_0747 and Cthe_0748 and Cthe_0749 and Cthe_0750
R_FACOAL180_ISO_	Cthe_1232
R_NAPRT	Cthe_1177
R_CHTNASE1	Cthe_0269 and Cthe_0270 and Cthe_0271
R_PGMT	Cthe_1265
R_DHFR	Cthe_1226
R_UDPDPS	Cthe_1001
R_MEPCT	Cthe_2941
R_ADOCBIK	Cthe_3151
R_G1SATi	Cthe_2530
R_PPAKr	Cthe_1028
R_GLNS	Cthe_0202 or Cthe_1539 or Cthe_0863
R_HXPRT	Cthe_2254
R_ACt2r	.
R_celloptx	Cthe_2446 and Cthe_0393
R_ME1	Cthe_0344 or (Cthe_0699 and Cthe_0700 and Cthe_0701)
R_G1PTT	Cthe_2562
R_ACNMSr	.
R_NADK	Cthe_0816
R_SUCCt2r	.
R_IPPMIa	Cthe_2210 and Cthe_2211
R_IPPMIb	Cthe_2210 and Cthe_2211
R_HISt2r	.
R_GHMT2r	Cthe_1058
R_bglu2_ext	Cthe_0212 or Cthe_1256
R_DAPE	Cthe_3100
R_PYNP1r	Cthe_0678
R_TKT1	(Cthe_2443 or Cthe_2704) and (Cthe_2444 or Cthe_2705)
R_TKT2	(Cthe_2443 or Cthe_2704) and (Cthe_2444 or Cthe_2705)
R_cellotettx	Cthe_2446 and Cthe_0393
R_DHAD1	Cthe_2713
R_URIK2	Cthe_0332
R_URIK1	Cthe_0332
R_DHAD2	Cthe_2713
R_FACOAL100	Cthe_1232
R_CITt2r	.
R_CMPSAS	Cthe_2639
R_SADT	Cthe_2536 and Cthe_2537
R_ADKd	Cthe_2924
R_METAT	Cthe_2251
R_OMPDC	Cthe_0951
R_DHPPDA	Cthe_0104
R_VALt2r	.
R_CYSS	Cthe_1560 or Cthe_1840
R_SERAT	Cthe_2066
R_CBPS	(Cthe_0950 or Cthe_1867) and Cthe_0949
R_NADOX	(Cthe_0341 and Cthe_0342) or (Cthe_0429 and Cthe_0430)
R_ADK1	Cthe_2924

www.manaraa.com

R00765	Cthe_0067
R_FACOAL130	Cthe_1232
R_RNDR3	Cthe_0053
R_CELLOME	.
R_RIBabc	(Cthe_1579 and Cthe_1580 and Cthe_1581) or (Cthe_0391 and Cthe_0392 and Cthe_0393)
R_MDH	Cthe_1053 or Cthe_0345
R_RPE	Cthe_0576
R_KAS17	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_ORPT	.
R_KAS11	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS12	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_KAS13	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_RPI	Cthe_2597
R_OBTFL	Cthe_0505
R_ACS	Cthe_0551
R_IPPS	Cthe_2519 or Cthe_1391
R_UGT2_CT	.
R_DHDPS	Cthe_0962
R_NIT_n1p4	Cthe_1566 and Cthe_1565 and Cthe_1573
R_ASNS2	Cthe_0069
R_ASNS1	Cthe_0556
R_UDCPDP	Cthe_2305
R_ATPS2Na_i	.
R_ALATA_L	.
R_AHC	Cthe_1200
R_SLDx	Cthe_0276
R_NMNATr	Cthe_1241
R_PFK_2	Cthe_1261 or Cthe_0347 or Cthe_1956
R_AMID4	Cthe_1032
R_BTSr	Cthe_0020 or Cthe_1839
R_CS	Cthe_3027
R_UDPGDr	Cthe_1352
R_NNDPR	Cthe_2354
R_IPDPS	Cthe_0714
R_ACKr	Cthe_1028
R_CYSTRS	Cthe_2065
R_GLUSy	Cthe_0197 and Cthe_0198 and Cthe_0201
R_NADPOX	Cthe_3003
R_ADPT	Cthe_1345 or Cthe_2254
R_FBA	Cthe_0349
R_Cut1	Cthe_1848
R_HISTP	Cthe_0724
R_GLNTRS	Cthe_0917
R_IG3PS	Cthe_2886
R_PNTK	Cthe_2588
R_ADNCYC	Cthe_2577
R_TECA1S45	.
R_HISTD	Cthe_2882
R_cellhp	Cthe_0275
R_DHPS	Cthe_2581
R_CDPDSP_CT	.
R_BTS2	Cthe_0020 or Cthe_1839

المنارة للاستشارات
www.manaraa.com

R_TYRTRS	Cthe_0723
R_CHORM	Cthe_0711
R_MCOATA	Cthe_0935
R_AGPR	Cthe_1863
R_KAS10	Cthe_0932 and Cthe_0933 and Cthe_0934 and Cthe_0935 and Cthe_0936 and Cthe_0937 and Cthe_0938
R_FACOAL161_ISO_	Cthe_1232
R_UREA	Cthe_1818 and Cthe_1816 and Cthe_1817
R_TYRt2r	.
R_PRAI	Cthe_0871
R_FRUpts	.
R_FACOAL170_ISO_	Cthe_1232
R_celltp	Cthe_0275
R_HSDxi	Cthe_0290 or Cthe_1376
R_PRASCS	Cthe_2885
R_UAG2EMA	Cthe_2640 or Cthe_2601
R_YUMPS	Cthe_0988 or Cthe_2939 or Cthe_0908
R_LIPO3S24_CT	.
R_AOXSr	Cthe_0022
R_PRFGS	Cthe_0554
R_G1PACT	Cthe_2629
R_ALAALAr	Cthe_1938
R_GLUPRT	Cthe_1249
R_CHTNASE	Cthe_0269 and Cthe_0270 and Cthe_0271
R_PPDK	Cthe_1308
R_UAGDP	Cthe_2629
R_PNP	Cthe_1201 or Cthe_1575
R_ADMDCr	Cthe_0715
R_RBFK	Cthe_0987
R_PSP_L	.
R_DBTSr	Cthe_0021
R_H2CO3D2	.
R_GTPCII	Cthe_0106
R_SUCOAS	.
R_bglu6_ext	Cthe_0212 or Cthe_1256
R_cellbp	Cthe_0275
R_CYSTS	Cthe_1560
R_Nat_4	.
R_NNAM	.
R_PNTOt2	.
R_FE3abc	Cthe_1586 and Cthe_1587 and Cthe_1588
R_HISTRS	Cthe_1332 or Cthe_2880
R_CYSTL	Cthe_1559
R_GLU5K	Cthe_1766
R_MTAP	Cthe_0622
R_bglu4_ext	Cthe_0212 or Cthe_1256
R_ICDHyr	Cthe_0285
R_GTPCI	Cthe_2203
R_SHSL2r	Cthe_1559 or Cthe_2799
R_FDXHASE	Cthe_3019 and Cthe_3020 and Cthe_3021 and Cthe_3022 and Cthe_3023 and Cthe_3024
R_DCMPDA	Cthe_2599
R_ACALD	Cthe_0423
R_CBIAT	Cthe_0965
R_PYNP2r	Cthe_0678
R_MMCD	Cthe_0699 and Cthe_0700 and Cthe_0701
R_bglu3_ext	Cthe_0212 or Cthe_1256

المنارة للاستشارات
www.manaraa.com

R_DXPRIi	Cthe_0999
R_PHETRS	Cthe_0214 and Cthe_0215
R_PRATPP	Cthe_2889
R_PRAIS	Cthe_1248
R_IOR2	Cthe_0614

D. Biomass Composition Calculations

CELLULAR COMPOSITION

Metabolite Class	**% w/w composition ***
Protein	52.85
DNA	2.60
RNA	6.55
Lipid	7.60
Lipoteichoic acid	3.04
Cell wall components	22.42
Ion and metabolite	4.94
Cellulosome (Zhang and Lynd, 2003)	20.00
	120.00

PROTEIN

Amino Acid	**Count (all ORFs) †**	**% Prevalence**	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol/gDW**
Alanine (A)	68886.00	6.38	89.05	568.10	4.35		0.2580
Arginine (R)	47155.00	4.37	175.11	764.71	5.85		0.1766
Asparagine (N)	58251.00	5.39	132.05	712.36	5.45		0.2182
Aspartic acid (D)	61489.00	5.69	132.04	751.90	5.75		0.2303
Cysteine (C)	12654.00	1.17	121.02	141.82	1.09		0.0474
Glutamate (E)	82288.00	7.62	146.05	1113.00	8.52		0.3082
Glutamine (Q)	27732.00	2.57	146.07	375.15	2.87		0.1039
Glycine (G)	72154.00	6.68	75.03	501.36	3.84		0.2703
Histidine (H)	15430.00	1.43	155.07	221.59	1.70		0.0578
Isoleucine (I)	91626.00	8.49	131.09	1112.36	8.51		0.3432
Leucine (L)	93461.00	8.66	131.09	1134.64	8.68		0.3501
Lysine (K)	87426.00	8.10	147.11	1191.08	9.12		0.3275
Methionine (M)	27111.00	2.51	149.05	374.23	2.86		0.1016
Phenylalanine (F)	45765.00	4.24	165.08	699.66	5.35		0.1714
Proline (P)	36619.00	3.39	115.06	390.20	2.99		0.1372
Serine (S)	65270.00	6.04	105.04	634.93	4.86		0.2445
Threonine (T)	54463.00	5.04	119.06	600.52	4.60		0.2040
Tryptophan (W)	9570.00	0.89	204.09	180.88	1.38		0.0358
Tyrosine (Y)	45643.00	4.23	181.07	765.38	5.86		0.1710
Valine (V)	76806.00	7.11	117.08	832.79	6.37		0.2877
Sum	**1079799**	**100**	**2736.31**	**13066.64**	**100**		

DNA

G+C content (%):	35.52						
DNA	**% Prevalence**	**MW (g/mol)**	**%P*MW**	**% (by weight)**			**mmol/gDW**
dATP	32.24	487.00	15700.91	32.53			0.0174
dCTP	17.76	462.99	8222.70	17.04			0.0096
dGTP	17.76	503.00	8933.20	18.51			0.0096
dTTP	32.24	477.99	15410.38	31.93			0.0174
Sum	**100**		**48267.18**	**100**			

RNA

RNA	**Count (all ORFs) †**	**% Prevalence**	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol/gDW**
ATP	1170579.0	30.5	503.00	15320.06	30.99		0.0404
CTP	751237.0	19.5	478.98	9362.55	18.94		0.0259

www.manaraa.com

GTP	747135.0	19.4	518.99	10089.14	20.41		0.0258
UTP	1174350.0	30.6	479.97	14665.80	29.67		0.0405
Sum	**3843301**	**100**		**49437.56**	**100**		

LIPIDS

Components (id)	**Content % (w/w) ***	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol / gDW**
Monoglucosyldiacylglycerol (m12dg_CT)	8.1	718.9	5823.283	7.72		0.0086
Diglucosyldiacylglycerol (d12dg_CT)	12.8	880.8	11274.6112	14.95		0.0110
Triglucosyldiacylglycerol (t12dg_CT)	9.1	1043.0	9491.027	12.58		0.0066
Cardiolipin (cdlp_CT)	0.8	1327.099	1061.6792	1.41		0.0005
Phosphatidylglycerol (pgly_CT)	16.5	709.6	11708.3505	15.52		0.0177
Lysylphosphatidylglycerol (lysylpgly_CT)	2.4	838.8	2013.072	2.67		0.0022
Phosphatidylethanolmaine (psetha_CT)	50.2	678.6	34065.0674	45.16		0.0562
Sum	**99.9**		**75437.1**	**100.0**		

Fatty Acids (Herrero et al., 1982)	**Metabolite abbr.**	**% (by weight)**	**Normalized % w/w**	**MW (g/mol)**	**Averaged MW (g/mol)**		**% molar composition**
C10:0 branched	dca / dcacoa	7.4	7.42971	171.3	12.7271		0.05104068
C13:0 branched	trdca / trdacoa	2.6	2.61044	213.34	5.56911		0.022334334
C14:0 branched (iso)	fa1 / fa1coa	2.5	2.51004	227.4	5.70783		0.022890635
C14:0 normal	ttdca / tdcoa	1	1.00401	227.4	2.2831		0.009156254
C14:1	ttdce / ttdcecoa	2.4	2.40963	225.4	5.43132		0.021781738
C16:0 branched (iso)	fa6 / fa6coa	31.4	31.5261	255.4	80.5176		0.322907337
C16:0 normal	hdca / pmtcoa	22.3	22.3895	255.4	57.1829		0.229325911
C17:0 branched (anteiso)	fa12 / fa12coa	12.3	12.3493	269.4	33.2692		0.133422807
C16:1	hdcea / hdcoa	12.3	12.3493	253.4	31.2933		0.125498662
C18:0 branched (iso)	ocdcaiso / strcoaiso	4.2	4.2168	283.5	11.9548		0.047943499
C18:0 normal	ocdca / strcoa	1.2	1.2048	283.5	3.41566		0.013698143
Sum		**99.6**	**100**	**2665.44**	**249.3522**		**1**
					"average" fatty acid		

LIPOTEICHOIC ACIDS

Components (id)	**Content % (w/w) ***	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol / gDW**
Lipoteichoic acid (n=24), linked, glucose substituted	19	8445.31	1604.61	22.6794		0.0007
Lipoteichoic acid (n=24), linked, N-acetyl-D-glucosamine	19	9430.58	1791.81	25.325		0.0006
Lipoteichoic acid (n=24), linked, D-alanine substituted	40	6692.1	2676.87	37.8348		0.0018
Lipoteichoic acid (n=24), linked, unsubstituted	22	4553.93	1001.86	14.160		0.0015
Sum	**100.0**		**7075.2**	**100.0**		

CELL WALL

Components (id)	**Content**	**MW**	**%P*MW**	**% (mole)**		**mmol / gDW**

المنارة للاستشارات

www.manaraa.com

	% (w/w) *	(g/mol)				
Peptidoglycan subunit	45	991	445.95	6.27862		0.1018
Teichoic acid						
Glycerol teichoic acid (n=45), unlinked, unsubstituted	11.9	7373.6	877.45	12.3539		0.0036
Glycerol teichoic acid (n=45), unlinked, D-ala substituted	11.9	11382.8	1354.5	19.071		0.0023
Glycerol teichoic acid (n=45), unlinked, glucose substituted	11.9	14688	1747.8	24.6086		0.0018
Minor teichoic acid (acetylgalactosamine glucose phosphate, n=30)	19.3	13869.6	2676.8	37.6877		0.0031
Sum	**100.0**		**7102.7**	**100.0**		
IONS AND METABOLITES						
Components (id)	**Content % (w/w) ***	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol / gDW**
K	86	39.1	33.626	5.436445038		1.0865
Mg	7.7	24.3	1.8711	0.30250795		0.1565
Fe(+3)	0.6	55.9	0.3354	0.054225411		0.0053
Ca	0.4	40.1	0.1604	0.025932486		0.0049
Phosphate	4.3	96	4.128	0.667389672		0.0221
Diphospate	0.5	174.9	0.8745	0.141383786		0.0014
Sum	**99.5**					
Components (id)	**Content % (w/w) ***	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol / gDW**
NAD	61.9	662.4	410.0256	66.29041927		0.0462
AMP	9.3	345.2	32.1036	5.190312761		0.0133
ATP	8.7	503.2	43.7784	7.07782268		0.0085
ADP	6.3	424.2	26.7246	4.320669097		0.0073
CMP	1.9	321.2	6.1028	0.9866632		0.0029
NADP	4	740.4	29.616	4.78813288		0.0027
CTP	1.5	479.1	7.1865	1.161869157		0.0015
GMP	1.1	361.2	3.9732	0.642362559		0.0015
GTP	1.3	519.1	6.7483	1.091023674		0.0012
CDP	0.6	400.2	2.4012	0.38821126		0.0007
NADPH	0.9	741.4	6.6726	1.078784963		0.0006
GDP	0.5	440.2	2.201	0.355844154		0.0006
Sum	**98.0**		**618.5**	**100.0**		
CELLULOSOME						
Components (id)	**Content % (w/w)**	**MW (g/mol)**	**%P*MW**	**% (by weight)**		**mmol / gDW**
Cellulosome [e]	100	128.6019662	128.6019662	100		1.5552

NOTES

* In the absence of data specific to *C. thermocellum* data was adapted from genome scale model of *B. subtilis* (Oh et al., 2007)

† Counts of amino acids or ribonucleotides were performed on published sequence of *C. thermocellum*, including only open reading frames.

REFERENCES

Gold N, Martin V (2007) Global View of the Clostridium thermocellum Cellulosome Revealed by Quantitative Proteomic Analysis *J Bacteriol* **189:**19 pp 6787 - 6795

Herrero A, Gomez R, Roberts M (1982) Ethanol-induced changes in the membrane lipid composition of Clostridium thermocellum *Biochim Biophys Acta - Biomembranes* **693:**1 pp 195 - 204

Oh Y, Palsson B, Park S, Schilling C, Mahadevin R (2007) Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data *J Biol Chem* **282:**39 pp 28791 - 28799

Zhang Y, Lynd L (2003) Quantification of Cell and Cellulase Mass Concentrations During Anaerobic Cellulose Fermentation: Development of an Enzyme-Linked Immunosorbent Assay-Based Method with Application to Clostridium thermocellum Batch Cultures. *Anal Chem* **75:**2 pp 219-227

E. Cellulosome Composition Calculations

المنارة للاستشارات
www.manaraa.com

GenInfo identifier	Protein	(Putative) function or activity	peptide ions	emPAI		% composition
125975556	CipA	Scaffoldin	42	5.92		12.43697479
125972933	CelK	Exoglucanase (GH9)	39	4.12		8.655462185
125974579	CelS	Exoglucanase (GH48)	29	3.56		7.478991597
125973097	CelR	Endoglucanase (GH9)	28	3.19		6.701680672
125975557	OlpB	Cell-surface anchor	27	2.75		5.777310924
125973339		GH5	15	2.59		5.441176471
125972791	CelA	Endoglucanase (GH8)	14	2.46		5.168067227
125973315	CelE	Endoglucanase (GH5), CE2	24	2.24		4.705882353
125973142	CelJ	Endoglucanase (GH9), GH44	42	2.16		4.537815126
125974342	XynC	Xylanase (GH10)	18	1.57		3.298319328
125974464	XynZ	Xylanase (GH10), CE1	18	1.57		3.298319328
125972934	CbhA	Exoglucanase (GH9)	30	1.51		3.172268908
125975294	CelT	Endoglucanase (GH9)	14	1.29		2.710084034
125975353	CelG	Endoglucanase (GH5)	9	1.15		2.415966387
125975452	XynA	Xylanase (GH11), CE4	12	1.15		2.415966387
125973912	XghA	Xyloglucanase (GH74)	21	1.13		2.37394958
125973263		GH9	15	0.94		1.974789916
125973062	CelF	Endoglucanase (GH9)	12	0.9		1.890756303
125973254		Cell-surface anchor	21	0.82		1.722689076
125974678		GH5	11	0.71		1.491596639
125972735	LicBb	Lichenase (GH16)	4	0.62		1.302521008
125975293	ManAb	Mannanase (GH26)	7	0.61		1.281512605
125973143	CelQb	Endoglucanase (GH9)	9	0.58		1.218487395
125972556		GH26	5	0.53		1.113445378
125973055	CelB	Endoglucanase (GH5)	5	0.53		1.113445378
125975558	Orf2p	Cell-surface anchor	7	0.5		1.050420168
125972796		GH9	5	0.49		1.029411765
125975243		GH9	8	0.44		0.924369748
125972926		GH5	3	0.33		0.693277311
125972567	CelNb	Endoglucanase (GH9)	4	0.27		0.567226891
125973343	CelD	Endoglucanase (GH9)	4	0.23		0.483193277
125972954		GH9	4	0.21		0.441176471
125972792	ChiA	Chitinase (GH18)	2	0.2		0.420168067
125973914		GH53	2	0.17		0.357142857
125973158		Endopygalactorunase	2	0.16		0.336134454
			SUM	47.6		

	% prevalence in cellulosome (molar)		MW (g/mol)	Averaged MW (g/mol cellome)
Alanine (A)	7.39613	0.073961334	89.05	6.58626
Arginine (R)	3.11852	0.031185177	175.11	5.46084
Asparagine (N)	5.65007	0.056500707	132.05	7.46092
Aspartic acid (D)	7.22148	0.072214845	132.04	9.53525
Cysteine (C)	0.84151	0.008415129	121.02	1.0184

www.manaraa.com

Glutamate (E)	5.40692	0.054069231	146.05	7.89681
Glutamine (Q)	2.39086	0.02390862	146.07	3.49233
Glycine (G)	8.29389	0.0829389	75.03	6.22291
Histidine (H)	1.37434	0.013743373	155.07	2.13118
Isoleucine (I)	6.08785	0.060878523	131.09	7.98057
Leucine (L)	6.58014	0.065801384	131.09	8.6259
Lysine (K)	5.65472	0.056547196	147.11	8.31866
Methionine (M)	1.9202	0.019202024	149.05	2.86206
Phenylalanine (F)	3.66772	0.03667715	165.08	6.05466
Proline (P)	6.28555	0.062855534	115.06	7.23216
Serine (S)	6.15343	0.061534329	105.04	6.46357
Threonine (T)	7.33637	0.073363664	119.06	8.73468
Tryptophan (W)	2.17608	0.021760768	204.09	4.44116
Tyrosine (Y)	5.49146	0.05491455	181.07	9.94338
Valine (V)	6.95276	0.069527559	117.08	8.14029
				128.602
				"averaged" amino acid

Cellulosome Reaction			
Alanine (A)	ala-L[c]	0.073961334	reactant
Arginine (R)	arg-L[c]	0.031185177	reactant
Asparagine (N)	asn-L[c]	0.056500707	reactant
Aspartic acid (D)	asp-L[c]	0.072214845	reactant
Cysteine (C)	cys-L[c]	0.008415129	reactant
Glutamate (E)	glu-L[c]	0.054069231	reactant
Glutamine (Q)	gln-L[c]	0.02390862	reactant
Glycine (G)	gly[c]	0.0829389	reactant
Histidine (H)	his-L[c]	0.013743373	reactant
Isoleucine (I)	iso-L[c]	0.060878523	reactant
Leucine (L)	leu-L[c]	0.065801384	reactant
Lysine (K)	lys-L[c]	0.056547196	reactant
Methionine (M)	met-L[c]	0.019202024	reactant
Phenylalanine (F)	phe-L[c]	0.03667715	reactant
Proline (P)	pro-L[c]	0.062855534	reactant
Serine (S)	ser-L[c]	0.061534329	reactant
Threonine (T)	thr-L[c]	0.073363664	reactant
Tryptophan (W)	trp-L[c]	0.021760768	reactant
Tyrosine (Y)	tyr-L[c]	0.05491455	reactant
Valine (V)	val-L[c]	0.069527559	reactant
ATP	atp[c]	1	reactant
Cellulosome	cellome[e]	1	product
ADP	adp[c]	1	product
Inorganic phosphate	pi[c]	1	product

المنارة للاستشارات
www.manaraa.com

F. Changes to original *i*SR432 model, as of May 2011.

```
Added 13 reactions:
R_NADOX     hydrogen:NAD+ oxidoreductase  True  Miscellaneous     1.12.1.2
      [c] : h2 + nad <==> h + nadh  (Cthe_0341 and Cthe_0342) or (Cthe_0429
and Cthe_0430)
R_PEPCKr    Phosphoenolpyruvatecarboxykinase(GTP)     True
      Glycolysis/Gluconeogenesis    .     [c] : gtp + oaa <==> co2 + gdp +
pep   Cthe_2874
R_NADPOX    hydrogen:NADP+ oxidoreductase True  Miscellaneous     .     [c] :
h2 + nadp <==> h + nadph      Cthe_3003
R_BIFHASE   Bifurcating hydrogenase True  Miscellaneous     .     [c] : 3 h +
2 fdred + nadh <==> fdox + nad + 2 h2      None
R_FRUpts    D-fructosetransportviaPEP:PyrPTS    False Transport,
Extracellular     .     fru[e] + pep[c] --> f1p[c] + pyr[c] None
R_ACGAMK    N-acetylglucosaminekinase     False Aminosugar Metabolism   .
      [c] : acgam + atp --> acgam6p + adp + h    None
R_XANt2r    xanthinetransportinviaprotonsymport, reversible True  Transport,
Extracellular     .     h[e] + xan[e] <==> h[c] + xan[c]    None
R_NTD9      _5'-nucleotidase(GMP)   False Nucleotides 3.1.3.5     [c] : gmp +
h2o --> gsn + pi  Cthe_2359
R_NTD8      _5'-nucleotidase(dGMP)  False Nucleotides 3.1.3.5     [c] : dgmp
+ h2o --> dgsn + pi     Cthe_2359
R00765      D-glucosamine-6-phosphate aminohydrolase (ketol isomerizing)
      True  Aminosugar Metabolism   .     [c] : gam6p + h2o <==> f6p + nh4
      Cthe_0067
R_NTD4      _5'-nucleotidase(CMP)   False Pyrimidine Catabolism   .     [c] :
cmp + h2o --> cytd + pi None
R_NTD7      _5'-nucleotidase(AMP)   False Nucleotides .     [c] : amp + h2o -
-> adn + pi None
R_NTD6      _5'-nucleotidase(dAMP)  False Nucleotides .     [c] : damp + h2o
--> dad-2 + pi    None

Removed 33 reactions:
R_EXCH_ser_DASH_l_e     ser-l[e] exchange flux  True  ExchangeFlux      .
      ser-l[e] <==> ser-l[b]  None
R_EXCH_ile_DASH_l_e     ile-l[e] exchange flux  True  ExchangeFlux      .
      ile-l[e] <==> ile-l[b]  None
R_EXCH_gly_e      gly[e] exchange flux    True  ExchangeFlux      .
      gly[e] <==> gly[b]      None
R_EXCH_gln_DASH_l_e     gln-l[e] exchange flux  True  ExchangeFlux      .
      gln-l[e] <==> gln-l[b]  None
R_EXCH_tyr_DASH_l_e     tyr-l[e] exchange flux  True  ExchangeFlux      .
      tyr-l[e] <==> tyr-l[b]  None
R_EXCH_met_DASH_l_e     met-l[e] exchange flux  True  ExchangeFlux      .
      met-l[e] <==> met-l[b]  None
R_SPODM     superoxidedismutase     False Miscellaneous     .     [c] : 2 h +
2 o2s --> h2o2 + o2     Cthe_3008
R_NADH6     NADHdehydrogenase(ubiquinone-8&#38;3.5protons)  False
      Miscellaneous     1.6.5.3     4.5 h[c] + nadh[c] + q8[c] --> 3.5 h[e] +
nad[c] + q8h2[c]  Cthe_3019 and Cthe_3020 and Cthe_3021 and Cthe_3022 and
Cthe_3023 and Cthe_3024
R_EXCH_asn_DASH_l_e     asn-l[e] exchange flux  True  ExchangeFlux      .
      asn-l[e] <==> asn-l[b]  None
R_XANt2     xanthinetransportinviaprotonsymport False Transport,
Extracellular     .     h[e] + xan[e] --> h[c] + xan[c]     None
```

المنارة للاستشارات

www.manaraa.com

```
R_EXCH_pro_DASH_l_e     pro-l[e] exchange flux  True  ExchangeFlux      .
      pro-l[e] <==> pro-l[b]  None
R_EXCH_leu_DASH_l_e     leu-l[e] exchange flux  True  ExchangeFlux      .
      leu-l[e] <==> leu-l[b]  None
R_EXCH_ala_DASH_l_e     ala-l[e] exchange flux  True  ExchangeFlux      .
      ala-l[e] <==> ala-l[b]  None
R_EXCH_trp_DASH_l_e     trp-l[e] exchange flux  True  ExchangeFlux      .
      trp-l[e] <==> trp-l[b]  None
R_EXCH_cys_DASH_l_e     cys-l[e] exchange flux  True  ExchangeFlux      .
      cys-l[e] <==> cys-l[b]  None
R_EXCH_lys_DASH_l_e     lys-l[e] exchange flux  True  ExchangeFlux      .
      lys-l[e] <==> lys-l[b]  None
R_EXCH_glu_DASH_l_e     glu-l[e] exchange flux  True  ExchangeFlux      .
      glu-l[e] <==> glu-l[b]  None
R_PYK pyruvatekinase    False Glycolysis/Gluconeogenesis    .     [c] : adp +
h + pep --> atp + pyr   None
R_EXCH_asp_DASH_l_e     asp-l[e] exchange flux  True  ExchangeFlux      .
      asp-l[e] <==> asp-l[b]  None
R_PEPCK     Phosphoenolpyruvatecarboxykinase(GTP)     False
      Glycolysis/Gluconeogenesis    .     [c] : gtp + oaa --> co2 + gdp + pep
      Cthe_2874
R_EXCH_arg_DASH_l_e     arg-l[e] exchange flux  True  ExchangeFlux      .
      arg-l[e] <==> arg-l[b]  None
R_NFO_Na    NADH:Ferredoxin Oxidoreductase, Na-translocating      True
      Miscellaneous     3.D.6.1.1   nadh[c] + fdox[c] + na1[e] <==> nad[c] +
fdred[c] + na1[c] Cthe_2430 and Cthe_2431 and Cthe_2432 and Cthe_2433 and
Cthe_2434 and Cthe_2435
R_SBTD_Dr   D-sorbitoldehydrogenase(R)    True  Sugar Metabolism  .     [c] :
nad + sbt-d <==> fru + h + nadh     Cthe_0388
R07364      1,2-dihydroxy-5-(methylthio)pent-1-en-3-one:oxygen oxidoreductase
(formate-forming) True  Methionine Metabolism   .     [c] : C15606 + o2 <==>
C01180 + for      None
R_EXCH_thr_DASH_l_e     thr-l[e] exchange flux  True  ExchangeFlux      .
      thr-l[e] <==> thr-l[b]  None
R_NFO_h     NADH:Ferredoxin Oxidoreductase, H-translocating True
      Miscellaneous     3.D.6.1.1   nadh[c] + fdox[c] + h[e] <==> nad[c] +
fdred[c] + h[c]   Cthe_2430 and Cthe_2431 and Cthe_2432 and Cthe_2433 and
Cthe_2434 and Cthe_2435
R_EXCH_val_DASH_l_e     val-l[e] exchange flux  True  ExchangeFlux      .
      val-l[e] <==> val-l[b]  None
R_MCITS      _2-methylcitratesynthase      False Miscellaneous     2.3.3.5
       [c] : h2o + oaa + ppcoa --> 2mcit + coa + h      Cthe_3027
R07392      S-Methyl-5-thio-D-ribulose-1-phosphate hydro-lyase    True
      Methionine Metabolism    .     [c] : 5mdru1p <==> C15650 + h2o     None
R07395      5-(Methylthio)-2,3-dioxopentyl-phosphate phosphohydrolase
(isomerizing)     True  Methionine Metabolism   .     [c] : C15650 + h2o <==>
C15606 + pi None
R07396      .     False Methionine Metabolism   .     [c] : C01180 + glu-l --
> met-l + akg     None
R_EXCH_phe_DASH_l_e     phe-l[e] exchange flux  True  ExchangeFlux      .
      phe-l[e] <==> phe-l[b]  None
R_EXCH_his_DASH_l_e     his-l[e] exchange flux  True  ExchangeFlux      .
      his-l[e] <==> his-l[b]  None

Modified 24 reactions:
```

www.manaraa.com

```
old:  R_GLUt2r    L-glutamatetransportviaprotonsymport,reversible True
      Transport, Extracellular      .     glu-l[e] + h[e] <==> glu-l[c] +
h[c]  None
new:  R_GLUt2r    L-glutamatetransportviaprotonsymport,reversible True
      Transport, Extracellular      .     glu-l[e] + h[e] <==> glu-l[c] +
h[c]  None

old:  R_H2CO3D2   carboxylicaciddissociation    True  Miscellaneous     .
      [c] : h + hco3 <==> h2co3     spontaneous
new:  R_H2CO3D2   carboxylicaciddissociation    True  Miscellaneous     .
      [c] : h + hco3 <==> h2co3     None

old:  R_PHEt2r    L-phenylalaninereversibletransportviaprotonsymport    True
      Transport, Extracellular      .     h[e] + phe-l[e] <==> h[c] + phe-
l[c]  None
new:  R_PHEt2r    L-phenylalaninereversibletransportviaprotonsymport    True
      Transport, Extracellular      .     h[e] + phe-l[e] <==> h[c] + phe-
l[c]  None

old:  R_UAGCVT    UDP-N-acetylglucosamine1-carboxyvinyltransferase      False
      Cell Envelope Biosynthesis    .     [c] : pep + uacgam --> pi + uaccg
      Cthe_2615 or Cthe_0973 or Cthe_0441 or Cthe_2328
new:  R_UAGCVT    UDP-N-acetylglucosamine1-carboxyvinyltransferase      False
      Cell Envelope Biosynthesis    .     [c] : pep + uacgam --> pi + uaccg
      (Cthe_2615 or Cthe_0973) or (Cthe_0441 or Cthe_2328)

old:  R_BIOMASScell     Biomass flux describing cell mass alone, excluding
extracellular protein   False Biomass     .     [c] : 0 glycogen +
0.257980584 ala-l + 0.176597196 arg-l + 0.218152121 asn-l + 0.230278549 asp-l
+ 0.047389692 cys-l + 0.308171564 glu-l + 0.103857352 gln-l + 0.270219364 gly
+ 0.057785913 his-l + 0.343142715 ile-l + 0.350014856 leu-l + 0.327413561
lys-l + 0.101531685 met-l + 0.171391596 phe-l + 0.137139492 pro-l +
0.244438532 ser-l + 0.203965923 thr-l + 0.035839999 trp-l + 0.170934701 tyr-l
+ 0.287641274 val-l + 0.017366666 datp + 0.009566749 dctp + 0.009566749 dgtp
+ 0.017366666 dttp + 0.008562798 m12dg_ct + 0.011044141 d12dg_ct +
0.006631063 t12dg_ct + 4.58142E-4 cdlp_ct + 0.017672003 pgly_ct + 0.002174587
lysylpgly_ct + 0.05622271 psetha_ct + 6.83646E-4 lipo1_24_ct + 6.12248E-4
lipo2_24_ct + 0.001816095 lipo3_24_ct + 0.001467494 lipo4_24_ct + 0.101806256
peptido_ct + 0.003618287 gtca1_45_ct + 0.00234387 gtca2_45_ct + 0.001816435
gtca3_45_ct + 0.003119816 tcam_ct + 0.7063 k + 0.1017 mg2 + 0.0034 fe3 +
0.0032 ca2 + 9.0E-4 ppi + 0.0162 nad + 0.0047 amp + 0.0010 cmp + 9.0E-4 nadp
+ 5.0E-4 gmp + 3.0E-4 cdp + 2.0E-4 nadph + 2.0E-4 gdp + 0.0415 utp + 150 atp
+ 0.0394 ctp + 0.0674 gtp + 0 coa --> cellmass + 150 pi + 150 h + 150 adp
      None
new:  R_BIOMASScell     Biomass flux describing cell mass alone, excluding
extracellular protein   False Biomass     .     [c] : 0 glycogen +
0.257980584 ala-l + 0.176597196 arg-l + 0.218152121 asn-l + 0.230278549 asp-l
+ 0.047389692 cys-l + 0.308171564 glu-l + 0.103857352 gln-l + 0.270219364 gly
+ 0.057785913 his-l + 0.343142715 ile-l + 0.350014856 leu-l + 0.327413561
lys-l + 0.101531685 met-l + 0.171391596 phe-l + 0.137139492 pro-l +
0.244438532 ser-l + 0.203965923 thr-l + 0.035839999 trp-l + 0.170934701 tyr-l
+ 0.287641274 val-l + 0.017366666 datp + 0.009566749 dctp + 0.009566749 dgtp
+ 0.017366666 dttp + 0.008562798 m12dg_ct + 0.011044141 d12dg_ct +
0.006631063 t12dg_ct + 4.58142E-4 cdlp_ct + 0.017672003 pgly_ct + 0.002174587
lysylpgly_ct + 0.05622271 psetha_ct + 6.83646E-4 lipo1_24_ct + 6.12248E-4
lipo2_24_ct + 0.001816095 lipo3_24_ct + 0.001467494 lipo4_24_ct + 0.101806256
peptido_ct + 0.003618287 gtca1_45_ct + 0.00234387 gtca2_45_ct + 0.001816435
```

المنارة للاستشارات
www.manaraa.com

```
gtca3_45_ct + 0.003119816 tcam_ct + 0.7063 k + 0.1017 mg2 + 0.0034 fe3 +
0.0032 ca2 + 9.0E-4 ppi + 0.0162 nad + 0.0047 amp + 0.0010 cmp + 9.0E-4 nadp
+ 5.0E-4 gmp + 3.0E-4 cdp + 2.0E-4 nadph + 2.0E-4 gdp + 0.0415 utp + 150 atp
+ 0.0394 ctp + 0.0674 gtp + 0 coa --> cellmass + 150 pi + 150 h + 150 adp
      None

old:  R_HISt2r    L-histidinereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     h[e] + his-l[e] <==> h[c] + his-
l[c]  None
new:  R_HISt2r    L-histidinereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     h[e] + his-l[e] <==> h[c] + his-
l[c]  None

old:  R_TYRt2r    L-tyrosinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     h[e] + tyr-l[e] <==> h[c] + tyr-
l[c]  None
new:  R_TYRt2r    L-tyrosinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     h[e] + tyr-l[e] <==> h[c] + tyr-
l[c]  None

old:  R_ARGt2r    L-arganinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     arg-l[e] + h[e] <==> arg-l[c] +
h[c]  None
new:  R_ARGt2r    L-arganinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     arg-l[e] + h[e] <==> arg-l[c] +
h[c]  None

old:  R_LYSt2r    L-lysinereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + lys-l[e] <==> h[c] + lys-
l[c]  None
new:  R_LYSt2r    L-lysinereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + lys-l[e] <==> h[c] + lys-
l[c]  None

old:  R_CYSt2r    L-cysteinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     cys-l[e] + h[e] <==> cys-l[c] +
h[c]  None
new:  R_CYSt2r    L-cysteinereversibletransportviaprotonsymport   True
      Transport, Extracellular      .     cys-l[e] + h[e] <==> cys-l[c] +
h[c]  None

old:  R_H2CO3D    carboxylicaciddissociation    True  Miscellaneous     .
      [c] : co2 + h2o <==> h2co3    spontaneous
new:  R_H2CO3D    carboxylicaciddissociation    True  Miscellaneous     .
      [c] : co2 + h2o <==> h2co3    None

old:  R_ASNt2r    L-asparaginereversibletransportviaprotonsymport True
      Transport, Extracellular      .     asn-l[e] + h[e] <==> asn-l[c] +
h[c]  None
new:  R_ASNt2r    L-asparaginereversibletransportviaprotonsymport True
      Transport, Extracellular      .     asn-l[e] + h[e] <==> asn-l[c] +
h[c]  None

old:  R_OMCDC      _2-Oxo-4-methyl-3-carboxypentanoatedecarboxylation    False
      Valine, Leucine, and Isoleucine Metabolism      .      [c] : 3c4mop + h
--> 4mop + co2    spontaneous
```

www.manaraa.com

```
new:  R_OMCDC     _2-Oxo-4-methyl-3-carboxypentanoatedecarboxylation    False
      Valine, Leucine, and Isoleucine Metabolism      .     [c] : 3c4mop + h
--> 4mop + co2    None

old:  R_SERt2r    L-serinereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + ser-l[e] <==> h[c] + ser-
l[c]  None
new:  R_SERt2r    L-serinereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + ser-l[e] <==> h[c] + ser-
l[c]  None

old:  R_FDXHASE   ferredoxinhydrogenase   True  Miscellaneous     1.12.7.2
      [c] : 2 fdred + 2 h <==> h2 + 2 fdox      Cthe_3003
new:  R_FDXHASE   ferredoxinhydrogenase   True  Miscellaneous     1.12.7.2
      [c] : 2 fdred + 2 h <==> h2 + 2 fdox      Cthe_3019 and Cthe_3020 and
Cthe_3021 and Cthe_3022 and Cthe_3023 and Cthe_3024

old:  R_GLNt2r    L-glutaminereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     gln-l[e] + h[e] <==> gln-l[c] +
h[c]  None
new:  R_GLNt2r    L-glutaminereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     gln-l[e] + h[e] <==> gln-l[c] +
h[c]  None

old:  R_TRPt2r    L-tryptophanreversibletransportviaprotonsymport True
      Transport, Extracellular      .     h[e] + trp-l[e] <==> h[c] + trp-
l[c]  None
new:  R_TRPt2r    L-tryptophanreversibletransportviaprotonsymport True
      Transport, Extracellular      .     h[e] + trp-l[e] <==> h[c] + trp-
l[c]  None

old:  R_THRt2r    L-threoninereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     h[e] + thr-l[e] <==> h[c] + thr-
l[c]  None
new:  R_THRt2r    L-threoninereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     h[e] + thr-l[e] <==> h[c] + thr-
l[c]  None

old:  R_ASPt2r    L-aspartatereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     asp-l[e] + h[e] <==> asp-l[c] +
h[c]  None
new:  R_ASPt2r    L-aspartatereversibletransportviaprotonsymport  True
      Transport, Extracellular      .     asp-l[e] + h[e] <==> asp-l[c] +
h[c]  None

old:  R_PYRt2r    pyruvatereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + pyr[e] <==> h[c] + pyr[c]
      None
new:  R_PYRt2r    pyruvatereversibletransportviaprotonsymport     True
      Transport, Extracellular      .     h[e] + pyr[e] <==> h[c] + pyr[c]
      None

old:  R_GLYt2r    glycinereversibletransportviaprotonsymport      True
      Transport, Extracellular      .     gly[e] + h[e] <==> gly[c] + h[c]
      None
```

المنارة للاستشارات
www.manaraa.com

```
new:  R_GLYt2r    glycinereversibletransportviaprotonsymport      True
      Transport, Extracellular      .     gly[e] + h[e] <==> gly[c] + h[c]
      None

old:  R_BIOMASS   Total biomass flux including extracellular protein    False
      Biomass     .     [c] : cellmass --> biomass    None
new:  R_BIOMASS   Total biomass flux including extracellular protein    False
      Biomass     .     cellmass[c] + 1.3 cellome[e] --> biomass[c]     None

old:  R_ALAt2r    L-alaninereversibletransportviaprotonsymport    True
      Transport, Extracellular      .     ala-l[e] + h[e] <==> ala-l[c] +
h[c]  None
new:  R_ALAt2r    L-alaninereversibletransportviaprotonsymport    True
      Transport, Extracellular      .     ala-l[e] + h[e] <==> ala-l[c] +
h[c]  None

old:  R_G5SADr    L-glutamate5-semialdehydedehydratase,reversible True
      Arginine and Proline Metabolism     .     [c] : glu5sa <==> 1pyr5c + h
+ h2o spontaneous
new:  R_G5SADr    L-glutamate5-semialdehydedehydratase,reversible True
      Arginine and Proline Metabolism     .     [c] : glu5sa <==> 1pyr5c + h
+ h2o None
```

المنارة للاستشارات
www.manaraa.com

APPENDIX 2: DATA

A. Single-gene knockouts affecting ethanol production.

Gene	Reactions affected	Growth	Ethanol	EtOH LB	EtOH UB
Cthe_3003	R_FDXHASE	0.2648	25.9697	25.24921505	26.63161505
Cthe_2431	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_2430	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_2433	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_2432	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_2435	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_2434	R_NFO_Na,R_NFO_h	0.2920	22.7121	22.48501815	23.40661814
Cthe_1029	R_PTA2,R_PTAr	0.2687	14.2586	11.81206001	15.49846005
Cthe_1028	R_ACKr,R_PPAKr	0.2687	14.2586	11.81206001	15.49846005

B. RNAseq query processing

1. Pre-processing summary

```
Blast File: cth_08.31.2009.out
Annotation File: cth_annotation_IMG.txt
Minimum Length: 30
Maximum e val: 0.001
Minimum agreement: 95.0

Overlaps     0      0
High e val   0      0
Short reads 17      0.007398639
Hits   128114       55.7570113
Multiple Identical Alignments 36366 15.82699372
Low agreement       2512   1.093257664
Query Not Matched to Gene Locus      62521 27.21001689
Total queries       229772      100
not ctherm   242    0.105321797
```

2. RNAseq scores

Gene locus	Strand	Description	Gene Length	Unique Reads	Normalized Hit Count
Cthe_2348	-	Ig-related protein	3110	10751	3.456913183
Cthe_3078	+	cellulosome anchoring protein, cohesin region	6941	10580	1.524276041
Cthe_0423	+	iron-containing alcohol dehydrogenase(EC:1.2.1.10)	2621	3768	1.437619229
Cthe_3077	+	cellulosome anchoring protein, cohesin region	5561	7082	1.273511958
Cthe_0426	+	putative PAS/PAC sensor protein	1670	1930	1.155688623

www.manaraa.com

Cthe_1020	-	extracellular solute-binding protein, family 1	1379	1532	1.110949964
Cthe_0424	+	aminoglycoside phosphotransferase	749	824	1.100133511
Cthe_0427	+	serine phosphatase	1169	991	0.847733105
Cthe_0429	+	NADH dehydrogenase (quinone)(EC:1.6.99.5)	1874	1558	0.831376734
Cthe_2909	+	ribosomal protein S3	680	523	0.769117647
Cthe_2918	+	ribosomal protein L6	551	416	0.754990926
Cthe_2993	+	hypothetical protein	1046	736	0.703632887
Cthe_0422	+	CoA-binding protein	683	471	0.689604685
Cthe_2932	+	DNA-directed RNA polymerase, alpha subunit(EC:2.7.7.6)	947	606	0.639915523
Cthe_2906	+	ribosomal protein L2	827	473	0.571946796
Cthe_2729	+	translation elongation factor G	2093	1153	0.550883899
Cthe_1964	-	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	1529	820	0.536298234
Cthe_3079	+	cellulosome anchoring protein, cohesin region	2066	1098	0.531461762
Cthe_2931	+	ribosomal protein S4	626	323	0.515974441
Cthe_0505	+	formate acetyltransferase(EC:2.3.1.54)	2228	1115	0.500448833
Cthe_0412	+	glycoside hydrolase, family 9	2687	1294	0.481577968
Cthe_2724	+	DNA-directed RNA polymerase, beta subunit(EC:2.7.7.6)	3752	1725	0.459754797
Cthe_0389	+	PfkB	971	428	0.440782698
Cthe_1091	-	metal dependent phosphohydrolase	1574	649	0.412325286
Cthe_0390	+	ROK domain containing protein	1214	489	0.402800659
Cthe_2925	+	methionine aminopeptidase, type I(EC:3.4.11.18)	770	305	0.396103896
Cthe_0430	+	hydrogenase, Fe-only	1700	668	0.392941176
Cthe_3091	+	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, beta subunit	557	196	0.351885099
Cthe_1090	-	hypothetical protein	782	274	0.350383632
Cthe_2392	+	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	1184	412	0.347972973
Cthe_2904	+	ribosomal protein L4/L1e	626	204	0.325878594
Cthe_2924	+	adenylate kinases(EC:2.7.4.3)	653	211	0.323124043
Cthe_2522	+	membrane associated protein	1622	519	0.319975339
Cthe_2903	+	ribosomal protein L3	638	198	0.310344828
Cthe_2923	+	preprotein translocase, SecY subunit	1283	387	0.301636789
Cthe_0388	+	Alcohol dehydrogenase GroES-like protein	1034	308	0.297872340
Cthe_2722	+	ribosomal protein L10	536	159	0.296641791
Cthe_2253	-	ATP-dependent metalloprotease FtsH(EC:3.6.4.6)	1799	531	0.295163980
Cthe_2915	+	ribosomal protein L5	548	160	0.291970803
Cthe_2728	+	ribosomal protein S7	470	135	0.287234043
Cthe_2518	+	ketol-acid reductoisomerase(EC:1.1.1.86)	995	283	0.284422111
Cthe_2174	-	transcription termination factor Rho	1961	550	0.280469148
Cthe_0498	+	protein of unknown function DUF342	1628	449	0.275798526
Cthe_2908	+	ribosomal protein L22	377	92	0.244031830
Cthe_1104	-	prepilin-type cleavage/methylation	530	128	0.241509434
Cthe_2519	+	2-isopropylmalate synthase/homocitrate synthase family protein	1619	386	0.238418777
Cthe_2730	+	translation elongation factor Tu	1202	284	0.236272879

المنارة للاستشارات

www.manaraa.com

Cthe_1019	-	binding-protein-dependent transport systems inner membrane component	983	216	0.219735504
Cthe_2855	+	hypothetical protein	1238	272	0.219709208
Cthe_1205	-	putative serine protein kinase, PrkA	1928	420	0.217842324
Cthe_0720	+	aminotransferase, class V(EC:2.8.1.7)	1184	257	0.217060811
Cthe_0981	-	S-adenosyl-methyltransferase MraW	941	203	0.215727949
Cthe_2393	+	thiamine pyrophosphate enzyme-like TPP-binding	935	200	0.213903743
Cthe_2922	+	ribosomal protein L15	440	94	0.213636364
Cthe_1344	-	(p)ppGpp synthetase I, SpoT/RelA(EC:2.7.6.5)	2174	456	0.209751610
Cthe_0344	-	Malate dehydrogenase (oxaloacetate-decarboxylating)(EC:1.1.1.38)	1172	245	0.209044369
Cthe_2725	+	DNA-directed RNA polymerase, beta' subunit(EC:2.7.7.6)	3497	730	0.208750357
Cthe_0341	+	NADH dehydrogenase (quinone)(EC:1.6.99.5)	1793	371	0.206915784
Cthe_2093	-	hypothetical protein	1058	218	0.206049149
Cthe_1093	-	Methenyltetrahydrofolate cyclohydrolase(EC:3.5.4.9)	857	176	0.205367561
Cthe_0425	+	hypothetical protein	245	49	0.200000000
Cthe_2920	+	ribosomal protein S5-like protein	500	99	0.198000000
Cthe_1391	-	2-isopropylmalate synthase	1535	299	0.194788274
Cthe_2929	+	ribosomal protein S13	371	72	0.194070081
Cthe_3003	-	hydrogenase, Fe-only(EC:1.12.7.2)	1934	371	0.191830403
Cthe_2917	+	ribosomal protein S8	398	76	0.190954774
Cthe_2854	+	intein	1847	351	0.190037899
Cthe_2907	+	ribosomal protein S19	284	53	0.186619718
Cthe_0991	-	translation initiation factor IF-2	3107	577	0.185709688
Cthe_1347	-	GatB/YqeY	443	82	0.185101580
Cthe_2739	+	trigger factor	1286	238	0.185069984
Cthe_0577	+	thiamine pyrophosphokinase	638	117	0.183385580
Cthe_0814	-	DNA repair protein RecN	1712	300	0.175233645
Cthe_0722	+	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase(EC:2.1.1.61)	1079	189	0.175162187
Cthe_0428	+	NADH dehydrogenase (ubiquinone), 24 kDa subunit	497	87	0.175050302
Cthe_0506	+	pyruvate formate-lyase activating enzyme	716	125	0.174581006
Cthe_2919	+	ribosomal protein L18	368	63	0.171195652
Cthe_1228	-	threonyl-tRNA synthetase	1907	326	0.170949135
Cthe_2444	+	Transketolase-like protein	941	160	0.170031881
Cthe_0052	+	hypothetical protein	1259	214	0.169976172
Cthe_2446	+	ABC-type sugar transport system periplasmic component-like protein	1118	190	0.169946333
Cthe_1868	+	carbamoyl-phosphate synthase, large subunit	3203	541	0.168904152
Cthe_0349	+	fructose-1,6-bisphosphate aldolase, class II(EC:4.1.2.13)	929	152	0.163616792
Cthe_1963	-	glycoside hydrolase, family 10	2513	409	0.162753681
Cthe_2874	+	Phosphoenolpyruvate carboxykinase (GTP)(EC:4.1.1.32)	1817	293	0.161254816
Cthe_2443	+	Transketolase-like protein	836	132	0.157894737
Cthe_0866	-	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	1040	164	0.157692308

www.manaraa.com

Cthe_2447	+	ABC transporter related protein	1487	233	0.156691325
Cthe_0618	-	ferrous iron transport protein B	2165	338	0.156120092
Cthe_3080	+	cellulosome anchoring protein, cohesin region	1343	209	0.155621742
Cthe_2741	+	ATP-dependent Clp protease, ATP-binding subunit ClpX	1295	200	0.154440154
Cthe_2521	+	DNA-directed RNA polymerase sigma factor	707	109	0.154172560
Cthe_3120	+	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	3536	543	0.153563348
Cthe_2914	+	ribosomal protein L24	350	52	0.148571429
Cthe_0681	+	IMP dehydrogenase/GMP reductase	1493	221	0.148024113
Cthe_0342	+	hydrogenase, Fe-only	1748	257	0.147025172
Cthe_2713	+	dihydroxy-acid dehydratase(EC:4.2.1.9)	1664	242	0.145432692
Cthe_1504	+	Linocin_M18 bacteriocin protein	812	118	0.145320197
Cthe_0138	+	Phosphoglycerate kinase(EC:2.7.2.3)	1193	172	0.144174350
Cthe_2445	+	Alcohol dehydrogenase GroES-like protein	1043	149	0.142857143
Cthe_2210	-	3-isopropylmalate dehydratase, small subunit	569	81	0.142355009
Cthe_0140	+	phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent(EC:5.4.2.1)	1535	218	0.142019544
Cthe_2905	+	Ribosomal protein L25/L23	353	50	0.141643059
Cthe_1185	-	hypothetical protein	1280	181	0.141406250
Cthe_3004	-	ferredoxin	1481	207	0.139770425
Cthe_1960	-	Peptidoglycan-binding domain 1	1082	148	0.136783734
Cthe_0152	+	aldo/keto reductase	947	127	0.134107709
Cthe_2714	+	acetolactate synthase, large subunit, biosynthetic type(EC:2.2.1.6)	1667	221	0.132573485
Cthe_1094	-	Radical SAM-like protein	1919	254	0.132360604
Cthe_2994	+	hypothetical protein	2024	266	0.131422925
Cthe_2913	+	ribosomal protein L14	368	48	0.130434783
Cthe_2740	+	ATP-dependent Clp protease, proteolytic subunit ClpP(EC:3.4.21.92)	584	75	0.128424658
Cthe_1005	-	translation elongation factor Ts	647	83	0.128284389
Cthe_0285	+	isocitrate dehydrogenase, NADP-dependent(EC:1.1.1.42)	1208	152	0.125827815
Cthe_1867	+	carbamoyl-phosphate synthase, small subunit(EC:6.3.5.5)	1064	133	0.125000000
Cthe_2727	+	ribosomal protein S12	428	53	0.123831776
Cthe_2930	+	ribosomal protein S11	404	50	0.123762376
Cthe_0701	+	conserved carboxylase region containing protein	1397	172	0.123120974
Cthe_0472	+	flagellar hook capping protein	971	119	0.122554068
Cthe_3122	+	S-layer-like domain containing protein	2078	253	0.121751684
Cthe_2442	+	carbohydrate kinase, FGGY(EC:2.7.1.30)	1484	179	0.120619946
Cthe_0143	+	Phosphopyruvate hydratase(EC:4.2.1.11)	1301	156	0.119907763
Cthe_2211	-	3-isopropylmalate dehydratase, large subunit(EC:4.2.1.33)	1259	148	0.117553614
Cthe_0163	+	GTP1/OBG subdomain containing protein	1274	149	0.116954474
Cthe_3093	+	Adenylosuccinate synthase(EC:6.3.4.4)	1274	148	0.116169545
Cthe_1557	-	ABC transporter related protein	821	95	0.115712546
Cthe_0699	+	carboxyl transferase(EC:6.4.1.3)	1550	177	0.114193548
Cthe_0615	+	Phenylacetate--CoA ligase(EC:6.2.1.30)	1301	148	0.113758647

المنارة للاستشارات

www.manaraa.com

Cthe_1204	-	protein of unknown function DUF444	1268	143	0.112776025
Cthe_2933	+	ribosomal protein L17	524	58	0.110687023
Cthe_2992	+	RNA polymerase, sigma-24 subunit, ECF subfamily	524	58	0.110687023
Cthe_2705	+	Transketolase, central region	935	103	0.110160428
Cthe_0989	-	phosphoesterase, RecJ-like protein	956	105	0.109832636
Cthe_1866	+	acetylornithine and succinylornithine aminotransferases(EC:2.6.1.11)	1202	132	0.109816972
Cthe_3035	-	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	1175	128	0.108936170
Cthe_1555	-	NLPA lipoprotein	884	95	0.107466063
Cthe_2390	+	pyruvate/ketoisovalerate oxidoreductase, gamma subunit(EC:1.2.7.1)	578	62	0.107266436
Cthe_0994	-	NusA antitermination factor	1214	130	0.107084020
Cthe_2737	+	excinuclease ABC, C subunit	1877	198	0.105487480
Cthe_1092	-	hypothetical protein	677	71	0.104874446
Cthe_0214	+	phenylalanyl-tRNA synthetase, alpha subunit(EC:6.1.1.20)	1019	106	0.104023553
Cthe_0501	+	hypothetical protein	629	65	0.103338633
Cthe_0053	+	ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent	2384	246	0.103187919
Cthe_1329	-	putative CoA-substrate-specific enzyme activase	4301	439	0.102069286
Cthe_0146	+	ribonuclease R(EC:3.1.13.1)	2273	232	0.102067752
Cthe_0863	+	glutamine synthetase, catalytic region	2117	216	0.102031176
Cthe_0215	+	phenylalanyl-tRNA synthetase, beta subunit	2387	242	0.101382488
Cthe_2882	+	Histidinol dehydrogenase(EC:1.1.1.23)	1307	132	0.100994644
Cthe_2265	+	H+-transporting two-sector ATPase, C (AC39) subunit	1019	102	0.100098135
Cthe_3157	+	pyruvate carboxyltransferase	1379	138	0.100072516
Cthe_0768	-	nucleic acid binding protein, containing KH domain	230	23	0.100000000
Cthe_2516	+	acetolactate synthase, large subunit, biosynthetic type(EC:2.2.1.6)	1628	162	0.099508600
Cthe_1345	-	adenine phosphoribosyltransferase(EC:2.4.2.7)	515	51	0.099029126
Cthe_0574	+	serine/threonine protein kinase	2111	209	0.099005211
Cthe_2078	-	protein of unknown function UPF0029	659	65	0.098634294
Cthe_1346	-	single-stranded-DNA-specific exonuclease RecJ	2483	243	0.097865485
Cthe_2517	+	acetolactate synthase, small subunit	512	50	0.097656250
Cthe_1385	-	preprotein translocase, SecA subunit	2732	265	0.096998536
Cthe_1797	+	3-phosphoshikimate 1-carboxyvinyltransferase	1271	122	0.095987411
Cthe_1245	-	phosphoribosylamine--glycine ligase(EC:6.3.4.13)	1259	119	0.094519460
Cthe_2598	+	uracil phosphoribosyltransferase	629	59	0.093799682
Cthe_2355	-	L-aspartate oxidase(EC:1.4.3.16)	1598	149	0.093241552
Cthe_2065	-	cysteinyl-tRNA synthetase(EC:6.1.1.16)	1406	130	0.092460882
Cthe_0155	+	hypothetical protein	1061	98	0.092365693
Cthe_0567	+	peptide deformylase(EC:3.5.1.88)	512	47	0.091796875
Cthe_1485	-	beta-lactamase-like protein	1253	115	0.091779729
Cthe_1317	-	hypothetical protein	884	81	0.091628959

المنارة للاستشارات

www.manaraa.com

Cthe_1261	-	6-phosphofructokinase(EC:2.7.1.11)	974	89	0.091375770
Cthe_0106	+	GTP cyclohydrolase II(EC:3.5.4.25)	1241	113	0.091055600
Cthe_0555	+	PpiC-type peptidyl-prolyl cis-trans isomerase	1241	112	0.090249799
Cthe_0162	+	ribosomal protein L27	278	25	0.089928058
Cthe_0105	+	riboflavin synthase, alpha subunit	656	58	0.088414634
Cthe_0732	+	Chorismate synthase(EC:4.2.3.5)	1184	104	0.087837838
Cthe_2606	+	ATP synthase F1, alpha subunit(EC:3.6.3.15)	1520	133	0.087500000
Cthe_2267	+	Sodium-transporting two-sector ATPase(EC:3.6.3.15)	1769	154	0.087054833
Cthe_0345	-	L-lactate dehydrogenase(EC:1.1.1.27)	956	83	0.086820084
Cthe_1163	-	phosphoglucosamine mutase(EC:5.4.2.10)	1349	117	0.086730912
Cthe_0578	-	glycoside hydrolase, family 9	2210	189	0.085520362
Cthe_0269	+	glycoside hydrolase, family 8	1433	121	0.084438241
Cthe_0490	+	CheA signal transduction histidine kinases	2078	174	0.083734360
Cthe_3200	+	alanyl-tRNA synthetase(EC:6.1.1.7)	2642	221	0.083648751
Cthe_0963	+	Dihydrodipicolinate reductase(EC:1.3.1.26)	758	63	0.083113456
Cthe_1006	-	ribosomal protein S2	758	63	0.083113456
Cthe_1101	-	hypothetical protein	1292	107	0.082817337
Cthe_0154	+	protein of unknown function DUF1292	290	24	0.082758621
Cthe_2366	-	single-stranded nucleic acid binding R3H	620	51	0.082258065
Cthe_0418	+	Polyribonucleotide nucleotidyltransferase(EC:2.7.7.8)	2102	170	0.080875357
Cthe_0139	+	Triose-phosphate isomerase(EC:5.3.1.1)	755	61	0.080794702
Cthe_1945	-	Thioredoxin-disulfide reductase(EC:1.8.1.9)	893	72	0.080627100
Cthe_0500	+	hypothetical protein	584	47	0.080479452
Cthe_0962	+	dihydrodipicolinate synthase(EC:4.2.1.52)	878	70	0.079726651
Cthe_2723	+	ribosomal protein L7/L12	389	31	0.079691517
Cthe_0934	-	3-oxoacyl-(acyl-carrier-protein) reductase(EC:1.1.1.100)	743	59	0.079407806
Cthe_1248	-	phosphoribosylformylglycinamidine cyclo-ligase(EC:6.3.3.1)	1022	81	0.079256360
Cthe_0993	-	protein of unknown function DUF448	278	22	0.079136691
Cthe_1789	-	ATPase AAA-2	2456	194	0.078990228
Cthe_0721	+	nitrogen-fixing NifU-like protein	449	35	0.077951002
Cthe_3086	+	3-phosphoshikimate 1-carboxyvinyltransferase	1271	99	0.077891424
Cthe_3158	+	putative aconitate hydratase	1928	150	0.077800830
Cthe_1249	-	amidophosphoribosyltransferase(EC:2.4.2.14)	1466	114	0.077762619
Cthe_0865	-	3-methyl-2-oxobutanoate dehydrogenase (ferredoxin)(EC:1.2.7.7)	746	58	0.077747989
Cthe_0236	+	aldo/keto reductase	1130	87	0.076991150
Cthe_1796	+	Prephenate dehydrogenase	1097	84	0.076572470
Cthe_0471	+	flagellar hook-length control protein	1586	121	0.076292560
Cthe_2736	+	phosphoenolpyruvate-protein phosphotransferase	1706	130	0.076201641
Cthe_3065	+	hypothetical protein	1241	94	0.075745367
Cthe_0958	-	queuine tRNA-ribosyltransferase(EC:2.4.2.29)	1124	85	0.075622776
Cthe_2619	+	cell shape determining protein, MreB/Mrl family	1034	78	0.075435203
Cthe_2209	-	3-isopropylmalate dehydrogenase(EC:1.1.1.85)	1088	81	0.074448529
Cthe_1003	-	ribosome recycling factor	551	41	0.074410163

www.manaraa.com

Cthe_1795	+	phospho-2-dehydro-3-deoxyheptonate aldolase(EC:2.5.1.54)	1025	76	0.074146341
Cthe_0573	+	protein serine/threonine phosphatases	737	54	0.073270014
Cthe_0017	+	protein of unknown function DUF1538	1862	136	0.073039742
Cthe_0988	-	tRNA pseudouridine synthase B(EC:4.2.1.70)	893	65	0.072788354
Cthe_1246	-	phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase(EC:2.1.2.3)	1544	112	0.072538860
Cthe_0816	-	NAD(+) kinase(EC:2.7.1.23)	869	63	0.072497123
Cthe_1777	-	amidohydrolase	1202	87	0.072379368
Cthe_2962	-	oligopeptide/dipeptide ABC transporter, ATPase subunit	1028	74	0.071984436
Cthe_0101	+	iron-containing alcohol dehydrogenase	1169	84	0.071856287
Cthe_2708	+	hypothetical protein	1421	102	0.071780436
Cthe_0935	-	malonyl CoA-acyl carrier protein transacylase	935	67	0.071657754
Cthe_2825	+	intein	1847	130	0.070384407
Cthe_0158	+	ribonuclease, Rne/Rng family	1493	105	0.070328198
Cthe_2255	-	tRNA(Ile)-lysidine synthetase	1412	99	0.070113314
Cthe_0765	-	ribosomal protein L19	344	24	0.069767442
Cthe_3096	-	hypothetical protein	935	65	0.069518717
Cthe_2721	+	ribosomal protein L1	695	48	0.069064748
Cthe_3228	+	copper amine oxidase-like protein	800	55	0.068750000
Cthe_0766	-	tRNA (guanine-N1)-methyltransferase(EC:2.1.1.31)	743	51	0.068640646
Cthe_1107	-	type II secretion system protein E	2363	162	0.068556919
Cthe_1265	-	phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I	1736	119	0.068548387
Cthe_2585	+	biotin--acetyl-CoA-carboxylase ligase	1016	69	0.067913386
Cthe_2607	+	ATP synthase F1, gamma subunit(EC:3.6.3.15)	887	60	0.067643743
Cthe_0613	+	thiamine pyrophosphate enzyme-like TPP-binding	1805	121	0.067036011
Cthe_2709	+	hypothetical protein	1478	99	0.066982409
Cthe_0137	+	glyceraldehyde-3-phosphate dehydrogenase, type I(EC:1.2.1.12)	1010	67	0.066336634
Cthe_2926	+	hypothetical protein	272	18	0.066176471
Cthe_0375	+	GMP synthase, large subunit	1535	101	0.065798046
Cthe_0063	+	Rubredoxin-type Fe(Cys)4 protein	335	22	0.065671642
Cthe_0324	+	valyl-tRNA synthetase(EC:6.1.1.9)	2654	172	0.064807837
Cthe_1862	-	ABC transporter related protein	1112	72	0.064748201
Cthe_2593	+	peptide chain release factor 1	1082	70	0.064695009
Cthe_1028	-	acetate kinase(EC:2.7.2.1)	1199	77	0.064220183
Cthe_0347	-	phosphofructokinase	1247	80	0.064153970
Cthe_0617	+	AAA ATPase, central region	1298	83	0.063944530
Cthe_0093	+	septum site-determining protein MinD	800	51	0.063750000
Cthe_1018	-	binding-protein-dependent transport systems inner membrane component	869	55	0.063291139
Cthe_0904	-	protein-export membrane protein SecD	1337	84	0.062827225
Cthe_0410	+	protein of unknown function DUF711	1358	85	0.062592047
Cthe_0234	+	AMP-dependent synthetase and ligase	2531	158	0.062425919
Cthe_2912	+	ribosomal protein S17	257	16	0.062256809

www.manaraa.com

Cthe_0961	+	aspartate-semialdehyde dehydrogenase(EC:1.2.1.11)	998	62	0.062124248
Cthe_2256	-	replicative DNA helicase	1337	83	0.062079282
Cthe_1050	-	recA protein	1049	65	0.061963775
Cthe_0287	+	multi-sensor hybrid histidine kinase	3482	215	0.061746123
Cthe_0896	-	DNA primase	1802	111	0.061598224
Cthe_1024	-	small GTP-binding protein	1322	81	0.061270802
Cthe_0673	+	ribonuclease H	443	27	0.060948081
Cthe_0937	-	fatty acid/phospholipid synthesis protein PlsX	1019	62	0.060843965
Cthe_2268	+	Sodium-transporting two-sector ATPase(EC:3.6.3.15)	1382	84	0.060781476
Cthe_1965	-	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	563	34	0.060390764
Cthe_2961	-	extracellular solute-binding protein, family 5	1625	98	0.060307692
Cthe_2934	+	ABC transporter related protein	899	54	0.060066741
Cthe_2629	+	UDP-N-acetylglucosamine pyrophosphorylase(EC:2.3.1.157)	1403	84	0.059871703
Cthe_0911	-	protein of unknown function DUF152	839	50	0.059594756
Cthe_2325	-	Recombinase	2093	124	0.059245103
Cthe_2888	+	imidazoleglycerol phosphate synthase, cyclase subunit	761	45	0.059132720
Cthe_1412	-	tryptophan synthase, beta subunit(EC:4.2.1.20)	1184	70	0.059121622
Cthe_0504	+	Putative enzyme of poly-gamma-glutamate biosynthesis (capsule formation)-like protein	1235	73	0.059109312
Cthe_0089	+	rod shape-determining protein MreC	863	51	0.059096176
Cthe_0711	+	chorismate mutase(EC:5.4.99.5)	356	21	0.058988764
Cthe_1767	-	hypothetical protein	1001	59	0.058941059
Cthe_2264	+	H+-transporting two-sector ATPase, E subunit	596	35	0.058724832
Cthe_1352	-	UDP-glucose 6-dehydrogenase(EC:1.1.1.22)	1331	77	0.057851240
Cthe_2591	+	hypothetical protein	416	24	0.057692308
Cthe_1911	-	Carbohydrate binding family 6	3872	223	0.057592975
Cthe_0712	+	cytidylate kinase(EC:2.7.4.14)	683	39	0.057101025
Cthe_0932	-	beta-ketoacyl synthase(EC:2.3.1.41)	1235	70	0.056680162
Cthe_2826	+	hypothetical protein	1238	70	0.056542811
Cthe_0646	+	anaerobic ribonucleoside-triphosphate reductase	2090	118	0.056459330
Cthe_1556	-	binding-protein-dependent transport systems inner membrane component	710	40	0.056338028
Cthe_2889	+	phosphoribosyl-AMP cyclohydrolase	662	37	0.055891239
Cthe_1790	-	ATP:guanido phosphotransferase	1022	57	0.055772994
Cthe_2597	+	sugar-phosphate isomerases, RpiB/LacA/LacB family(EC:5.3.1.26)	449	25	0.055679287
Cthe_1942	-	hypothetical protein	899	50	0.055617353
Cthe_2376	+	DNA gyrase, B subunit(EC:5.99.1.3)	1925	107	0.055584416
Cthe_0903	-	protein-export membrane protein SecF	956	53	0.055439331
Cthe_2927	+	translation initiation factor IF-1	218	12	0.055045872
Cthe_0309	+	excinuclease ABC, B subunit	1982	109	0.054994955
Cthe_2910	+	ribosomal protein L16	437	24	0.054919908
Cthe_1917	-	ATPase, P-type (transporting), HAD superfamily, subfamily IC	2717	149	0.054839897

المنارة للاستشارات
www.manaraa.com

Cthe_0104	+	riboflavin biosynthesis protein RibD(EC:1.1.1.193,EC:3.5.4.26)	1097	60	0.054694622
Cthe_0187	+	hypothetical protein	1172	64	0.054607509
Cthe_1099	-	hypothetical protein	2549	139	0.054531189
Cthe_2704	+	Transketolase-like protein	845	46	0.054437870
Cthe_2311	-	hypothetical protein	719	39	0.054242003
Cthe_0833	-	exodeoxyribonuclease VII, large subunit(EC:3.1.11.6)	1235	66	0.053441296
Cthe_3092	+	conserved hypothetical protein, CF-27 family	206	11	0.053398058
Cthe_1203	-	SpoVR	1388	74	0.053314121
Cthe_1025	-	protein of unknown function DUF512	1337	71	0.053103964
Cthe_0474	+	protein of unknown function DUF1078-like protein	1322	70	0.052950076
Cthe_2208	+	LexA DNA-binding region containing protein	587	31	0.052810903
Cthe_2838	+	intein	1841	97	0.052688756
Cthe_1312	-	glycyl-tRNA synthetase(EC:6.1.1.14)	1388	73	0.052593660
Cthe_0877	+	GTP-binding protein YchF	1094	57	0.052102377
Cthe_1768	-	NifU-related domain containing protein	692	36	0.052023121
Cthe_0571	+	sun protein	1367	71	0.051938552
Cthe_0714	+	hydroxymethylbutenyl pyrophosphate reductase	2084	108	0.051823417
Cthe_0156	+	Radical SAM	1841	95	0.051602390
Cthe_1946	-	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	1283	66	0.051441933
Cthe_2706	+	ABC transporter related protein	1109	57	0.051397656
Cthe_0793	-	hypothetical protein	1268	65	0.051261830
Cthe_2278	+	extracellular solute-binding protein, family 3	863	44	0.050984936
Cthe_0868	-	type IV pilus assembly PilZ	710	36	0.050704225
Cthe_0016	+	Ferritin and Dps	479	24	0.050104384
Cthe_1384	-	FolC bifunctional protein	1286	64	0.049766719
Cthe_1368	-	S-layer-like domain containing protein	2132	106	0.049718574
Cthe_0071	-	Cellulose 1,4-beta-cellobiosidase(EC:3.2.1.91)	2816	140	0.049715909
Cthe_2089	-	glycoside hydrolase, family 48	2225	110	0.049438202
Cthe_3094	-	glycosyl transferase, family 2	974	48	0.049281314
Cthe_1433	-	short-chain dehydrogenase/reductase SDR	773	38	0.049159120
Cthe_2947	+	prolyl-tRNA synthetase	1718	84	0.048894063
Cthe_0614	+	pyruvate ferredoxin/flavodoxin oxidoreductase	575	28	0.048695652
Cthe_1331	-	aspartyl-tRNA synthetase	1787	87	0.048684947
Cthe_0217	+	Glucose-6-phosphate isomerase(EC:5.3.1.9)	1346	65	0.048291233
Cthe_2066	-	serine O-acetyltransferase(EC:2.3.1.30)	746	36	0.048257373
Cthe_1762	-	efflux transporter, RND family, MFP subunit	1265	61	0.048221344
Cthe_0917	-	glutaminyl-tRNA synthetase(EC:6.1.1.18)	1742	84	0.048220436
Cthe_0723	+	tyrosyl-tRNA synthetase(EC:6.1.1.1)	1226	59	0.048123980
Cthe_2608	+	ATP synthase F1, beta subunit(EC:3.6.3.15)	1394	67	0.048063128
Cthe_1237	-	leucyl-tRNA synthetase	2477	119	0.048041986
Cthe_2367	-	60 kDa inner membrane insertion protein	875	42	0.048000000
Cthe_1095	-	cell divisionFtsK/SpoIIIE	2426	116	0.047815334
Cthe_0361	+	hypothetical protein	566	27	0.047703180
Cthe_1787	-	glycoside hydrolase 15-related	1934	92	0.047569804

المنارة للاستشارات
www.manaraa.com

Cthe_2633	+	hypothetical protein	1094	52	0.047531993
Cthe_2183	-	UTP-glucose-1-phosphate uridylyltransferase(EC:2.7.7.9)	869	41	0.047180667
Cthe_1801	-	ABC transporter related protein	764	36	0.047120419
Cthe_1286	-	peptidase S1 and S6, chymotrypsin/Hap	1529	72	0.047089601
Cthe_0460	+	DNA topoisomerase I(EC:5.99.1.2)	2108	99	0.046963947
Cthe_0419	+	peptidase M16-like protein	1259	59	0.046862589
Cthe_0957	-	preprotein translocase, YajC subunit	299	14	0.046822742
Cthe_2448	+	inner-membrane translocator	983	46	0.046795524
Cthe_0654	+	biotin and thiamin synthesis associated protein	1376	64	0.046511628
Cthe_2883	+	histidinol-phosphate aminotransferase	1076	50	0.046468401
Cthe_0982	-	MraZ protein	431	20	0.046403712
Cthe_0997	-	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase(EC:1.17.4.3)	1058	49	0.046313800
Cthe_2938	+	putative glucokinase, ROK family	953	44	0.046169990
Cthe_2707	+	ABC-type transport system involved in multi-copper enzyme maturation, permease component	722	33	0.045706371
Cthe_2407	+	AAA ATPase, central region	1622	74	0.045622688
Cthe_1869	+	ornithine carbamoyltransferase(EC:2.1.3.3)	923	42	0.045503792
Cthe_1211	-	pyridoxal-phosphate dependent TrpB-like enzyme	1364	62	0.045454545
Cthe_2372	+	DNA polymerase III, beta subunit(EC:2.7.7.7)	1100	50	0.045454545
Cthe_0551	-	AMP-dependent synthetase and ligase	1664	75	0.045072115
Cthe_0762	-	protein of unknown function DUF1121	653	29	0.044410413
Cthe_1839	+	Radical SAM	1061	47	0.044297832
Cthe_1103	-	prepilin-type cleavage/methylation	566	25	0.044169611
Cthe_0157	+	hypothetical protein	839	37	0.044100119
Cthe_3164	+	two component transcriptional regulator, AraC family	1634	72	0.044063647
Cthe_0391	+	ABC transporter related protein	1499	66	0.044029353
Cthe_0831	-	Polyprenyl synthetase(EC:2.5.1.10)	887	39	0.043968433
Cthe_0998	-	putative membrane-associated zinc metalloprotease	1274	56	0.043956044
Cthe_1873	-	HMG-I and HMG-Y, DNA-binding	1343	59	0.043931497
Cthe_2860	+		1710	75	0.043859649
Cthe_2279	+	polar amino acid ABC transporter, inner membrane subunit	710	31	0.043661972
Cthe_3190	+	helicase, RecD/TraA family(EC:3.1.11.5)	2132	93	0.043621013
Cthe_0821	-	coagulation factor 5/8 type-like protein	1676	73	0.043556086
Cthe_1783	-	ribosomal protein L13	437	19	0.043478261
Cthe_2404	+	transcriptional regulator, GntR family	692	30	0.043352601
Cthe_2964	-	binding-protein-dependent transport systems inner membrane component	1085	47	0.043317972
Cthe_2916	+	ribosomal protein S14	185	8	0.043243243
Cthe_2586	+	amidohydrolase 2	788	34	0.043147208
Cthe_0927	-	chromosome segregation protein SMC	3572	154	0.043113102
Cthe_1912	-	copper amine oxidase-like protein	1607	69	0.042937150
Cthe_0791	-	protein of unknown function DUF552	467	20	0.042826552
Cthe_0763	-	GTP-binding protein	875	37	0.042285714

المنارة للاستشارات
www.manaraa.com

Cthe_0230	+	hypothetical protein	1706	72	0.042203986
Cthe_1022	-	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))(EC:1.1.1.94)	1019	43	0.042198234
Cthe_0178	-	argininosuccinate lyase(EC:4.3.2.1)	1376	58	0.042151163
Cthe_0206	+	UvrD/REP helicase	2231	94	0.042133572
Cthe_2354	-	nicotinate-nucleotide pyrophosphorylase(EC:2.4.2.19)	833	35	0.042016807
Cthe_3230	+	hypothetical protein	2888	121	0.041897507
Cthe_2897	+	transcription elongation factor GreA	479	20	0.041753653
Cthe_0060	+	SEC-C motif containing protein	503	21	0.041749503
Cthe_1765	-	hypothetical protein	1151	48	0.041702867
Cthe_0915	+	hypothetical protein	1202	50	0.041597338
Cthe_1332	-	histidyl-tRNA synthetase(EC:6.1.1.21)	1253	52	0.041500399
Cthe_1838	-	glycoside hydrolase, family 10	1859	77	0.041420118
Cthe_2696	+	carbohydrate kinase, YjeF related protein	1547	64	0.041370394
Cthe_0946	-	Phosphoglycerate mutase	701	29	0.041369472
Cthe_0607	+	peptidase M42	1049	43	0.040991420
Cthe_2686	+	type IV pilus assembly protein PilM	1103	45	0.040797824
Cthe_0183	+	protein of unknown function DUF208	545	22	0.040366972
Cthe_2963	-	oligopeptide/dipeptide ABC transporter, ATPase subunit	1016	41	0.040354331
Cthe_1845	-	Homoserine O-succinyltransferase(EC:2.3.1.46)	917	37	0.040348964
Cthe_0951	-	orotidine 5'-phosphate decarboxylase	944	38	0.040254237
Cthe_0239	+	cellulosome enzyme, dockerin type I	3155	127	0.040253566
Cthe_3159	+	transcriptional regulator, GntR family	671	27	0.040238450
Cthe_2333	-	two component transcriptional regulator, winged helix family	698	28	0.040114613
Cthe_0787	-	isoleucyl-tRNA synthetase(EC:6.1.1.5)	2795	112	0.040071556
Cthe_2815	-	lysyl-tRNA synthetase(EC:6.1.1.6)	1601	64	0.039975016
Cthe_1371	-	YD repeat protein	5879	235	0.039972784
Cthe_1398	-	cellulosome enzyme, dockerin type I	2528	101	0.039952532
Cthe_1979	-	intein	1352	54	0.039940828
Cthe_1432	-	transposase IS200-like protein	476	19	0.039915966
Cthe_1872	-	Tn7-like transposition protein C	1655	66	0.039879154
Cthe_2090	-	hypothetical protein	578	23	0.039792388
Cthe_3188	+	copper amine oxidase-like protein	1487	59	0.039677202
Cthe_1807	-	hypothetical protein	1664	66	0.039663462
Cthe_1935	-	arginyl-tRNA synthetase(EC:6.1.1.19)	1694	67	0.039551358
Cthe_1070	-	metal dependent phosphohydrolase	2201	87	0.039527488
Cthe_3229	+	transposase, IS4	1217	48	0.039441249
Cthe_0931	-	Ribonuclease III(EC:3.1.26.3)	710	28	0.039436620
Cthe_2719	+	NusG antitermination factor	533	21	0.039399625
Cthe_0880	+	phospho-2-dehydro-3-deoxyheptonate aldolase(EC:2.5.1.54)	1016	40	0.039370079
Cthe_2834	+	intein	1685	66	0.039169139
Cthe_0869	-	hypothetical protein	668	26	0.038922156
Cthe_1761	-	protein of unknown function DUF214	1208	47	0.038907285
Cthe_3189	+	hypothetical protein	617	24	0.038897893

المنارة للاستشارات
www.manaraa.com

Cthe_1773	-	peptidase S16, lon-like protein	2444	95	0.038870704
Cthe_0568	+	methionyl-tRNA formyltransferase(EC:2.1.2.9)	935	36	0.038502674
Cthe_2075	-	protein of unknown function DUF28	728	28	0.038461538
Cthe_2258	-	phosphoesterase, RecJ-like protein	2009	77	0.038327526
Cthe_3165	+	PpiC-type peptidyl-prolyl cis-trans isomerase	1148	44	0.038327526
Cthe_0674	+	NUDIX hydrolase	548	21	0.038321168
Cthe_2109	-	copper amine oxidase-like protein	2537	97	0.038234135
Cthe_0338	+	NADH-quinone oxidoreductase, E subunit	497	19	0.038229376
Cthe_0936	-	3-oxoacyl-(acyl-carrier-protein) synthase III(EC:2.3.1.41)	995	38	0.038190955
Cthe_0069	+	Aspartate--ammonia ligase(EC:6.3.1.1)	1022	39	0.038160470
Cthe_2506	+	S-layer-like domain containing protein	3041	116	0.038145347
Cthe_1484	-	transcriptional regulator	710	27	0.038028169
Cthe_2887	+	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase(EC:5.3.1.16)	713	27	0.037868163
Cthe_0286	+	response regulator receiver sensor signal transduction histidine kinase	1274	48	0.037676609
Cthe_0761	-	Ribonuclease H(EC:3.1.26.4)	770	29	0.037662338
Cthe_2720	+	ribosomal protein L11	425	16	0.037647059
Cthe_1098	-	hypothetical protein	1289	48	0.037238169
Cthe_2149	-	amine oxidase	1289	48	0.037238169
Cthe_1569	+	O-acetylhomoserine aminocarboxypropyltransferase(EC:2.5.1.49)	1316	49	0.037234043
Cthe_0895	-	RNA polymerase, sigma 38 subunit, RpoS	1076	40	0.037174721
Cthe_0986	-	peptidase M16-like protein	1292	48	0.037151703
Cthe_1802	-	cobalt ABC transporter, inner membrane subunit CbiQ	809	30	0.037082818
Cthe_0283	-	aldo/keto reductase	947	35	0.036958817
Cthe_0205	+	hypothetical protein	1958	72	0.036772217
Cthe_2096	-	methionyl-tRNA synthetase(EC:6.1.1.10)	1964	72	0.036659878
Cthe_2361	-	DNA gyrase, A subunit	2483	91	0.036649215
Cthe_0495	+	RNA polymerase, sigma 28 subunit	656	24	0.036585366
Cthe_0770	-	signal recognition particle protein	1340	49	0.036567164
Cthe_1251	-	Xanthine/uracil/vitamin C permease	1397	51	0.036506800
Cthe_0987	-	riboflavin biosynthesis protein RibF	938	34	0.036247335
Cthe_1212	-	hypothetical protein	938	34	0.036247335
Cthe_0179	-	Argininosuccinate synthase(EC:6.3.4.5)	1217	44	0.036154478
Cthe_0275	+	glycosyltransferase 36	2435	88	0.036139630
Cthe_0760	-	hypothetical protein	1634	59	0.036107711
Cthe_0959	-	S-adenosylmethionine--tRNA-ribosyltransferase- isomerase	1025	37	0.036097561
Cthe_1260	-	4-hydroxybenzoyl-CoA thioesterase	416	15	0.036057692
Cthe_2989	+	glycosyltransferase 36	2954	106	0.035883548
Cthe_0655	+	cysteine desulfurase family protein	1145	41	0.035807860
Cthe_1373	-	YD repeat protein	5753	206	0.035807405
Cthe_2410	+	protein of unknown function DUF342	1592	57	0.035804020
Cthe_0949	-	carbamoyl-phosphate synthase, large subunit	3218	115	0.035736482
Cthe_1041	-	UDP-N-acetylmuramoylalanine--D-glutamate ligase	1400	50	0.035714286
Cthe_2378	+	parB-like partition protein	869	31	0.035673188

المنارة للاستشارات

www.manaraa.com

Cthe_1446	+	peptidase S1 and S6, chymotrypsin/Hap	2537	90	0.035474970
Cthe_0444	+	cell division protein FtsA	1241	44	0.035455278
Cthe_0499	+	hypothetical protein	311	11	0.035369775
Cthe_1029	-	phosphate acetyltransferase(EC:2.3.1.8)	1076	38	0.035315985
Cthe_2431	+	electron transport complex, RnfABCDGE type, D subunit	965	34	0.035233161
Cthe_0541	+	peptidase M24	1193	42	0.035205365
Cthe_2588	+	putative transcriptional acitvator, Baf family	767	27	0.035202086
Cthe_0539	+	ABC transporter related protein	683	24	0.035139092
Cthe_1313	-	phosphopantothenoylcysteine decarboxylase/phosphopantothenate--cysteine ligase(EC:4.1.1.36,EC:6.3.2.5)	1196	42	0.035117057
Cthe_1452	+	KWG repeat containing protein	1655	58	0.035045317
Cthe_0290	+	Homoserine dehydrogenase(EC:1.1.1.3)	1286	45	0.034992224
Cthe_1852	-	Hsp33 protein	890	31	0.034831461
Cthe_2513	+	peptidase S1 and S6, chymotrypsin/Hap	1178	41	0.034804754
Cthe_2262	+	V-type ATPase, 116 kDa subunit	1955	68	0.034782609
Cthe_3095	-	glycosyl transferase, family 39	2906	101	0.034755678
Cthe_1060	-	AIR synthase related protein-like protein	983	34	0.034587996
Cthe_1848	-	copper-translocating P-type ATPase(EC:3.6.3.4)	2231	77	0.034513671
Cthe_0950	-	carbamoyl-phosphate synthase, small subunit(EC:6.3.5.5)	1073	37	0.034482759
Cthe_0249	+	peptidase C11, clostripain	2408	83	0.034468439
Cthe_2283	-	methyl-accepting chemotaxis sensory transducer	4427	152	0.034334764
Cthe_1307	-	cellulosome anchoring protein, cohesin region	1895	65	0.034300792
Cthe_0335	+	hydrogenase large subunit-like protein	1346	46	0.034175334
Cthe_2856	+	hypothetical protein	527	18	0.034155598
Cthe_2884	+	Imidazoleglycerol-phosphate dehydratase(EC:4.2.1.19)	587	20	0.034071550
Cthe_2857	+	hypothetical protein	470	16	0.034042553
Cthe_0507	-	Exonuclease, RNase T and DNA polymerase III	647	22	0.034003091
Cthe_2581	+	dihydropteroate synthase(EC:2.5.1.15)	1181	40	0.033869602
Cthe_2711	+	hypothetical protein	1181	40	0.033869602
Cthe_1543	-	aspartyl-tRNA synthetase	1745	59	0.033810888
Cthe_2203	-	GTP cyclohydrolase I(EC:3.5.4.16)	563	19	0.033747780
Cthe_0070	+	asparaginyl-tRNA synthetase(EC:6.1.1.22)	1394	47	0.033715925
Cthe_0536	+	glycoside hydrolase, family 5	1691	57	0.033707865
Cthe_1376	-	Homoserine dehydrogenase(EC:1.1.1.3)	1247	42	0.033680834
Cthe_0445	+	cell division protein FtsZ	1130	38	0.033628319
Cthe_3027	+	Citrate (Si)-synthase(EC:2.3.3.1)	1370	46	0.033576642
Cthe_2872	-	glycoside hydrolase, family 5	1700	57	0.033529412
Cthe_0271	+	type 3a, cellulose-binding	926	31	0.033477322
Cthe_1164	-	FAD dependent oxidoreductase	1586	53	0.033417402
Cthe_2030	-	intein	1676	56	0.033412888
Cthe_3087	+	response regulator receiver protein	809	27	0.033374536
Cthe_1778	-	copper amine oxidase-like protein	959	32	0.033368092
Cthe_0914	+	hypothetical protein	2552	85	0.033307210

www.manaraa.com

Cthe_1842	-	O-acetylhomoserine/O-acetylserine sulfhydrylase(EC:2.5.1.49)	1292	43	0.033281734
Cthe_0240	+	hypothetical protein	782	26	0.033248082
Cthe_1035	-	DNA ligase, NAD-dependent(EC:6.5.1.2)	1991	66	0.033149171
Cthe_1786	-	DNA repair protein RadA	1364	45	0.032991202
Cthe_2383	-	hypothetical protein	914	30	0.032822757
Cthe_2391	+	pyruvate ferredoxin/flavodoxin oxidoreductase, delta subunit	305	10	0.032786885
Cthe_1315	-	Guanylate kinase(EC:2.7.4.8)	611	20	0.032733224
Cthe_0082	+	ATP-dependent protease La(EC:3.4.21.53)	2447	80	0.032693094
Cthe_0700	+	biotin/lipoyl attachment protein	398	13	0.032663317
Cthe_2095	-	hydrolase, TatD family	767	25	0.032594524
Cthe_1931	-	S-layer-like domain containing protein	1565	51	0.032587859
Cthe_2844	+	intein	1205	39	0.032365145
Cthe_1861	-	transcriptional regulator, CdaR	1085	35	0.032258065
Cthe_0977	-	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanyl ligase	1373	44	0.032046613
Cthe_1570	+	extracellular solute-binding protein, family 3	1061	34	0.032045240
Cthe_2600	+	glycosyl transferase, family 4	1130	36	0.031858407
Cthe_0339	+	histidine kinase	566	18	0.031802120
Cthe_3119	+	flavin reductase-like, FMN-binding	662	21	0.031722054
Cthe_2364	-	glucose inhibited division protein A	1892	60	0.031712474
Cthe_0624	+	glycoside hydrolase, family 9-like Ig-like	4805	152	0.031633715
Cthe_3148	+	ABC transporter related protein	1898	60	0.031612223
Cthe_2018	-	hypothetical protein	1235	39	0.031578947
Cthe_2371	+	chromosomal replication initiator protein DnaA	1331	42	0.031555222
Cthe_0276	+	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	959	30	0.031282586
Cthe_1004	-	uridylate kinase	707	22	0.031117397
Cthe_0312	+	ATPase AAA-2	2411	75	0.031107424
Cthe_1763	-	ABC transporter related protein	740	23	0.031081081
Cthe_0906	-	Radical SAM	1352	42	0.031065089
Cthe_2880	+	histidyl-tRNA synthetase(EC:6.1.1.21)	1256	39	0.031050955
Cthe_0202	+	glutamine synthetase, type I(EC:6.3.1.2)	1322	41	0.031013616
Cthe_0365	+	peptide chain release factor 2	1073	33	0.030754893
Cthe_1031	-	glutamyl-tRNA(Gln) amidotransferase, B subunit	1433	44	0.030704815
Cthe_2193	-	Carbohydrate binding family 6	2846	87	0.030569220
Cthe_1565	+	Nitrogenase(EC:1.18.6.1)	1475	45	0.030508475
Cthe_0572	+	radical SAM enzyme, Cfr family	1049	32	0.030505243
Cthe_0467	+	Flagellar biosynthesis/type III secretory pathway protein-like protein	791	24	0.030341340
Cthe_2430	+	electron transport complex, RnfABCDGE type, C subunit	1319	40	0.030326005
Cthe_0687	+	chromosome segregation and condensation protein ScpA	794	24	0.030226700
Cthe_0794	-	aluminium resistance protein	1292	39	0.030185759
Cthe_0372	-	glutamate synthase (NADPH), homotetrameric	1394	42	0.030129125
Cthe_0402	+	copper amine oxidase-like protein	2198	66	0.030027298
Cthe_2269	+	V-type ATPase, D subunit(EC:3.6.3.15)	668	20	0.029940120

المنارة للاستشارات

www.manaraa.com

Cthe_1269	+	hypothetical protein	1283	38	0.029618083
Cthe_2911	+	ribosomal protein L29	203	6	0.029556650
Cthe_2328	-	UDP-N-acetylglucosamine 1-carboxyvinyltransferase(EC:2.5.1.7)	1253	37	0.029529130
Cthe_1943	-	deoxyribose-phosphate aldolase(EC:4.1.2.4)	683	20	0.029282577
Cthe_1235	-	Cellulose 1,4-beta-cellobiosidase(EC:3.2.1.91)	20657	604	0.029239483
Cthe_0945	-	metallophosphoesterase	992	29	0.029233871
Cthe_1351	-	acetobutylicum phosphotransbutyrylase	479	14	0.029227557
Cthe_1257	-	Carbohydrate-binding, CenC-like protein	3152	92	0.029187817
Cthe_2117	+	putative sulfonate transport system substrate-binding protein	1028	30	0.029182879
Cthe_2381	+	seryl-tRNA synthetase(EC:6.1.1.11)	1271	37	0.029110936
Cthe_1566	+	Nitrogenase(EC:1.18.6.1)	1343	39	0.029039464
Cthe_1322	-	chaperone protein DnaK	1826	53	0.029025192
Cthe_2086	-	peptidase U32	2519	73	0.028979754
Cthe_0160	+	ribosomal protein L21	311	9	0.028938907
Cthe_0873	-	anthranilate phosphoribosyltransferase(EC:2.4.2.18)	1037	30	0.028929605
Cthe_0952	-	dihydroorotase, multifunctional complex type(EC:3.5.2.3)	1280	37	0.028906250
Cthe_0974	-	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	1109	32	0.028854824
Cthe_2584	+	hypothetical protein	626	18	0.028753994
Cthe_0777	-	DNA mismatch repair protein MutS	2612	75	0.028713629
Cthe_2330	-	hypothetical protein	1466	42	0.028649386
Cthe_2441	-	transcriptional regulator, DeoR family	803	23	0.028642590
Cthe_0784	-	hypothetical protein	875	25	0.028571429
Cthe_0299	+	MATE efflux family protein	1370	39	0.028467153
Cthe_1833	-	copper amine oxidase-like protein	809	23	0.028430161
Cthe_1841	-	hypothetical protein	809	23	0.028430161
Cthe_0088	+	cell shape determining protein, MreB/Mrl family	1022	29	0.028375734
Cthe_2839	+	hypothetical protein	1238	35	0.028271405
Cthe_2605	+	ATP synthase F1, delta subunit	566	16	0.028268551
Cthe_1256	-	glycoside hydrolase, family 3-like protein	2267	64	0.028231142
Cthe_0815	-	arginine repressor, ArgR	461	13	0.028199566
Cthe_1775	-	peptidase M22, glycoprotease	710	20	0.028169014
Cthe_3166	+	glucose-1-phosphate adenylyltransferase(EC:2.7.7.27)	1280	36	0.028125000
Cthe_2697	+	alanine racemase(EC:5.1.1.1)	1175	33	0.028085106
Cthe_1782	-	ribosomal protein S9	392	11	0.028061224
Cthe_0201	+	glutamate synthase, alpha subunit-like protein	722	20	0.027700831
Cthe_3109	+	Uracil-DNA glycosylase superfamily	578	16	0.027681661
Cthe_1923	-	CTP synthase(EC:6.3.4.2)	1628	45	0.027641278
Cthe_0560	-	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	1775	49	0.027605634
Cthe_0067	+	Silent information regulator protein Sir2	725	20	0.027586207
Cthe_1064	-	aminotransferase, class V	1163	32	0.027515047
Cthe_1938	-	D-alanine--D-alanine ligase(EC:6.3.2.4)	1130	31	0.027433628
Cthe_1162	-	glucosamine--fructose-6-phosphate	1823	50	0.027427318

المنارة للاستشارات
www.manaraa.com

		aminotransferase, isomerizing(EC:2.6.1.16)			
Cthe_0054	+	protein of unknown function DUF187	2189	60	0.027409776
Cthe_0451	+	protein of unknown function DUF115	1871	51	0.027258151
Cthe_0186	+	UDP-glucose 4-epimerase(EC:5.1.3.2)	1064	29	0.027255639
Cthe_0001	+	Recombinase	1619	44	0.027177270
Cthe_2326	-	protein of unknown function DUF163	479	13	0.027139875
Cthe_1341	-	Radical SAM	1517	41	0.027027027
Cthe_1940	-	RDD domain containing protein	740	20	0.027027027
Cthe_2594	+	zinc/iron permease	740	20	0.027027027
Cthe_0311	+	excinuclease ABC, A subunit	2828	76	0.026874116
Cthe_1540	-	glutamyl-tRNA(Gln) amidotransferase, B subunit	1415	38	0.026855124
Cthe_0461	+	gid protein	1310	35	0.026717557
Cthe_1541	-	glutamyl-tRNA(Gln) amidotransferase, A subunit	1460	39	0.026712329
Cthe_2280	+	ABC transporter related protein	824	22	0.026699029
Cthe_0413	+	glycoside hydrolase, family 9	3674	98	0.026673925
Cthe_3150	+	cobalamin biosynthesis protein CobD	1055	28	0.026540284
Cthe_0161	+	protein of unknown function DUF464	341	9	0.026392962
Cthe_1105	-	type II secretion system protein	1214	32	0.026359143
Cthe_2399	-	Formate--tetrahydrofolate ligase(EC:6.3.4.3)	1670	44	0.026347305
Cthe_1804	-	cell wall hydrolase/autolysin	722	19	0.026315789
Cthe_1568	+	Radical SAM	875	23	0.026285714
Cthe_2104	-	PSP1	875	23	0.026285714
Cthe_2694	+	O-antigen polymerase	1598	42	0.026282854
Cthe_2474	+	phage terminase, large subunit, PBSX family	1256	33	0.026273885
Cthe_1199	-	amidohydrolase	1295	34	0.026254826
Cthe_0203	+	response regulator receiver and ANTAR domain protein	572	15	0.026223776
Cthe_0713	+	1-acyl-sn-glycerol-3-phosphate acyltransferase	611	16	0.026186579
Cthe_1308	-	pyruvate, phosphate dikinase(EC:2.7.9.1)	2651	69	0.026027914
Cthe_0153	+	Holliday junction resolvase YqgF	425	11	0.025882353
Cthe_1760	+	penicillin-binding protein, 1A family	2474	64	0.025869038
Cthe_0176	+	hypothetical protein	350	9	0.025714286
Cthe_0543	+	glycoside hydrolase, family 9	2219	57	0.025687247
Cthe_2973	+	sugar fermentation stimulation protein	701	18	0.025677603
Cthe_3000	+	phosphate transporter	1052	27	0.025665399
Cthe_2171	-	type III restriction enzyme, res subunit	2969	76	0.025597844
Cthe_0776	-	DNA mismatch repair protein MutL	2267	58	0.025584473
Cthe_2437	+	shikimate kinase	509	13	0.025540275
Cthe_1320	-	ribosomal protein L11 methyltransferase	941	24	0.025504782
Cthe_0771	-	putative helix-turn-helix protein, YlxM/p13-like protein	353	9	0.025495751
Cthe_1247	-	phosphoribosylglycinamide formyltransferase	629	16	0.025437202
Cthe_1102	-	fimbrial assembly protein	551	14	0.025408348
Cthe_0184	+	type IV pilus assembly protein PilM	1103	28	0.025385313
Cthe_1172	+	protein of unknown function DUF362	1148	29	0.025261324
Cthe_3147	+	ABC transporter related protein	1745	44	0.025214900

المنارة للاستشارات
www.manaraa.com

Cthe_1947	-	Redoxin	596	15	0.025167785
Cthe_1864	+	acetylglutamate kinase(EC:2.7.2.8)	914	23	0.025164114
Cthe_0559	+	single-strand binding protein/Primosomal replication protein n	716	18	0.025139665
Cthe_0485	+	flagellar biosynthesis protein FlhA	2030	51	0.025123153
Cthe_2329	-	hypothetical protein	836	21	0.025119617
Cthe_0465	+	flagellar M-ring protein FliF	1553	39	0.025112685
Cthe_2812	+	glycoside hydrolase, family 9	1835	46	0.025068120
Cthe_2120	-	RNA polymerase sigma factor	761	19	0.024967148
Cthe_1250	-	phosphoribosylaminoimidazole carboxylase, catalytic subunit(EC:4.1.1.21)	521	13	0.024952015
Cthe_2334	-	polysaccharide biosynthesis protein CapD	1844	46	0.024945770
Cthe_3090	+	hydro-lyases, Fe-S type, tartrate/fumarate subfamily, alpha subunit	842	21	0.024940618
Cthe_1316	-	protein of unknown function DUF370	281	7	0.024911032
Cthe_3227	+	copper amine oxidase-like protein	803	20	0.024906600
Cthe_0575	+	ribosome small subunit-dependent GTPase A	884	22	0.024886878
Cthe_0462	+	flagellar basal-body rod protein FlgB	404	10	0.024752475
Cthe_0978	-	UDP-N-acetylmuramyl-tripeptide synthetases(EC:6.3.2.13)	1457	36	0.024708305
Cthe_0678	+	pyrimidine-nucleoside phosphorylase(EC:2.4.2.2)	1301	32	0.024596464
Cthe_1431	+	hypothetical protein	1955	48	0.024552430
Cthe_1989	-		1674	41	0.024492234
Cthe_0769	-	ribosomal protein S16	245	6	0.024489796
Cthe_0433	+	glycoside hydrolase, family 9	2369	58	0.024482904
Cthe_3149	+	aminoacyl-histidine dipeptidase(EC:3.4.13.3)	1430	35	0.024475524
Cthe_1363	-	lipopolysaccharide biosynthesis(EC:2.7.10.1)	1394	34	0.024390244
Cthe_2433	+	electron transport complex, RnfABCDGE type, E subunit	656	16	0.024390244
Cthe_0267	-	type 3a, cellulose-binding	2015	49	0.024317618
Cthe_0306	+	DNA gyrase/topoisomerase IV, subunit A	2183	53	0.024278516
Cthe_2876	+	ATP-dependent DNA helicase PcrA	2225	54	0.024269663
Cthe_0858	+	protein of unknown function DUF1432	989	24	0.024266936
Cthe_0325	+	NAD+ synthetase	1937	47	0.024264326
Cthe_0043	+	glycoside hydrolase, family 9	2228	54	0.024236984
Cthe_1319	-	protein of unknown function DUF558	743	18	0.024226110
Cthe_0554	+	phosphoribosylformylglycinamidine synthase(EC:6.3.5.3)	3770	91	0.024137931
Cthe_1382	-	Patatin	1124	27	0.024021352
Cthe_2327	-	beta-lactamase-like protein	791	19	0.024020228
Cthe_1381	-	threonine synthase(EC:4.2.3.1)	1499	36	0.024016011
Cthe_2877	+	S-layer-like domain containing protein	1757	42	0.023904382
Cthe_0310	-	hypothetical protein	1046	25	0.023900574
Cthe_1844	-	transcriptional regulator, BadM/Rrf2 family	461	11	0.023861171
Cthe_2604	+	ATP synthase F0, B subunit	545	13	0.023853211
Cthe_2179	-	Pectate lyase/Amb allergen	2768	66	0.023843931
Cthe_0947	-	dihydroorotate dehydrogenase family protein	923	22	0.023835320
Cthe_0175	+	polysaccharide deacetylase	1136	27	0.023767606
Cthe_3002	+	hypothetical protein	548	13	0.023722628

المنارة للاستشارات

www.manaraa.com

Cthe_1274	+	nucleoside recognition	1307	31	0.023718439
Cthe_1270	+	proteinase inhibitor I4, serpin	1223	29	0.023712183
Cthe_2369	-	ribonuclease P protein component	380	9	0.023684211
Cthe_2615	+	UDP-N-acetylglucosamine 1-carboxyvinyltransferase(EC:2.5.1.7)	1268	30	0.023659306
Cthe_1835	-	Viral A-type inclusion protein repeat containing protein	3308	78	0.023579202
Cthe_1458	-	extracellular solute-binding protein, family 3	809	19	0.023485785
Cthe_0251	+	transglutaminase-like protein	2555	60	0.023483366
Cthe_2529	+	Porphobilinogen synthase(EC:4.2.1.24)	980	23	0.023469388
Cthe_2107	-	thymidylate kinase	683	16	0.023426061
Cthe_0910	-	extracellular solute-binding protein, family 5	1667	39	0.023395321
Cthe_0459	+	DNA protecting protein DprA	1112	26	0.023381295
Cthe_2408	+	phage shock protein A, PspA	728	17	0.023351648
Cthe_0847	-	translation elongation factor P	557	13	0.023339318
Cthe_1362	-	NusG antitermination factor	557	13	0.023339318
Cthe_0477	+	flagellar motor switch protein FliM	986	23	0.023326572
Cthe_1200	-	adenosylhomocysteinase(EC:3.3.1.1)	1244	29	0.023311897
Cthe_2819	+	methyl-accepting chemotaxis sensory transducer	4424	103	0.023282098
Cthe_0262	+	gamma-glutamyl phosphate reductase(EC:1.2.1.41)	1295	30	0.023166023
Cthe_0926	-	signal recognition particle-docking protein FtsY	908	21	0.023127753
Cthe_0100	+	hypothetical protein	782	18	0.023017903
Cthe_2630	+	ribose-phosphate pyrophosphokinase(EC:2.7.6.1)	962	22	0.022869023
Cthe_0627	+	hypothetical protein	875	20	0.022857143
Cthe_0653	+	putative transcriptional regulator, CopG family	263	6	0.022813688
Cthe_2886	+	imidazole glycerol phosphate synthase, glutamine amidotransferase subunit	614	14	0.022801303
Cthe_0942	-	MiaB-like tRNA modifying enzyme YliG	1361	31	0.022777370
Cthe_1106	-	twitching motility protein	1055	24	0.022748815
Cthe_1324	-	heat-inducible transcription repressor HrcA	1055	24	0.022748815
Cthe_2019	-	intein	1847	42	0.022739578
Cthe_2476	+	phage putative head morphogenesis protein, SPP1 gp7 family	1409	32	0.022711143
Cthe_0999	-	1-deoxy-D-xylulose 5-phosphate reductoisomerase(EC:1.1.1.267)	1145	26	0.022707424
Cthe_0250	+	protein of unknown function UPF0027	1190	27	0.022689076
Cthe_2345	-	DegT/DnrJ/EryC1/StrS aminotransferase	1148	26	0.022648084
Cthe_0599	+	thiazole biosynthesis protein ThiH	1109	25	0.022542831
Cthe_2601	+	UDP-N-acetylglucosamine 2-epimerase(EC:5.1.3.14)	1154	26	0.022530329
Cthe_0466	+	flagellar motor switch protein FliG	1022	23	0.022504892
Cthe_0566	+	primosomal protein N'	2447	55	0.022476502
Cthe_2189	+	diguanylate cyclase with GAF sensor	1247	28	0.022453889
Cthe_2257	-	ribosomal protein L9	446	10	0.022421525
Cthe_2921	+	ribosomal protein L30	179	4	0.022346369
Cthe_0933	-	acyl carrier protein	224	5	0.022321429
Cthe_1846	-	periplasmic sensor signal transduction histidine kinase	1391	31	0.022286125

المنارة للاستشارات
www.manaraa.com

Cthe_1503	+	hypothetical protein	359	8	0.022284123
Cthe_0837	-	protein of unknown function DUF322	404	9	0.022277228
Cthe_1225	-	translation initiation factor IF-3	494	11	0.022267206
Cthe_1375	-	aspartate kinase(EC:2.7.2.4)	1349	30	0.022238695
Cthe_0327	+	S-layer-like domain containing protein	2573	57	0.022153129
Cthe_0185	+	cell wall hydrolase/autolysin	2846	63	0.022136332
Cthe_0232	+	hypothetical protein	995	22	0.022110553
Cthe_0111	+	PHP-like protein	725	16	0.022068966
Cthe_1361	-	NAD-dependent epimerase/dehydratase	1043	23	0.022051774
Cthe_0233	+	GCN5-related N-acetyltransferase	590	13	0.022033898
Cthe_0829	-	hypothetical protein	728	16	0.021978022
Cthe_1774	-	ribosomal-protein-alanine acetyltransferase	458	10	0.021834061
Cthe_1284	-	glycogen/starch synthases, ADP-glucose type(EC:2.4.1.21)	1466	32	0.021828104
Cthe_3154	+	hypothetical protein	551	12	0.021778584
Cthe_0647	+	anaerobic ribonucleoside-triphosphate reductase activating protein	689	15	0.021770682
Cthe_1803	-	cobalamin (vitamin B12) biosynthesis CbiM protein	1058	23	0.021739130
Cthe_2130	-	hypothetical protein	1292	28	0.021671827
Cthe_2965	-	binding-protein-dependent transport systems inner membrane component	923	20	0.021668472
Cthe_2332	-	multi-sensor signal transduction histidine kinase	1802	39	0.021642619
Cthe_0634	-	serine/threonine protein kinase	2087	45	0.021562051
Cthe_1151	-	hypothetical protein	1067	23	0.021555764
Cthe_2515	+	Guanine deaminase(EC:3.5.4.3)	464	10	0.021551724
Cthe_0188	+	ribonuclease PH(EC:2.7.7.56)	791	17	0.021491783
Cthe_2406	+	hypothetical protein	512	11	0.021484375
Cthe_2119	-	glycoside hydrolase, family 10	2282	49	0.021472393
Cthe_2972	-	glycoside hydrolase, family 11	2051	44	0.021452950
Cthe_3191	+	ATPase involved in chromosome partitioning-like protein	1220	26	0.021311475
Cthe_0901	-	pantoate--beta-alanine ligase(EC:6.3.2.1)	845	18	0.021301775
Cthe_3103	+	2-hydroxyglutaryl-CoA dehydratase, D-component	986	21	0.021298174
Cthe_0274	-	glycoside hydrolase, family 9	1691	36	0.021289178
Cthe_2614	+	hypothetical protein	752	16	0.021276596
Cthe_2191	+	1,4-alpha-glucan branching enzyme(EC:2.4.1.18)	2213	47	0.021238138
Cthe_0686	+	tryptophanyl-tRNA synthetase(EC:6.1.1.2)	989	21	0.021233569
Cthe_1860	-	cell division ATP-binding protein FtsE	707	15	0.021216407
Cthe_0182	+	Holliday junction DNA helicase RuvB	992	21	0.021169355
Cthe_0890	-	RNA polymerase, sigma-24 subunit, ECF subfamily	662	14	0.021148036
Cthe_1764	-	Outer membrane protein-like protein	1088	23	0.021139706
Cthe_0091	+	Peptidoglycan glycosyltransferase(EC:2.4.1.129)	2129	45	0.021136684
Cthe_0337	+	PHP-like protein	710	15	0.021126761
Cthe_1575	+	purine nucleoside phosphorylase(EC:2.4.2.1)	710	15	0.021126761
Cthe_1937	-	glutamate racemase(EC:5.1.1.3)	806	17	0.021091811

المنارة للاستشارات

www.manaraa.com

Cthe_2092	-	dimethyladenosine transferase(EC:2.1.1.48)	854	18	0.021077283
Cthe_0905	-	metal dependent phosphohydrolase	665	14	0.021052632
Cthe_1071	-	PhoH-like protein	1001	21	0.020979021
Cthe_2148	-	Carbohydrate-binding, CenC-like protein	764	16	0.020942408
Cthe_0064	-	Na/Pi-cotransporter II-related protein	956	20	0.020920502
Cthe_0616	+	amino acid-binding ACT	431	9	0.020881671
Cthe_1264	-	DNA polymerase III, alpha subunit(EC:2.7.7.7)	3545	74	0.020874471
Cthe_0441	+	UDP-N-acetylglucosamine 1-carboxyvinyltransferase(EC:2.5.1.7)	863	18	0.020857474
Cthe_1910	-	hypothetical protein	1247	26	0.020850040
Cthe_1961	+	Nucleotidyl transferase	2450	51	0.020816327
Cthe_0629	+	type II secretion system protein E	1445	30	0.020761246
Cthe_0710	+	hypothetical protein	434	9	0.020737327
Cthe_2685	+	putative TIM-barrel protein, nifR3 family	968	20	0.020661157
Cthe_0486	+	GTP-binding signal recognition particle SRP54, G-domain	1211	25	0.020644096
Cthe_1014	-	MutS2 family protein	2381	49	0.020579588
Cthe_3196	-	Uncharacterized conserved protein UCP033563	1361	28	0.020573108
Cthe_0468	+	flagellar protein export ATPase FliI(EC:3.6.3.15)	1313	27	0.020563595
Cthe_1177	-	putative nicotinate phosphoribosyltransferase	1460	30	0.020547945
Cthe_1326	-	small GTP-binding protein	1811	37	0.020430701
Cthe_1564	+	GCN5-related N-acetyltransferase	539	11	0.020408163
Cthe_2944	+	periplasmic sensor signal transduction histidine kinase	1814	37	0.020396913
Cthe_1956	-	Tagatose-6-phosphate kinase(EC:2.7.1.144)	932	19	0.020386266
Cthe_2014	-	hypothetical protein	638	13	0.020376176
Cthe_1063	-	thiamine biosynthesis/tRNA modification protein ThiI	1178	24	0.020373514
Cthe_0611	+	RNA-metabolising metallo-beta-lactamase	1625	33	0.020307692
Cthe_2892	+	chaperonin GroEL	1625	33	0.020307692
Cthe_2365	-	tRNA modification GTPase TrmE	1379	28	0.020304569
Cthe_2010	-	hypothetical protein	1238	25	0.020193861
Cthe_1863	+	N-acetyl-gamma-glutamyl-phosphate reductase(EC:1.2.1.38)	1040	21	0.020192308
Cthe_2990	+	peptide methionine sulfoxide reductase	596	12	0.020134228
Cthe_2576	+	hypothetical protein	1292	26	0.020123839
Cthe_0373	-	oxidoreductase FAD/NAD(P)-binding	845	17	0.020118343
Cthe_0368	+	NCAIR mutase (PurE)-related protein	746	15	0.020107239
Cthe_0243	+	copper amine oxidase-like protein	848	17	0.020047170
Cthe_2681	+	serine phosphatase(EC:3.1.3.16)	2396	48	0.020033389
Cthe_0328	-	peptide chain release factor 3	1601	32	0.019987508
Cthe_2403	+	4-diphosphocytidyl-2C-methyl-D-erythritol kinase	851	17	0.019976498
Cthe_0435	+	cellulosome enzyme, dockerin type I	1052	21	0.019961977
Cthe_2595	+	Sua5/YciO/YrdC/YwlC family protein	1055	21	0.019905213
Cthe_0569	+	protein of unknown function DUF116	755	15	0.019867550
Cthe_1483	-	hypothetical protein	1058	21	0.019848771
Cthe_2738	+	metallophosphoesterase	707	14	0.019801980
Cthe_1791	-	UvrB/UvrC protein	506	10	0.019762846

المنارة للاستشارات
www.manaraa.com

Cthe_3160	+	putative RNA methylase, NOL1/NOP2/sun family	1370	27	0.019708029
Cthe_0115	+	glycogen debranching enzyme, putative	1979	39	0.019706923
Cthe_3097	-	glycosyl transferase, family 2	965	19	0.019689119
Cthe_0576	+	Ribulose-phosphate 3-epimerase(EC:5.1.3.1)	662	13	0.019637462
Cthe_0601	+	thiamine-phosphate pyrophosphorylase(EC:2.5.1.3)	1070	21	0.019626168
Cthe_1832	-	hypothetical protein	1529	30	0.019620667
Cthe_2871	+	TipAS antibiotic-recognition	512	10	0.019531250
Cthe_2865	+	phage / plasmid primase, P4 family	2153	42	0.019507664
Cthe_2535	+	adenylylsulfate reductase, thioredoxin dependent(EC:1.8.4.8)	719	14	0.019471488
Cthe_1229	-	Spore germination protein-like protein	617	12	0.019448947
Cthe_0491	+	CheW protein	464	9	0.019396552
Cthe_0995	-	protein of unknown function DUF150	464	9	0.019396552
Cthe_1757	+	peptidase M23B	929	18	0.019375673
Cthe_0492	+	CheC, inhibitor of MCP methylation	620	12	0.019354839
Cthe_2902	+	ribosomal protein S10	311	6	0.019292605
Cthe_1936	-	hypothetical protein	467	9	0.019271949
Cthe_0743	-	MATE efflux family protein	1454	28	0.019257221
Cthe_1840	-	cysteine synthase A(EC:2.5.1.47)	935	18	0.019251337
Cthe_2632	+	transcription-repair coupling factor	3536	68	0.019230769
Cthe_1859	-	protein of unknown function DUF214	884	17	0.019230769
Cthe_2363	-	methyltransferase GidB	728	14	0.019230769
Cthe_2060	-	RNA methyltransferase, TrmH family, group 3	833	16	0.019207683
Cthe_1579	+	ABC transporter related protein	1562	30	0.019206146
Cthe_2809	+	glycoside hydrolase, family 16	3965	76	0.019167718
Cthe_0129	+	metal dependent phosphohydrolase	1517	29	0.019116678
Cthe_0948	-	oxidoreductase FAD/NAD(P)-binding	785	15	0.019108280
Cthe_0081	-	arginine biosynthesis bifunctional protein ArgJ(EC:2.3.1.35)	1205	23	0.019087137
Cthe_0570	+	peptidase, membrane zinc metallopeptidase, putative	683	13	0.019033675
Cthe_0918	-	cellulosome enzyme, dockerin type I	3629	69	0.019013502
Cthe_1138	-	protein of unknown function DUF87	1685	32	0.018991098
Cthe_2579	+	iron-containing alcohol dehydrogenase	1106	21	0.018987342
Cthe_2695	+	hypothetical protein	632	12	0.018987342
Cthe_3107	+	Radical SAM	632	12	0.018987342
Cthe_1321	-	chaperone protein DnaJ	1160	22	0.018965517
Cthe_3123	+	von Willebrand factor, type A	3902	74	0.018964634
Cthe_3124	+	AMP-dependent synthetase and ligase	1637	31	0.018937080
Cthe_2031	-	transposase IS200-like protein	476	9	0.018907563
Cthe_1037	-	cell wall hydrolase/autolysin	848	16	0.018867925
Cthe_1389	+	metal dependent phosphohydrolase(EC:2.7.7.25)	1328	25	0.018825301
Cthe_0489	+	response regulator receiver modulated CheB methylesterase(EC:3.1.1.61)	1064	20	0.018796992
Cthe_1089	-	Stage V sporulation protein S	266	5	0.018796992
Cthe_1939	-	magnesium transporter	1334	25	0.018740630
Cthe_0985	-	peptidase M16-like protein	1283	24	0.018706157

المنارة للاستشارات

www.manaraa.com

Cthe_2168	-	Propeptide, PepSY amd peptidase M4	2195	41	0.018678815
Cthe_2039	-	UvrD/REP helicase	3755	70	0.018641811
Cthe_0305	+	DNA topoisomerase (ATP-hydrolyzing)(EC:5.99.1.3)	1985	37	0.018639798
Cthe_1776	-	protein of unknown function UPF0079	485	9	0.018556701
Cthe_0550	+	heat shock protein Hsp90	1904	35	0.018382353
Cthe_1183	-	ATPase associated with various cellular activities, AAA_3	929	17	0.018299247
Cthe_0436	+	Tetratricopeptide TPR_2	3170	58	0.018296530
Cthe_0625	+	glycoside hydrolase, family 9	2132	39	0.018292683
Cthe_0087	+	maf protein	602	11	0.018272425
Cthe_0360	+	thioredoxin	329	6	0.018237082
Cthe_0266	+	methyl-accepting chemotaxis sensory transducer	1487	27	0.018157364
Cthe_1383	-	Tetratricopeptide TPR_2	1157	21	0.018150389
Cthe_2587	+	hypothetical protein	497	9	0.018108652
Cthe_2885	+	phosphoribosylaminoimidazole-succinocarboxamide synthase(EC:6.3.2.6)	884	16	0.018099548
Cthe_0656	+	type IV pilus assembly protein PilM	1106	20	0.018083183
Cthe_0478	+	CheC, inhibitor of MCP methylation	1220	22	0.018032787
Cthe_0351	+	PpiC-type peptidyl-prolyl cis-trans isomerase	1391	25	0.017972682
Cthe_0979	-	Peptidoglycan glycosyltransferase(EC:2.4.1.129)	2171	39	0.017964072
Cthe_0628	+	hypothetical protein	836	15	0.017942584
Cthe_2362	-	parB-like partition protein	836	15	0.017942584
Cthe_0198	+	Glutamate synthase (NADPH)(EC:1.4.1.13)	1505	27	0.017940199
Cthe_0922	-	diaminopimelate dehydrogenase(EC:1.4.1.16)	1004	18	0.017928287
Cthe_1618	-	hypothetical protein	2288	41	0.017919580
Cthe_0254	+	protein of unknown function DUF58	1340	24	0.017910448
Cthe_1957	-	extracellular solute-binding protein, family 1	1340	24	0.017910448
Cthe_2163	-	anti-sigma-factor antagonist	335	6	0.017910448
Cthe_0295	-	phosphoserine aminotransferase(EC:2.6.1.52)	1118	20	0.017889088
Cthe_0789	-	RNA-binding S4	791	14	0.017699115
Cthe_2743	+	putative metalloendopeptidase, glycoprotease family(EC:3.4.24.57)	1019	18	0.017664377
Cthe_2144	-	DNA polymerase III, subunits gamma and tau(EC:2.7.7.7)	1643	29	0.017650639
Cthe_1955	-	RNA binding S1	2156	38	0.017625232
Cthe_3034	+	hypothetical protein	341	6	0.017595308
Cthe_0996	-	DNA polymerase III, alpha subunit(EC:2.7.7.7)	4328	76	0.017560074
Cthe_1072	-	putative stage IV sporulation YqfD	1196	21	0.017558528
Cthe_1296	-	hypothetical protein	686	12	0.017492711
Cthe_2676	+	GumN	1430	25	0.017482517
Cthe_0114	+	hypothetical protein	1319	23	0.017437453
Cthe_2435	+	electron transport complex, RnfABCDGE type, B subunit	803	14	0.017434620
Cthe_0645	+	homocysteine S-methyltransferase	2414	42	0.017398509
Cthe_2224	-	putative methionyl-tRNA formyltransferase	692	12	0.017341040
Cthe_2760	+	glycoside hydrolase, family 9	2885	50	0.017331023
Cthe_2996	+	intracellular protease, PfpI family	578	10	0.017301038

المنارة للاستشارات
www.manaraa.com

Cthe_2418	-	ATPase	1043	18	0.017257910
Cthe_0827	-	hemolysin A	812	14	0.017241379
Cthe_2692	+	glycosyl transferase, group 1	1103	19	0.017225748
Cthe_0253	+	ATPase associated with various cellular activities, AAA_3	989	17	0.017189080
Cthe_2347	-	lipopolysaccharide biosynthesis	1280	22	0.017187500
Cthe_2881	+	ATP phosphoribosyltransferase(EC:2.4.2.17)	641	11	0.017160686
Cthe_3014	-	hydrogenase formation HypD protein	1049	18	0.017159199
Cthe_1756	-	Putative virion core protein (lumpy skin disease virus)-like protein	1166	20	0.017152659
Cthe_1871	-	Tn7-like transposition protein D	1877	32	0.017048482
Cthe_0792	-	Protein of unknown function UPF0001	704	12	0.017045455
Cthe_1348	-	ribosomal protein S21	176	3	0.017045455
Cthe_0870	-	NADPH-dependent FMN reductase	587	10	0.017035775
Cthe_0258	+	cellulosome enzyme, dockerin type I	1409	24	0.017033357
Cthe_1509	-	protein of unknown function DUF438	1235	21	0.017004049
Cthe_0408	+	diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(s)	2588	44	0.017001546
Cthe_1806	-	cellulosome enzyme, dockerin type I	6533	111	0.016990663
Cthe_0265	+	aminotransferase, class V	1121	19	0.016949153
Cthe_1008	+	aminodeoxychorismate lyase	1121	19	0.016949153
Cthe_3167	+	glucose-1-phosphate adenylyltransferase, GlgD subunit(EC:2.7.7.27)	1121	19	0.016949153
Cthe_RF0078	+	(miscRNA)	236	4	0.016949153
Cthe_0455	+	Mg chelatase, subunit ChlI	1538	26	0.016905072
Cthe_3039	+	cell divisionFtsK/SpoIIIE	2492	42	0.016853933
Cthe_2186	-	single-strand binding protein	416	7	0.016826923
Cthe_0518	-	type III restriction enzyme, res subunit	2975	50	0.016806723
Cthe_2726	+	ribosomal protein L7Ae/L30e/S12e/Gadd45	239	4	0.016736402
Cthe_0930	-	Radical SAM	1016	17	0.016732283
Cthe_2742	+	Endopeptidase La(EC:3.4.21.53)	1676	28	0.016706444
Cthe_1258	-	copper amine oxidase-like protein	779	13	0.016688062
Cthe_2083	-	DNA-directed DNA polymerase(EC:2.7.7.7)	1199	20	0.016680567
Cthe_0298	+	methyl-accepting chemotaxis sensory transducer	1802	30	0.016648169
Cthe_2470	+	DNA adenine methylase(EC:2.1.1.72)	842	14	0.016627078
Cthe_2703	+	protein of unknown function DUF161	905	15	0.016574586
Cthe_3066	+	ABC transporter related protein	725	12	0.016551724
Cthe_1800	+	Peptidoglycan-binding LysM	1511	25	0.016545334
Cthe_0764	-	signal peptidase I	665	11	0.016541353
Cthe_1342	-	beta-lactamase-like protein	605	10	0.016528926
Cthe_2718	+	preprotein translocase, SecE subunit	242	4	0.016528926
Cthe_2991	+	NADH:flavin oxidoreductase/NADH oxidase	1937	32	0.016520392
Cthe_0487	+	Cobyrinic acid a,c-diamide synthase	908	15	0.016519824
Cthe_1562	+	protein of unknown function DUF214	1211	20	0.016515277
Cthe_2360	+	glycoside hydrolase, family 9	2786	46	0.016511127
Cthe_0493	+	CheD, stimulates methylation of MCP proteins	485	8	0.016494845
Cthe_1948	-	cytochrome c biogenesis protein, transmembrane region	668	11	0.016467066

المنارة للاستشارات
www.manaraa.com

Cthe_3105	+	exsB protein	668	11	0.016467066
Cthe_0231	+	Radical SAM	1337	22	0.016454749
Cthe_0542	+	protein of unknown function DUF45	731	12	0.016415869
Cthe_3013	-	hydrogenase expression/formation protein HypE	977	16	0.016376663
Cthe_0369	+	protein of unknown function DUF111	1223	20	0.016353230
Cthe_2341	-	glycosyl transferase, family 2	1166	19	0.016295026
Cthe_2890	+	putative transcriptional regulator	1169	19	0.016253208
Cthe_1517	-	type I phosphodiesterase/nucleotide pyrophosphatase	1109	18	0.016230839
Cthe_0608	+	peptidase M42	1049	17	0.016205910
Cthe_0200	+	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	1235	20	0.016194332
Cthe_0033	+	small GTP-binding protein	2666	43	0.016129032
Cthe_1142	-	hypothetical protein	1055	17	0.016113744
Cthe_1297	-	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase(EC:2.4.2.21)	1055	17	0.016113744
Cthe_1805	-	diguanylate cyclase	932	15	0.016094421
Cthe_2937	+	cobalt transport protein	809	13	0.016069221
Cthe_3001	+	protein of unknown function DUF47	623	10	0.016051364
Cthe_0244	-	heavy metal translocating P-type ATPase	2123	34	0.016015073
Cthe_2118	+	ABC transporter related protein	752	12	0.015957447
Cthe_0783	-	RNA modification enzyme, MiaB family	1442	23	0.015950069
Cthe_2194	-	Carbohydrate binding family 6	1505	24	0.015946844
Cthe_0775	-	tRNA delta(2)-isopentenylpyrophosphate transferase(EC:2.5.1.8)	941	15	0.015940489
Cthe_0307	+	hypothetical protein	377	6	0.015915119
Cthe_1027	-	protein of unknown function DUF177	503	8	0.015904573
Cthe_2346	-	O-antigen polymerase	3029	48	0.015846814
Cthe_1001	-	undecaprenyl diphosphate synthase(EC:2.5.1.31)	758	12	0.015831135
Cthe_1036	-	hypothetical protein	695	11	0.015827338
Cthe_1930	-	carboxyl-terminal protease(EC:3.4.21.102)	1517	24	0.015820699
Cthe_2386	-	VanW	1709	27	0.015798713
Cthe_1918	-	Orn/Lys/Arg decarboxylase, major region	1460	23	0.015753425
Cthe_0208	+	single-stranded-DNA-specific exonuclease RecJ	2291	36	0.015713662
Cthe_0893	-	protein of unknown function DUF633	704	11	0.015625000
Cthe_3214	+	hypothetical protein	1985	31	0.015617128
Cthe_1236	-	Fibronectin, type III	707	11	0.015558699
Cthe_0600	+	thiamine biosynthesis protein ThiF	644	10	0.015527950
Cthe_2050	-	CRISPR-associated RAMP protein, SSO1426 family	839	13	0.015494636
Cthe_2106	-	protein of unknown function DUF327	452	7	0.015486726
Cthe_1519	+	putative RNA methylase	1229	19	0.015459723
Cthe_2081	-	small GTP-binding protein	1748	27	0.015446224
Cthe_1399	-	C_GCAxxG_C_C family protein	518	8	0.015444015
Cthe_2040	-	ATP-dependent exonuclease synthesis protein AddB (superfamily I helicase)	3440	53	0.015406977
Cthe_2870	+	protein of unknown function DUF21	1301	20	0.015372790

www.manaraa.com

Cthe_1058	-	Glycine hydroxymethyltransferase(EC:2.1.2.1)	1238	19	0.015347334
Cthe_1592	+	Glycerate kinase(EC:2.7.1.31)	1175	18	0.015319149
Cthe_0886	-	DNA polymerase I(EC:2.7.7.7)	2684	41	0.015275708
Cthe_0112	+	UDP-N-acetylenolpyruvoylglucosamine reductase(EC:1.1.1.158)	917	14	0.015267176
Cthe_1554	-	methylated-DNA--protein-cysteine methyltransferase	527	8	0.015180266
Cthe_0889	-	putative transmembrane anti-sigma factor	989	15	0.015166835
Cthe_2105	-	DNA-directed DNA polymerase(EC:2.7.7.7)	989	15	0.015166835
Cthe_0270	+	glycoside hydrolase, family 18	1454	22	0.015130674
Cthe_1629	-	peptidase S14, ClpP	728	11	0.015109890
Cthe_2218	-	flagellar hook-associated 2-like protein	2516	38	0.015103339
Cthe_0688	+	putative transcriptional regulator	596	9	0.015100671
Cthe_0494	+	hypothetical protein	398	6	0.015075377
Cthe_0056	+	Ig-like, group 2	13892	209	0.015044630
Cthe_2534	+	sulfate ABC transporter, ATPase subunit(EC:3.6.3.30)	1064	16	0.015037594
Cthe_1853	-	cold-shock DNA-binding domain protein	200	3	0.015000000
Cthe_1576	+	basic membrane lipoprotein	1067	16	0.014995314
Cthe_1186	-	Alpha/beta hydrolase fold-3	935	14	0.014973262
Cthe_0540	+	protein of unknown function DUF214	2606	39	0.014965464
Cthe_0856	+	branched-chain amino acid aminotransferase(EC:2.6.1.42)	1070	16	0.014953271
Cthe_2029	-	hypothetical protein	803	12	0.014943960
Cthe_1055	+	protein of unknown function DUF58	1208	18	0.014900662
Cthe_0874	-	glutamine amidotransferase of anthranilate synthase(EC:4.1.3.27)	605	9	0.014876033
Cthe_1425	+	Inorganic diphosphatase(EC:3.6.1.1)	2084	31	0.014875240
Cthe_0289	+	DEAD_2	2357	35	0.014849385
Cthe_2471	+	hypothetical protein	1280	19	0.014843750
Cthe_2530	+	glutamate-1-semialdehyde-2,1-aminomutase(EC:5.4.3.8)	1283	19	0.014809041
Cthe_1785	-	protein of unknown function DUF147	1082	16	0.014787431
Cthe_0744	-	copper amine oxidase-like protein	2852	42	0.014726508
Cthe_2412	+	SMC protein-like protein	2651	39	0.014711430
Cthe_2940	+	transcriptional regulator, CarD family	476	7	0.014705882
Cthe_2011	-		887	13	0.014656144
Cthe_0803	+	pseudouridine synthase, RluA family	956	14	0.014644351
Cthe_1559	+	Cystathionine gamma-synthase(EC:2.5.1.48)	1163	17	0.014617369
Cthe_1201	-	purine nucleoside phosphorylase I, inosine and guanosine-specific(EC:2.4.2.1)	827	12	0.014510278
Cthe_0393	+	sugar ABC transporter (sugar-binding protein)	965	14	0.014507772
Cthe_0009	+	YD repeat containing protein	5522	80	0.014487505
Cthe_0316	+	PA14	1934	28	0.014477766
Cthe_0512	-	phage integrase	1520	22	0.014473684
Cthe_1189	+	ABC transporter related protein	968	14	0.014462810
Cthe_1181	-	transglutaminase-like protein	2282	33	0.014460999
Cthe_2254	-	hypoxanthine phosphoribosyltransferase(EC:2.4.2.8)	554	8	0.014440433
Cthe_0207	+	hypothetical protein	833	12	0.014405762

المنارة للاستشارات

www.manaraa.com

Cthe_3156	-	methyl-accepting chemotaxis sensory transducer	1250	18	0.014400000
Cthe_1278	-	ATP-dependent DNA helicase RecG	2084	30	0.014395393
Cthe_1136	-	serine/threonine protein kinase with TPR repeats	1460	21	0.014383562
Cthe_0980	-	hypothetical protein	488	7	0.014344262
Cthe_0598	+	thiazole biosynthesis	767	11	0.014341591
Cthe_2212	-	transcriptional regulator, AraC family	908	13	0.014317181
Cthe_1234	+	methyl-accepting chemotaxis sensory transducer	770	11	0.014285714
Cthe_2873	-	hypothetical protein	701	10	0.014265335
Cthe_3163	+	Carbohydrate binding family 25	491	7	0.014256619
Cthe_1577	+	hypothetical protein	1544	22	0.014248705
Cthe_2160	-	hypothetical protein	2393	34	0.014208107
Cthe_1858	-	peptidase M23B	1127	16	0.014196983
Cthe_2342	-	hypothetical protein	1058	15	0.014177694
Cthe_2747	+	PHP-like protein	848	12	0.014150943
Cthe_2995	+	hypothetical protein	212	3	0.014150943
Cthe_0042	+	small GTP-binding protein	1202	17	0.014143095
Cthe_0884	-	Lytic transglycosylase, catalytic	566	8	0.014134276
Cthe_1365	-	hypothetical protein	566	8	0.014134276
Cthe_0260	+	peptidase S1 and S6, chymotrypsin/Hap	2123	30	0.014130947
Cthe_0976	-	phospho-N-acetylmuramoyl-pentapeptide-transferase(EC:2.7.8.13)	992	14	0.014112903
Cthe_0094	+	Septum formation topological specificity factor MinE	284	4	0.014084507
Cthe_0872	-	Indole-3-glycerol-phosphate synthase(EC:4.1.1.48)	782	11	0.014066496
Cthe_2051	-	CRISPR-associated RAMP protein, SSO1426 family	782	11	0.014066496
Cthe_1571	+	binding-protein-dependent transport systems inner membrane component	1067	15	0.014058107
Cthe_2124	-		1353	19	0.014042868
Cthe_0643	+	hypothetical protein	1073	15	0.013979497
Cthe_2422	-	hypothetical protein	1073	15	0.013979497
Cthe_2749	+	hypothetical protein	1505	21	0.013953488
Cthe_2159	-	hypothetical protein	860	12	0.013953488
Cthe_2666	-	ATPase, P-type (transporting), HAD superfamily, subfamily IC	2594	36	0.013878180
Cthe_1287	-	periplasmic sensor signal transduction histidine kinase	1514	21	0.013870542
Cthe_2702	-	polysaccharide pyruvyl transferase	2237	31	0.013857845
Cthe_1337	-	type II secretion system protein	869	12	0.013808976
Cthe_1302	-	hypothetical protein	1667	23	0.013797241
Cthe_0944	-	SMC protein-like protein	1451	20	0.013783598
Cthe_0538	+	periplasmic sensor signal transduction histidine kinase	1016	14	0.013779528
Cthe_0544	+	protein serine/threonine phosphatases	875	12	0.013714286
Cthe_3023	-	respiratory-chain NADH dehydrogenase, subunit 1	875	12	0.013714286
Cthe_1032	-	glutamyl-tRNA(Gln) amidotransferase, A subunit(EC:3.5.1.4)	1460	20	0.013698630

المنارة للاستشارات

www.manaraa.com

Cthe_1572	+	ABC transporter related protein	803	11	0.013698630
Cthe_3064	+	polysaccharide biosynthesis protein	1607	22	0.013690106
Cthe_3237	+	Relaxase/mobilization nuclease domain containing protein	1391	19	0.013659238
Cthe_2340	-	UDP-glucose/GDP-mannose dehydrogenase	1319	18	0.013646702
Cthe_1932	-	S-layer-like domain containing protein	3005	41	0.013643927
Cthe_2896	+	hypothetical protein	953	13	0.013641133
Cthe_0110	+	HPr kinase	956	13	0.013598326
Cthe_1457	-	polar amino acid ABC transporter, inner membrane subunit	662	9	0.013595166
Cthe_0340	+	ferredoxin	368	5	0.013586957
Cthe_2811	+	glycoside hydrolase, family 26	1775	24	0.013521127
Cthe_1962	+	pseudouridine synthase, RluA family	962	13	0.013513514
Cthe_0672	+	pyrroline-5-carboxylate reductase(EC:1.5.1.2)	815	11	0.013496933
Cthe_0322	-	glycoside hydrolase, family 3-like protein	1334	18	0.013493253
Cthe_1834	-	hypothetical protein	593	8	0.013490725
Cthe_2432	+	electron transport complex, RnfABCDGE type, G subunit	593	8	0.013490725
Cthe_1184	-	hypothetical protein	668	9	0.013473054
Cthe_1574	-	hypothetical protein	743	10	0.013458950
Cthe_0626	+	hypothetical protein	1412	19	0.013456091
Cthe_0665	+	HflK protein	968	13	0.013429752
Cthe_1182	-	protein of unknown function DUF58	1268	17	0.013406940
Cthe_2142	-	recombination protein RecR	599	8	0.013355593
Cthe_0706	+	transcriptional regulator, RpiR family	899	12	0.013348165
Cthe_0825	-	glycoside hydrolase, family 9	1949	26	0.013340174
Cthe_0741	-	adenylosuccinate lyase(EC:4.3.2.2)	1427	19	0.013314646
Cthe_2058	-	Appr-1-p processing protein	527	7	0.013282732
Cthe_2184	+	putative ribonuclease BN	830	11	0.013253012
Cthe_1292	+	proposed homoserine kinase(EC:5.4.2.1)	1208	16	0.013245033
Cthe_0458	+	exodeoxyribonuclease III Xth(EC:4.2.99.18)	755	10	0.013245033
Cthe_0745	+	glycoside hydrolase, family 9	2192	29	0.013229927
Cthe_1560	+	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit(EC:2.5.1.47)	908	12	0.013215859
Cthe_2242	-	flagellar hook-associated protein 3	908	12	0.013215859
Cthe_1165	-	YbbR-like protein	1211	16	0.013212221
Cthe_2974	-	hypothetical protein	1211	16	0.013212221
Cthe_1766	-	glutamate 5-kinase(EC:2.7.2.11)	833	11	0.013205282
Cthe_0332	+	phosphoribulokinase/uridine kinase	1673	22	0.013150030
Cthe_1780	-	hypothetical protein	533	7	0.013133208
Cthe_2734	+	VanW	839	11	0.013110846
Cthe_2121	-	hypothetical protein	917	12	0.013086150
Cthe_2936	+	ABC transporter related protein	845	11	0.013017751
Cthe_3007	-	ErfK/YbiS/YcfS/YnhG	692	9	0.013005780
Cthe_3069	+	two component transcriptional regulator, winged helix family	692	9	0.013005780
Cthe_1410	-	cation diffusion facilitator family transporter	1001	13	0.012987013
Cthe_1411	-	tryptophan synthase, alpha subunit(EC:4.2.1.20)	770	10	0.012987013

المنارة للاستشارات

www.manaraa.com

Cthe_3104	+	putative CoA-substrate-specific enzyme activase	770	10	0.012987013
Cthe_0278	+	pseudouridine synthase	695	9	0.012949640
Cthe_1617	-	phage putative tail component	773	10	0.012936611
Cthe_3052	+	YD repeat protein	8828	114	0.012913457
Cthe_2182	+	Ig-like, group 2	2789	36	0.012907852
Cthe_1631	-	hypothetical protein	1784	23	0.012892377
Cthe_2263	+	H+-transporting two-sector ATPase, C subunit	467	6	0.012847966
Cthe_1075	-	protein of unknown function DUF881	701	9	0.012838802
Cthe_1357	-	glycosyl transferase, group 1	1094	14	0.012797075
Cthe_0481	+	flagellar biosynthetic protein FliP	782	10	0.012787724
Cthe_1847	-	two component transcriptional regulator, winged helix family	704	9	0.012784091
Cthe_0040	+	Cellulase., Cellulose 1,4-beta-cellobiosidase(EC:3.2.1.4,EC:3.2.1.91)	2663	34	0.012767555
Cthe_1202	-	major facilitator superfamily MFS_1	1175	15	0.012765957
Cthe_2207	-	isoaspartyl dipeptidase	1175	15	0.012765957
Cthe_1671	+	Recombinase	1568	20	0.012755102
Cthe_0237	-	response regulator receiver protein	392	5	0.012755102
Cthe_3047	+	Peptidoglycan glycosyltransferase(EC:2.4.1.129)	1412	18	0.012747875
Cthe_1145	-	N-6 DNA methylase	1886	24	0.012725345
Cthe_2762	+	transcriptional regulator-like protein	1337	17	0.012715034
Cthe_0912	-	glycoside hydrolase, family 10	3233	41	0.012681720
Cthe_1983	-	transposase IS116/IS110/IS902	1184	15	0.012668919
Cthe_1423	-	protein of unknown function UPF0118	1106	14	0.012658228
Cthe_1771	+	Rubrerythrin	554	7	0.012635379
Cthe_2941	+	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase(EC:2.7.7.60)	713	9	0.012622721
Cthe_2251	-	Methionine adenosyltransferase(EC:2.5.1.6)	1193	15	0.012573345
Cthe_2427	+	protein of unknown function DUF1385	956	12	0.012552301
Cthe_2312	-	hypothetical protein	479	6	0.012526096
Cthe_1580	+	inner-membrane translocator	1118	14	0.012522361
Cthe_2087	-	metalloenzyme	884	11	0.012443439
Cthe_2301	-	CRISPR-associated autoregulator, DevR family	884	11	0.012443439
Cthe_3075	+	von Willebrand factor, type A	1610	20	0.012422360
Cthe_0606	+	DNA internalization-related competence protein ComEC/Rec2	2336	29	0.012414384
Cthe_2217	-	flagellar protein FliS	404	5	0.012376238
Cthe_0938	-	regulatory protein, DeoR	566	7	0.012367491
Cthe_RF0080	-	(miscRNA)	647	8	0.012364760
Cthe_0059	+	type 3a, cellulose-binding	1460	18	0.012328767
Cthe_0602	+	thiamine biosynthesis protein ThiC	1298	16	0.012326656
Cthe_0609	+	peptidase M42(EC:3.2.1.4)	977	12	0.012282497
Cthe_0630	+	hypothetical protein	977	12	0.012282497
Cthe_0854	-	shikimate 5-dehydrogenase(EC:1.1.1.25)	896	11	0.012276786
Cthe_2244	-	flagellar hook-associated protein FlgK	1469	18	0.012253233
Cthe_1798	-	CoA-binding protein	653	8	0.012251149
Cthe_1688	-	Radical SAM	980	12	0.012244898

المنارة للاستشارات
www.manaraa.com

Cthe_0431	+	hypothetical protein	245	3	0.012244898
Cthe_1303	-	glycosyl transferase, group 1	1226	15	0.012234910
Cthe_1397	-	glycosyl transferase, family 8	818	10	0.012224939
Cthe_2305	-	putative undecaprenol kinase(EC:2.7.1.66)	818	10	0.012224939
Cthe_1915	-	periplasmic sensor signal transduction histidine kinase	1391	17	0.012221423
Cthe_0828	-	deoxyxylulose-5-phosphate synthase(EC:2.2.1.7)	1883	23	0.012214551
Cthe_2802	+	NLPA lipoprotein	983	12	0.012207528
Cthe_2088	-	protein of unknown function DUF710	410	5	0.012195122
Cthe_2243	-	flagellar hook-associated protein FlgK	1559	19	0.012187300
Cthe_2338	-	glycosyl transferase, group 1	1232	15	0.012175325
Cthe_3101	+	aminotransferase, class I and II	1232	15	0.012175325
Cthe_2767	+	Methyltransferase type 12	1397	17	0.012168933
Cthe_0252	+	hypothetical protein	1151	14	0.012163336
Cthe_2380	+	TPR repeat domain containing protein	740	9	0.012162162
Cthe_0694	+	spermidine synthase(EC:2.5.1.16)	827	10	0.012091898
Cthe_1865	+	type IV pilus assembly PilZ	662	8	0.012084592
Cthe_2693	+	hypothetical protein	497	6	0.012072435
Cthe_2680	-	peptidase S41	2237	27	0.012069736
Cthe_2799	+	Cystathionine gamma-synthase(EC:2.5.1.48,EC:4.4.1.1)	1160	14	0.012068966
Cthe_1141	-	hypothetical protein	4061	49	0.012065994
Cthe_1700	-	Recombinase	1328	16	0.012048193
Cthe_2053	-	hypothetical protein	1247	15	0.012028869
Cthe_3186	+	Radical SAM	1751	21	0.011993147
Cthe_1794	-	translation elongation factor G	2087	25	0.011978917
Cthe_1007	+	GTP-binding protein TypA	1844	22	0.011930586
Cthe_1066	-	DNA repair protein RecO	755	9	0.011920530
Cthe_3117	+	zinc/iron permease	755	9	0.011920530
Cthe_3114	+	glycosyl transferase, group 1	1175	14	0.011914894
Cthe_1294	-	Recombinase	1427	17	0.011913104
Cthe_1219	+	hypothetical protein	1094	13	0.011882998
Cthe_2821	+	response regulator receiver modulated CheB methylesterase(EC:3.1.1.61)	1094	13	0.011882998
Cthe_2101	-	Protein of unknown function UPF0011	842	10	0.011876485
Cthe_3100	+	Diaminopimelate epimerase(EC:5.1.1.7)	842	10	0.011876485
Cthe_2900	+	RNA polymerase, sigma 28 subunit	758	9	0.011873351
Cthe_1434	-	hypothetical protein	590	7	0.011864407
Cthe_2800	+	aminotransferase, class I and II	1181	14	0.011854361
Cthe_2042	-		928	11	0.011853448
Cthe_0003	+	ankyrin repeat protein	929	11	0.011840689
Cthe_2472	+	hypothetical protein	845	10	0.011834320
Cthe_1512	-	Type II site-specific deoxyribonuclease(EC:3.1.21.4)	932	11	0.011802575
Cthe_0229	+	NAD-dependent epimerase/dehydratase(EC:5.1.3.12)	1019	12	0.011776251
Cthe_1062	-	VanW	1445	17	0.011764706
Cthe_0928	-	4-hydroxythreonine-4-phosphate dehydrogenase(EC:1.1.1.262)	1022	12	0.011741683

المنارة للاستشارات
www.manaraa.com

Cthe_0135	+	beta-ketoacyl synthase	8264	97	0.011737657
Cthe_2143	-	hypothetical protein	341	4	0.011730205
Cthe_0092	+	septum site-determining protein MinC	683	8	0.011713031
Cthe_0549	+	ABC-3	854	10	0.011709602
Cthe_1573	+	nitrogenase iron protein(EC:1.18.6.1)	854	10	0.011709602
Cthe_0605	+	NLP/P60	1112	13	0.011690647
Cthe_0357	-	alpha-glucan phosphorylases(EC:2.4.1.1)	2567	30	0.011686794
Cthe_0621	-	putative translation initiation factor, aIF-2BI family(EC:5.3.1.23)	1028	12	0.011673152
Cthe_2108	-	Orn/Lys/Arg decarboxylase, major region	1457	17	0.011667811
Cthe_1061	-	hypothetical protein	686	8	0.011661808
Cthe_2684	+	hypothetical protein	515	6	0.011650485
Cthe_1857	-	carboxyl-terminal protease(EC:3.4.21.102)	1289	15	0.011636928
Cthe_1500	-	ABC transporter related protein	1892	22	0.011627907
Cthe_2818	+	CheA signal transduction histidine kinases	1808	21	0.011615044
Cthe_0411	-	hypothetical protein	1034	12	0.011605416
Cthe_2998	+	ABC transporter related protein	1727	20	0.011580776
Cthe_2495	+	hypothetical protein	1382	16	0.011577424
Cthe_2808	+	transcriptional regulator, LacI family	1037	12	0.011571842
Cthe_0579	+	nitroreductase	521	6	0.011516315
Cthe_0695	+	putative agmatinase	869	10	0.011507480
Cthe_1899	-	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	869	10	0.011507480
Cthe_0547	+	periplasmic solute binding protein	956	11	0.011506276
Cthe_1799	+	ABC transporter related protein	1916	22	0.011482255
Cthe_1546	-	4Fe-4S ferredoxin, iron-sulfur binding	1133	13	0.011473963
Cthe_1472	+	Carbohydrate binding family 11	2702	31	0.011472983
Cthe_1753	-	transport system permease protein	1046	12	0.011472275
Cthe_0414	-	manganese containing catalase	611	7	0.011456628
Cthe_0331	+	hypothetical protein	437	5	0.011441648
Cthe_0834	-	NusB antitermination factor	437	5	0.011441648
Cthe_1275	-	H+-ATPase subunit H	437	5	0.011441648
Cthe_2699	+	transcriptional modulator of MazE/toxin, MazF	350	4	0.011428571
Cthe_3226	+	copper amine oxidase-like protein	788	9	0.011421320
Cthe_2449	+	Phosphoglycerate mutase	614	7	0.011400651
Cthe_2284	-	CheA signal transduction histidine kinases	2108	24	0.011385199
Cthe_0035	+	4Fe-4S ferredoxin, iron-sulfur binding	704	8	0.011363636
Cthe_2140	-	hypothetical protein	1322	15	0.011346445
Cthe_0916	+	hypothetical protein	884	10	0.011312217
Cthe_1197	-	intein	1415	16	0.011307420
Cthe_0975	-	stage V sporulation protein E	1151	13	0.011294526
Cthe_1241	-	nicotinate (nicotinamide) nucleotide adenylyltransferase(EC:2.7.7.18)	620	7	0.011290323
Cthe_1743	-	protein of unknown function DUF955	710	8	0.011267606
Cthe_2419	-	primosome, DnaD subunit	977	11	0.011258956
Cthe_1752	-	ABC transporter related protein	800	9	0.011250000
Cthe_1980	-	hypothetical protein	800	9	0.011250000
Cthe_2475	+	phage portal protein, SPP1 family	1433	16	0.011165387

www.manaraa.com

Cthe_1613	-	glycosyl hydrolase-like protein	2423	27	0.011143211
Cthe_0334	+	putative anti-sigma regulatory factor, serine/threonine protein kinase	449	5	0.011135857
Cthe_1343	-	D-tyrosyl-tRNA(Tyr) deacylase	449	5	0.011135857
Cthe_3020	-	NADH dehydrogenase (ubiquinone)(EC:1.6.5.3)	1079	12	0.011121409
Cthe_3162	+	SAM dependent methyltransferase	1169	13	0.011120616
Cthe_0638	-	FHA domain containing protein	4679	52	0.011113486
Cthe_3012	+	Carbohydrate binding family 6	1892	21	0.011099366
Cthe_0786	-	3-dehydroquinate synthase(EC:4.2.3.4)	1082	12	0.011090573
Cthe_2539	+	UBA/THIF-type NAD/FAD binding fold	812	9	0.011083744
Cthe_1259	-	CDP-alcohol phosphatidyltransferase	542	6	0.011070111
Cthe_1221	-	glycosyltransferase 36	8768	97	0.011062956
Cthe_3016	-	(NiFe) hydrogenase maturation protein HypF	2261	25	0.011057054
Cthe_2195	-	Carbohydrate binding family 6	2897	32	0.011045910
Cthe_0355	+	hypothetical protein	1178	13	0.011035654
Cthe_0151	+	protein of unknown function DUF965	272	3	0.011029412
Cthe_3232	+	YD repeat protein	5009	55	0.010980236
Cthe_0583	-	response regulator receiver protein	1184	13	0.010979730
Cthe_3043	+	hypothetical protein	638	7	0.010971787
Cthe_2320	-	DNA-cytosine methyltransferase(EC:2.1.1.37)	1004	11	0.010956175
Cthe_0836	-	hypothetical protein	548	6	0.010948905
Cthe_2057	-	CRISPR-associated protein, TM1812 family	1832	20	0.010917031
Cthe_3155	+	beta-lactamase-like protein	734	8	0.010899183
Cthe_2981	-	hypothetical protein	551	6	0.010889292
Cthe_2131	-	hypothetical protein	368	4	0.010869565
Cthe_0892	-	protein of unknown function DUF34	1106	12	0.010849910
Cthe_2761	+	glycoside hydrolase, family 9	2123	23	0.010833726
Cthe_0288	+	MCP methyltransferase, CheR-type(EC:2.1.1.80)	1109	12	0.010820559
Cthe_2374	+	DNA replication and repair protein RecF	1109	12	0.010820559
Cthe_1453	-		185	2	0.010810811
Cthe_3030	+	methyl-accepting chemotaxis sensory transducer	2498	27	0.010808647
Cthe_1051	-	integral membrane protein MviN	1577	17	0.010779962
Cthe_1599	-	multi-sensor signal transduction histidine kinase	1670	18	0.010778443
Cthe_2025	-	Resolvase helix-turn-helix region containing protein	557	6	0.010771993
Cthe_3053	+	hypothetical protein	557	6	0.010771993
Cthe_RF0073	+	(miscRNA)	279	3	0.010752688
Cthe_3183	+	TrkA-N	1397	15	0.010737294
Cthe_1355	-	glycosyl transferase, family 2	932	10	0.010729614
Cthe_1390	+	alpha/beta hydrolase fold	932	10	0.010729614
Cthe_0261	+	HAD-superfamily hydrolase, subfamily IA, variant 1(EC:3.1.3.18)	653	7	0.010719755
Cthe_2669	+	hypothetical protein	653	7	0.010719755
Cthe_3049	-	TPR repeat domain containing protein	1775	19	0.010704225
Cthe_0725	+	response regulator receiver modulated metal dependent phosphohydrolase	1496	16	0.010695187
Cthe_1602	-	phosphate ABC transporter, ATPase subunit(	749	8	0.010680908

المنارة للاستشارات

www.manaraa.com

		EC:3.6.3.27)			
Cthe_1849	-	Heavy metal transport/detoxification protein	1499	16	0.010673783
Cthe_0871	-	Phosphoribosylanthranilate isomerase(EC:5.3.1.24)	656	7	0.010670732
Cthe_2807	+	glycoside hydrolase, family 5	1031	11	0.010669253
Cthe_0370	+	conserved hypothetical radical SAM protein	941	10	0.010626993
Cthe_2016	-		659	7	0.010622155
Cthe_2196	-	Carbohydrate binding family 6	1601	17	0.010618364
Cthe_0511	+	histidine kinase	1322	14	0.010590015
Cthe_0612	+	SSS sodium solute transporter superfamily	1511	16	0.010589014
Cthe_0759	-	flagellar biosynthesis	284	3	0.010563380
Cthe_2339	-	glycosyl transferase, group 1	1043	11	0.010546500
Cthe_1428	-	glycoside hydrolase, family 1	1328	14	0.010542169
Cthe_2006	-		665	7	0.010526316
Cthe_2784	+	cobalamin B12-binding protein	665	7	0.010526316
Cthe_1325	-	putative oxygen-independent coproporphyrinogen III oxidase	1142	12	0.010507881
Cthe_1707	-	phage minor structural protein	1904	20	0.010504202
Cthe_0374	+	Glutamate dehydrogenase (NADP(+))(EC:1.4.1.4)	1334	14	0.010494753
Cthe_2411	+	metallophosphoesterase	1145	12	0.010480349
Cthe_1052	-	competence/damage-inducible protein CinA	1241	13	0.010475423
Cthe_2353	+	PGAP1-like protein	1148	12	0.010452962
Cthe_1539	-	glutamine synthetase, catalytic region	2105	22	0.010451306
Cthe_0362	+	transcriptional regulator, AsnC family	479	5	0.010438413
Cthe_1563	+	ABC transporter related protein	671	7	0.010432191
Cthe_2573	+	ABC transporter related protein	767	8	0.010430248
Cthe_1048	-	Rhomboid-like protein	1535	16	0.010423453
Cthe_3063	+	Acetyl xylan esterase	962	10	0.010395010
Cthe_0058	+	RNA polymerase sigma factor	770	8	0.010389610
Cthe_0557	+	polysaccharide deacetylase	770	8	0.010389610
Cthe_1455	-	hypothetical protein	674	7	0.010385757
Cthe_1600	-	two component transcriptional regulator, winged helix family	674	7	0.010385757
Cthe_0321	+	TPR repeat domain containing protein	965	10	0.010362694
Cthe_1068	-	cytidine deaminase	386	4	0.010362694
Cthe_2779	+	hypothetical protein	386	4	0.010362694
Cthe_3006	-	ErfK/YbiS/YcfS/YnhG	869	9	0.010356732
Cthe_0213	+	protein of unknown function DUF975	773	8	0.010349288
Cthe_2069	+	hypothetical protein	776	8	0.010309278
Cthe_0633	+	hypothetical protein	485	5	0.010309278
Cthe_0677	+	phosphopentomutase(EC:5.4.2.7)	1166	12	0.010291595
Cthe_1282	-	diguanylate cyclase with GAF sensor	1652	17	0.010290557
Cthe_0076	+	hypothetical protein	389	4	0.010282776
Cthe_3118	+	hemerythrin-like metal-binding protein	389	4	0.010282776
Cthe_2895	+	glycoside hydrolase, family 18	1751	18	0.010279840
Cthe_2064	+	spore germination B3 GerAC like	1271	13	0.010228167
Cthe_2387	-	peptidase M14, carboxypeptidase A	1271	13	0.010228167

المنارة للاستشارات
www.manaraa.com

Cthe_2434	+	electron transport complex, RnfABCDGE type, A subunit	587	6	0.010221465
Cthe_0404	+	type 3a, cellulose-binding	1568	16	0.010204082
Cthe_0778	-	hypothetical protein	392	4	0.010204082
Cthe_2899	+	putative anti-sigma regulatory factor, serine/threonine protein kinase	392	4	0.010204082
Cthe_2073	+	amidohydrolase(EC:3.5.1.14)	1181	12	0.010160881
Cthe_2190	-	N-acetylglucosamine-6-phosphate deacetylase(EC:3.5.1.25)	1181	12	0.010160881
Cthe_1914	-	TPR repeat domain containing protein	1280	13	0.010156250
Cthe_0039	+	methyl-accepting chemotaxis sensory transducer	1871	19	0.010154997
Cthe_2068	-	8-oxoguanine DNA glycosylase-like protein	887	9	0.010146561
Cthe_2528	+	uroporphyrin-III C-methyltransferase	1481	15	0.010128292
Cthe_0095	+	methylglyoxal synthase(EC:4.2.3.3)	395	4	0.010126582
Cthe_1486	-	NUDIX hydrolase	395	4	0.010126582
Cthe_1511	-	DNA methylase N-4/N-6	989	10	0.010111223
Cthe_2417	+	Abortive infection protein	1088	11	0.010110294
Cthe_1067	-	GTP-binding protein Era	893	9	0.010078387
Cthe_1253	-	phosphoenolpyruvate synthase	2582	26	0.010069713
Cthe_1239	-	hypothetical protein	596	6	0.010067114
Cthe_2416	+	Propeptide, PepSY amd peptidase M4	1394	14	0.010043042
Cthe_0174	+	sulfatase	1892	19	0.010042283
Cthe_0196	-	glutamine synthetase, catalytic region	2093	21	0.010033445
Cthe_1855	-	Methyltransferase type 11	698	7	0.010028653
Cthe_0632	+	hypothetical protein	602	6	0.009966777
Cthe_2862	+	hypothetical protein	602	6	0.009966777
Cthe_0736	-	cellulosome anchoring protein, cohesin region	3917	39	0.009956599
Cthe_2797	+	thiamine pyrophosphate enzyme-like TPP-binding	905	9	0.009944751
Cthe_0855	-	HAD superfamily (subfamily IIIA) phosphatase, TIGR01668	503	5	0.009940358
Cthe_3036	+	methyl-accepting chemotaxis sensory transducer	1712	17	0.009929907
Cthe_0729	+	cellulosome enzyme, dockerin type I	1613	16	0.009919405
Cthe_2428	+	modification methylase, HemK family	908	9	0.009911894
Cthe_1906	-	glycosyl transferase, family 28	1211	12	0.009909166
Cthe_2664	+	ABC-1	1721	17	0.009877978
Cthe_2238	-	aldehyde dehydrogenase	1418	14	0.009873061
Cthe_1929	-	hypothetical protein	608	6	0.009868421
Cthe_1630	-	phage portal protein, HK97 family	1322	13	0.009833585
Cthe_0226	+	hypothetical protein	407	4	0.009828010
Cthe_2379	+	hypothetical protein	509	5	0.009823183
Cthe_0068	-	Peptidylprolyl isomerase(EC:5.2.1.8)	611	6	0.009819967
Cthe_1927	-	hypothetical protein	815	8	0.009815951
Cthe_3212	+	protein of unknown function DUF324	1331	13	0.009767092
Cthe_2977	+	MAEBL, putative	1127	11	0.009760426
Cthe_0246	+	Carbohydrate binding family 6	2462	24	0.009748172
Cthe_2450	-	hypothetical protein	308	3	0.009740260
Cthe_2712	+	hypothetical protein	719	7	0.009735744

المنارة للاستشارات
www.manaraa.com

Cthe_0282	+	glycogen/starch synthases, ADP-glucose type(EC:2.4.1.21)	1442	14	0.009708738
Cthe_1831	-	hypothetical protein	1031	10	0.009699321
Cthe_0861	+	cell envelope-related transcriptional attenuator	1136	11	0.009683099
Cthe_3084	-	putative RNA methylase	1136	11	0.009683099
Cthe_0130	+	3-oxoacyl-(acyl-carrier-protein) synthase(EC:2.3.1.41)	1034	10	0.009671180
Cthe_0839	-	stage III sporulation protein AG	623	6	0.009630819
Cthe_0534	+	ABC-type bacteriocin transporter	2183	21	0.009619789
Cthe_1523	+	glycosyl transferase, group 1	1880	18	0.009574468
Cthe_2214	+	hypothetical protein	941	9	0.009564293
Cthe_0707	+	Phosphoglycerate mutase	629	6	0.009538951
Cthe_1409	+	periplasmic sensor signal transduction histidine kinase	1469	14	0.009530293
Cthe_1909	-	copper amine oxidase-like protein	1472	14	0.009510870
Cthe_2215	-	Mg2+ transporter protein, CorA-like protein	947	9	0.009503696
Cthe_0057	-	hypothetical protein	632	6	0.009493671
Cthe_2061	-	hypothetical protein	422	4	0.009478673
Cthe_1304	-	PhoH-like protein	1373	13	0.009468318
Cthe_1933	-	Zn-dependent hydrolase of the beta-lactamase fold	635	6	0.009448819
Cthe_2801	+	carbon-monoxide dehydrogenase, catalytic subunit(EC:1.2.99.2)	2012	19	0.009443340
Cthe_3005	+	Peptidoglycan-binding LysM	1271	12	0.009441385
Cthe_2331	-	hypothetical protein	530	5	0.009433962
Cthe_2172	-	ribosomal protein L31	212	2	0.009433962
Cthe_0210	-	transcriptional regulator, LacI family	1061	10	0.009425071
Cthe_0394	+	iron-containing alcohol dehydrogenase	1169	11	0.009409752
Cthe_0747	-	extracellular solute-binding protein, family 1	1064	10	0.009398496
Cthe_2222	-	glycosyltransferase 28-like protein	1064	10	0.009398496
Cthe_0610	+	histidinol-phosphate aminotransferase	1067	10	0.009372071
Cthe_1605	-	phosphate ABC transporter (binding protein)-like protein	854	8	0.009367681
Cthe_1074	-	polysaccharide deacetylase	962	9	0.009355509
Cthe_2429	-	protein of unknown function UPF0118	1070	10	0.009345794
Cthe_0005	+	hypothetical protein	857	8	0.009334889
Cthe_1309	+	Radical SAM	1073	10	0.009319664
Cthe_1816	-	urease, alpha subunit(EC:3.5.1.5)	1718	16	0.009313155
Cthe_1724	-	hypothetical protein	752	7	0.009308511
Cthe_2543	-	Spore germination protein-like protein	1397	13	0.009305655
Cthe_1836	-	hypothetical protein	968	9	0.009297521
Cthe_1825	-	multi-sensor hybrid histidine kinase	2798	26	0.009292352
Cthe_0147	+	protein of unknown function DUF523	431	4	0.009280742
Cthe_0519	-	DNA methylase N-4/N-6	1832	17	0.009279476
Cthe_0864	-	pyruvate ferredoxin/flavodoxin oxidoreductase	539	5	0.009276438
Cthe_2102	-	methyltransferase small	755	7	0.009271523
Cthe_2602	+	ATP synthase F0, A subunit	755	7	0.009271523
Cthe_0848	-	peptidase M24	1079	10	0.009267841
Cthe_2796	+	pyruvate flavodoxin/ferredoxin oxidoreductase-like protein	1187	11	0.009267060

المنارة للاستشارات

www.manaraa.com

Cthe_1492	-	NAD(P)H dehydrogenase (quinone)	1622	15	0.009247842
Cthe_3194	-	hypothetical protein	650	6	0.009230769
Cthe_0580	-	aminotransferase, class I and II	1193	11	0.009220453
Cthe_1501	-	ABC transporter, transmembrane region	1736	16	0.009216590
Cthe_2611	+	Fibronectin, type III	4886	45	0.009209988
Cthe_0019	+	regulatory protein, P-II family	326	3	0.009202454
Cthe_0277	+	hypothetical protein	875	8	0.009142857
Cthe_0683	+	diaminopimelate decarboxylase(EC:4.1.1.20)	1313	12	0.009139375
Cthe_0533	+	Radical SAM	1532	14	0.009138381
Cthe_2822	+	oxidoreductase-like protein	986	9	0.009127789
Cthe_1944	-	protein of unknown function DUF458	548	5	0.009124088
Cthe_2094	-	hypothetical protein	551	5	0.009074410
Cthe_2837	+	prophage Lp4 protein 7, DNA replication	551	5	0.009074410
Cthe_1115	-	Tn7-like transposition protein C	1655	15	0.009063444
Cthe_1393	-	multi-sensor signal transduction histidine kinase	1769	16	0.009044658
Cthe_2157	-	hypothetical protein	332	3	0.009036145
Cthe_1748	-	hypothetical protein	1553	14	0.009014810
Cthe_1273	+	alpha-L-arabinofuranosidase B	1445	13	0.008996540
Cthe_1358	-	glycosyltransferase	890	8	0.008988764
Cthe_1584	-	two component transcriptional regulator, AraC family	779	7	0.008985879
Cthe_1056	+	transglutaminase-like protein	2672	24	0.008982036
Cthe_2200	-	hypothetical protein	335	3	0.008955224
Cthe_0015	+	alpha-L-arabinofuranosidase B	2123	19	0.008949600
Cthe_2063	+	spore germination protein	1121	10	0.008920607
Cthe_2656	+	MazG family protein	785	7	0.008917197
Cthe_1010	+	peptidase U32	1235	11	0.008906883
Cthe_0241	+	hypothetical protein	674	6	0.008902077
Cthe_0488	+	type IV pilus assembly PilZ	674	6	0.008902077
Cthe_2508	+	TOPRIM domain containing protein	899	8	0.008898776
Cthe_1538	-	transcriptional regulator, XRE family	1124	10	0.008896797
Cthe_2537	+	sulfate adenylyltransferase, large subunit(EC:2.7.7.4)	1799	16	0.008893830
Cthe_0065	+	hypothetical protein	338	3	0.008875740
Cthe_0806	-	PAS/PAC sensor hybrid histidine kinase	2708	24	0.008862629
Cthe_0799	+	two component transcriptional regulator, winged helix family	677	6	0.008862629
Cthe_2590	+	glycoside hydrolase, family 10	1919	17	0.008858781
Cthe_1498	-	methyl-accepting chemotaxis sensory transducer	1808	16	0.008849558
Cthe_2824	+	hypothetical protein	791	7	0.008849558
Cthe_2928	+	ribosomal protein L36	113	1	0.008849558
Cthe_0284	+	protein of unknown function DUF815	1358	12	0.008836524
Cthe_1271	+	Carbohydrate binding family 6	2039	18	0.008827857
Cthe_0296	+	hypothetical protein	1475	13	0.008813559
Cthe_2778	+	cytidyltransferase-related domain	1022	9	0.008806262
Cthe_1240	-	metal dependent phosphohydrolase	569	5	0.008787346
Cthe_1913	-	hypothetical protein	797	7	0.008782936

www.manaraa.com

Cthe_1116	-	HMG-I and HMG-Y, DNA-binding	2165	19	0.008775982
Cthe_1040	-	DNA polymerase III, delta subunit	1028	9	0.008754864
Cthe_2356	-	quinolinate synthetase complex, A subunit	914	8	0.008752735
Cthe_1916	-	two component transcriptional regulator, winged helix family	686	6	0.008746356
Cthe_1372	-	hypothetical protein	458	4	0.008733624
Cthe_2488	+	phage tape measure protein	1835	16	0.008719346
Cthe_1821	-	inner-membrane translocator	1148	10	0.008710801
Cthe_3074	+	Cof-like hydrolase	920	8	0.008695652
Cthe_0235	+	Glutaredoxin-like protein, YruB-family	230	2	0.008695652
Cthe_2599	+	CMP/dCMP deaminase, zinc-binding	461	4	0.008676790
Cthe_1232	+	AMP-dependent synthetase and ligase	1730	15	0.008670520
Cthe_0318	-	ErfK/YbiS/YcfS/YnhG	692	6	0.008670520
Cthe_0272	-	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	1154	10	0.008665511
Cthe_0047	+	copper amine oxidase-like protein	809	7	0.008652658
Cthe_0812	-	response regulator receiver protein	809	7	0.008652658
Cthe_0878	-	CheW protein	578	5	0.008650519
Cthe_2945	+	histidine kinase	1850	16	0.008648649
Cthe_1039	+	ribosomal protein S20	347	3	0.008645533
Cthe_1139	-	hypothetical protein	1160	10	0.008620690
Cthe_0660	+	glycoside hydrolase, family 81	2321	20	0.008616975
Cthe_0350	+	signal peptidase I(EC:3.4.21.89)	581	5	0.008605852
Cthe_1919	-	MgtC/SapB transporter	698	6	0.008595989
Cthe_2344	-	Undecaprenyl-phosphate galactose phosphotransferase(EC:2.7.8.6)	698	6	0.008595989
Cthe_0640	+	cellulosome enzyme, dockerin type I	1748	15	0.008581236
Cthe_2147	+	glycoside hydrolase, family 5	1982	17	0.008577195
Cthe_2622	+	hypothetical protein	350	3	0.008571429
Cthe_2236	-	flagellin-like protein	818	7	0.008557457
Cthe_1581	+	inner-membrane translocator	935	8	0.008556150
Cthe_2229	-	N-acetylneuraminate synthase(EC:2.5.1.56)	1052	9	0.008555133
Cthe_2221	-	DegT/DnrJ/EryC1/StrS aminotransferase	1169	10	0.008554320
Cthe_0142	+	metal dependent phosphohydrolase	1637	14	0.008552230
Cthe_2810	+	Na/Pi-cotransporter II-related protein	1754	15	0.008551881
Cthe_1207	+	membrane protein-like protein	1523	13	0.008535785
Cthe_2504	-	hypothetical protein	1289	11	0.008533747
Cthe_0953	-	aspartate carbamoyltransferase(EC:2.1.3.2)	938	8	0.008528785
Cthe_1953	+	cell wall hydrolase, SleB	704	6	0.008522727
Cthe_2626	-	UDP-N-acetylmuramate--alanine ligase(EC:6.3.2.8)	1409	12	0.008516678
Cthe_1424	+	hypothetical protein	941	8	0.008501594
Cthe_0376	+	transcriptional repressor, CopY family	353	3	0.008498584
Cthe_0992	-	ribosomal protein L7Ae/L30e/S12e/Gadd45	353	3	0.008498584
Cthe_1223	-	ribosomal protein L20	353	3	0.008498584
Cthe_2249	-	helicase, RecD/TraA family(EC:3.1.11.5)	2240	19	0.008482143
Cthe_0212	-	Beta-glucosidase(EC:3.2.1.21)	1415	12	0.008480565
Cthe_2139	+	alpha-L-arabinofuranosidase B	2948	25	0.008480326

المنارة للاستشارات

www.manaraa.com

Cthe_2220	-	NAD-dependent epimerase/dehydratase	944	8	0.008474576
Cthe_0515	-	transposase IS66	1535	13	0.008469055
Cthe_2879	-	cellulosome enzyme, dockerin type I	1535	13	0.008469055
Cthe_1586	-	binding-protein-dependent transport systems inner membrane component	1655	14	0.008459215
Cthe_1405	-	Spore coat assembly protein-like protein	2129	18	0.008454674
Cthe_1890	-	cellulosome enzyme, dockerin type I	2132	18	0.008442777
Cthe_2180	+	Cl- channel, voltage gated	1307	11	0.008416220
Cthe_2181	+	predicted ATPase of the PP-loop superfamily implicated in cell cycle control	713	6	0.008415147
Cthe_1754	-	periplasmic binding protein	956	8	0.008368201
Cthe_1898	-	hypothetical protein	479	4	0.008350731
Cthe_1386	-	hypothetical protein	599	5	0.008347245
Cthe_2423	-	hypothetical protein	2996	25	0.008344459
Cthe_0026	+		1319	11	0.008339651
Cthe_2792	+	Phenylacetate--CoA ligase(EC:6.2.1.30)	1322	11	0.008320726
Cthe_1628	-	phage major capsid protein, HK97 family	1202	10	0.008319468
Cthe_2789	+	NLPA lipoprotein	1082	9	0.008317930
Cthe_3081	-	transglutaminase-like protein	842	7	0.008313539
Cthe_0366	-	hypothetical protein	2777	23	0.008282319
Cthe_1518	+	hypothetical protein	242	2	0.008264463
Cthe_2047	-	intein	1700	14	0.008235294
Cthe_0173	+	dehydrogenase (flavoproteins)	1094	9	0.008226691
Cthe_0902	-	3-methyl-2-oxobutanoate hydroxymethyltransferase(EC:2.1.2.11)	851	7	0.008225617
Cthe_1135	-	SEFIR domain containing protein	974	8	0.008213552
Cthe_0691	+	hypothetical protein	488	4	0.008196721
Cthe_0850	-		855	7	0.008187135
Cthe_3178	-	protein of unknown function DUF342	1589	13	0.008181246
Cthe_1901	+	transposase, mutator type	1223	10	0.008176615
Cthe_3161	+	Pseudouridine synthase, Rsu	734	6	0.008174387
Cthe_2150	-	integral membrane protein-like protein	1346	11	0.008172363
Cthe_0661	+	Ricin B lectin	1715	14	0.008163265
Cthe_1474	-	hypothetical protein	245	2	0.008163265
Cthe_2582	+	dihydroneopterin aldolase	368	3	0.008152174
Cthe_1360	-	polysaccharide biosynthesis protein CpsF	491	4	0.008146640
Cthe_0442	+	Polypeptide-transport-associated, FtsQ-type	860	7	0.008139535
Cthe_RF0090	-	(miscRNA)	246	2	0.008130081
Cthe_0396	+	ABC transporter related protein	1724	14	0.008120650
Cthe_0300	-	Allergen V5/Tpx-1 related	617	5	0.008103728
Cthe_2589	-	hypothetical protein	1358	11	0.008100147
Cthe_1751	-	RNA methyltransferase, TrmA family	1361	11	0.008082292
Cthe_0480	+	hypothetical protein	497	4	0.008048290
Cthe_2757	+	sodium/hydrogen exchanger	1616	13	0.008044554
Cthe_2138	+	glycoside hydrolase, family 43	1742	14	0.008036739
Cthe_0593	+	copper amine oxidase-like protein	998	8	0.008016032
Cthe_2559	+	DegT/DnrJ/EryC1/StrS aminotransferase	1373	11	0.008011653
Cthe_1170	-	Dihydrodipicolinate reductase(EC:1.3.1.26)	875	7	0.008000000

المنارة للاستشارات
www.manaraa.com

Cthe_1442	-	ATP-dependent transcriptional regulator-like protein	875	7	0.008000000
Cthe_0726	+	peptidase M18, aminopeptidase I	1502	12	0.007989348
Cthe_0281	+	hypothetical protein	251	2	0.007968127
Cthe_0881	-	diguanylate cyclase	1130	9	0.007964602
Cthe_1333	-	hypothetical protein	1130	9	0.007964602
Cthe_2663	+	methyl-accepting chemotaxis sensory transducer	1256	10	0.007961783
Cthe_2128	-	extracellular solute-binding protein, family 1	1382	11	0.007959479
Cthe_0443	+	protein of unknown function DUF1290	377	3	0.007957560
Cthe_2313	-	protein of unknown function DUF218	755	6	0.007947020
Cthe_0051	+	hypothetical protein	1259	10	0.007942812
Cthe_0581	-	Fibronectin-binding A-like protein	1766	14	0.007927520
Cthe_0631	+	hypothetical protein	884	7	0.007918552
Cthe_1820	-	ABC transporter related protein	758	6	0.007915567
Cthe_3177	-	hypothetical protein	1265	10	0.007905138
Cthe_2561	-	CDP-glucose 4,6-dehydratase	1139	9	0.007901668
Cthe_1709	-	Phage-related protein-like protein	2279	18	0.007898201
Cthe_2798	+	putative endoribonuclease L-PSP	380	3	0.007894737
Cthe_0824	+	copper amine oxidase-like protein	1775	14	0.007887324
Cthe_2116	+	binding-protein-dependent transport systems inner membrane component	761	6	0.007884363
Cthe_1678	-	hypothetical protein	635	5	0.007874016
Cthe_0835	-	hypothetical protein	254	2	0.007874016
Cthe_1985	-	phage major capsid protein, HK97	1271	10	0.007867821
Cthe_2382	-	major facilitator superfamily MFS_1	1400	11	0.007857143
Cthe_3024	-	NADH dehydrogenase (quinone)(EC:1.6.99.5)	1910	15	0.007853403
Cthe_0623	+	transcriptional regulators, TraR/DksA family	638	5	0.007836991
Cthe_0297	+	hypothetical protein	383	3	0.007832898
Cthe_3106	+	6-pyruvoyl tetrahydropterin synthase and hypothetical protein	383	3	0.007832898
Cthe_1413	-	hypothetical protein	1916	15	0.007828810
Cthe_2062	+	GerA spore germination protein	1535	12	0.007817590
Cthe_2055	-	hypothetical protein	1919	15	0.007816571
Cthe_1968	-	cell divisionFtsK/SpoIIIE	3200	25	0.007812500
Cthe_0731	+	Shikimate kinase(EC:2.7.1.71)	512	4	0.007812500
Cthe_2046	-	hypothetical protein	512	4	0.007812500
Cthe_1283	-	endonuclease III	641	5	0.007800312
Cthe_2123	-	hypothetical protein	1412	11	0.007790368
Cthe_1481	+	membrane protein-like protein	2312	18	0.007785467
Cthe_2111	+	protein of unknown function DUF214	2057	16	0.007778318
Cthe_0046	+	hypothetical protein	2702	21	0.007772021
Cthe_1521	-	hypothetical protein	515	4	0.007766990
Cthe_1527	-	periplasmic sensor signal transduction histidine kinase	1031	8	0.007759457
Cthe_3185	+	tRNA-guanine transglycosylases, various specificities	1034	8	0.007736944
Cthe_1356	-	glycosyltransferase	905	7	0.007734807
Cthe_2465	+	hypothetical protein	905	7	0.007734807

المنارة للاستشارات
www.manaraa.com

Cthe_2690	+	UbiA prenyltransferase	905	7	0.007734807
Cthe_2377	+	Cobyrinic acid a,c-diamide synthase	776	6	0.007731959
Cthe_0925	-	protein of unknown function DUF218	647	5	0.007727975
Cthe_2776	+	hypothetical protein	647	5	0.007727975
Cthe_0898	-	metal dependent phosphohydrolase	1037	8	0.007714561
Cthe_1111	-	Bile acid:sodium symporter	1037	8	0.007714561
Cthe_1231	+	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	1298	10	0.007704160
Cthe_0929	-	Stage V sporulation protein S	260	2	0.007692308
Cthe_2793	+	aminotransferase, class I and II	1172	9	0.007679181
Cthe_1632	-		1567	12	0.007657945
Cthe_2768	+	metallophosphoesterase	2612	20	0.007656968
Cthe_0758	-	protein of unknown function UPF0102	392	3	0.007653061
Cthe_2052	-	hypothetical protein	392	3	0.007653061
Cthe_2032	-	hypothetical protein	6014	46	0.007648819
Cthe_0561	+	ApbE-like lipoprotein	1046	8	0.007648184
Cthe_0483	+	type III secretion system inner membrane R protein	785	6	0.007643312
Cthe_2967	+	major facilitator superfamily MFS_1	1178	9	0.007640068
Cthe_3202	+	CRISPR-associated protein, Csh2 family	917	7	0.007633588
Cthe_1330	-	signal peptidase I	524	4	0.007633588
Cthe_0392	+	inner-membrane translocator	1049	8	0.007626311
Cthe_3113	+	Nucleotidyl transferase	1049	8	0.007626311
Cthe_3199	+	protein of unknown function UPF0118	1049	8	0.007626311
Cthe_2788	+	hypothetical protein	656	5	0.007621951
Cthe_2304	+	hypothetical protein	920	7	0.007608696
Cthe_0405	+	glycoside hydrolase, family 5	1580	12	0.007594937
Cthe_1023	-	protein of unknown function DUF205	659	5	0.007587253
Cthe_1443	-	hypothetical protein	923	7	0.007583965
Cthe_3184	+	cation transporter	1451	11	0.007580979
Cthe_1119	-	hypothetical protein	794	6	0.007556675
Cthe_1878	-	hypothetical protein	794	6	0.007556675
Cthe_2204	-	cyanophycin synthetase	2648	20	0.007552870
Cthe_1810	-	peptidase M56, BlaR1	2252	17	0.007548845
Cthe_2197	-	Carbohydrate binding family 6	2786	21	0.007537688
Cthe_1616	-	phage minor structural protein	2522	19	0.007533703
Cthe_2401	+	protein of unknown function DUF1021	266	2	0.007518797
Cthe_0320	+	hypothetical protein	1466	11	0.007503411
Cthe_0401	+	methyl-accepting chemotaxis sensory transducer	2267	17	0.007498897
Cthe_0756	-	Membrane dipeptidase(EC:3.4.13.19)	935	7	0.007486631
Cthe_3102	+	cell envelope-related transcriptional attenuator	935	7	0.007486631
Cthe_1110	-	protein tyrosine phosphatase	401	3	0.007481297
Cthe_0055	+	hypothetical protein	536	4	0.007462687
Cthe_0537	+	two component transcriptional regulator, winged helix family	671	5	0.007451565
Cthe_3073	+	HAD-superfamily hydrolase, subfamily IA, variant 3	671	5	0.007451565
Cthe_1513	-	DNA adenine methylase(EC:2.1.1.72)	941	7	0.007438895

المنارة للاستشارات
www.manaraa.com

Cthe_0800	+	periplasmic sensor signal transduction histidine kinase	1211	9	0.007431874
Cthe_0045	+	copper amine oxidase-like protein	809	6	0.007416564
Cthe_2205	-	cyanophycinase	809	6	0.007416564
Cthe_3133	+	alpha/beta hydrolase fold	809	6	0.007416564
Cthe_0116	+	protein of unknown function DUF199	944	7	0.007415254
Cthe_0133	+	AMP-dependent synthetase and ligase	1484	11	0.007412399
Cthe_2270	-	ABC transporter related protein	1622	12	0.007398274
Cthe_2943	+	ABC-2 type transporter	1217	9	0.007395234
Cthe_0169	-	dehydrogenase (flavoproteins)	1082	8	0.007393715
Cthe_0432	-	hypothetical protein	677	5	0.007385524
Cthe_2679	-	Tetratricopeptide TPR_2	1085	8	0.007373272
Cthe_1482	-	Cof-like hydrolase	815	6	0.007361963
Cthe_2451	-	SNF2-related protein	3263	24	0.007355195
Cthe_0166	+	4Fe-4S ferredoxin, iron-sulfur binding	680	5	0.007352941
Cthe_1926	-	intein	1226	9	0.007340946
Cthe_0084	-	hypothetical protein	545	4	0.007339450
Cthe_0564	+	Trans-hexaprenyltranstransferase(EC:2.5.1.30)	956	7	0.007322176
Cthe_1054	+	ATPase associated with various cellular activities, AAA_3	956	7	0.007322176
Cthe_2223	-	LmbE-like protein	683	5	0.007320644
Cthe_0846	-	hypothetical protein	410	3	0.007317073
Cthe_2755	+	ADP-ribosylation/Crystallin J1	821	6	0.007308161
Cthe_1905	-	glycosyl transferase, family 28	1232	9	0.007305195
Cthe_2612	+	Fibronectin, type III	3836	28	0.007299270
Cthe_0050	+	hypothetical protein	959	7	0.007299270
Cthe_2319	-	restriction endonuclease (HaeIII)	959	7	0.007299270
Cthe_1217	+	ATP-dependent Clp protease, ATP-binding subunit clpA	2330	17	0.007296137
Cthe_0670	+	spore germination protein	1097	8	0.007292616
Cthe_1406	-	hypothetical protein	686	5	0.007288630
Cthe_1418	-	hypothetical protein	686	5	0.007288630
Cthe_1030	-	TPR repeat domain containing protein	824	6	0.007281553
Cthe_0020	+	biotin synthase(EC:2.8.1.6)	962	7	0.007276507
Cthe_1591	-	hypothetical protein	413	3	0.007263923
Cthe_0145	+	metal dependent phosphohydrolase	551	4	0.007259528
Cthe_1693	-	hypothetical protein	551	4	0.007259528
Cthe_1969	-	hypothetical protein	1379	10	0.007251632
Cthe_1011	+	Peptidoglycan glycosyltransferase(EC:2.4.1.129)	1655	12	0.007250755
Cthe_0954	-	Uracil phosphoribosyltransferase(EC:2.4.2.9)	554	4	0.007220217
Cthe_1400	-	glycosyl hydrolase 53	1247	9	0.007217322
Cthe_0134	+	hypothetical protein	971	7	0.007209063
Cthe_2260	-	Prephenate dehydratase(EC:4.2.1.51)	836	6	0.007177033
Cthe_2868	+	transposase IS116/IS110/IS902	1256	9	0.007165605
Cthe_1721	-	phage portal protein, HK97 family	1259	9	0.007148531
Cthe_0907	-	hypothetical protein	140	1	0.007142857
Cthe_0734	-	peptidase M23B	1121	8	0.007136485

المنارة للاستشارات

www.manaraa.com

Cthe_0888	-	type IV pilus assembly PilZ	701	5	0.007132668
Cthe_2415	+	cell wall hydrolase, SleB	701	5	0.007132668
Cthe_1665	-	Hsp33 protein	842	6	0.007125891
Cthe_1991	+	transposase, mutator type	986	7	0.007099391
Cthe_3136	-	peptidase S8 and S53, subtilisin, kexin, sedolisin	1127	8	0.007098492
Cthe_1448	-	MATE efflux family protein	1412	10	0.007082153
Cthe_2511	+	MscS Mechanosensitive ion channel	848	6	0.007075472
Cthe_3038	+	hypothetical protein	707	5	0.007072136
Cthe_1590	+	peptidase M56, BlaR1	2264	16	0.007067138
Cthe_3172	+	protein of unknown function DUF214	1415	10	0.007067138
Cthe_3195	+	hypothetical protein	425	3	0.007058824
Cthe_1690	-	Radical SAM	992	7	0.007056452
Cthe_1334	+	FHA domain containing protein	1562	11	0.007042254
Cthe_0030	+	phosphatidate cytidylyltransferase	710	5	0.007042254
Cthe_1755	-	hypothetical protein	284	2	0.007042254
Cthe_2185	-	ribosomal protein S18	284	2	0.007042254
Cthe_2891	+	chaperonin Cpn10	284	2	0.007042254
Cthe_1144	-	restriction modification system DNA specificity domain	1421	10	0.007037298
Cthe_1109	-	permease	854	6	0.007025761
Cthe_1551	+		712	5	0.007022472
Cthe_2402	+	Peptidoglycan-binding LysM	1568	11	0.007015306
Cthe_1076	-	protein of unknown function DUF881	713	5	0.007012623
Cthe_0804	+	transcriptional regulator, MerR family	572	4	0.006993007
Cthe_1404	+	GTP-binding protein, HSR1-related	2150	15	0.006976744
Cthe_1415	-	ABC transporter related protein	860	6	0.006976744
Cthe_2863	+	intein	1004	7	0.006972112
Cthe_2623	+	Exopolysaccharide biosynthesis protein	1148	8	0.006968641
Cthe_2847	+	phage / plasmid primase, P4 family	2153	15	0.006967023
Cthe_2289	-	hypothetical protein	719	5	0.006954103
Cthe_3068	+	periplasmic sensor signal transduction histidine kinase	1439	10	0.006949270
Cthe_0795	-	alpha amylase, catalytic region	1727	12	0.006948466
Cthe_1639	-	DNA methylase N-4/N-6	1298	9	0.006933744
Cthe_0502	+	hypothetical protein	866	6	0.006928406
Cthe_1907	-	amino acid adenylation domain	1877	13	0.006925946
Cthe_1706	-	hypothetical protein	1157	8	0.006914434
Cthe_3040	+	hypothetical protein	434	3	0.006912442
Cthe_2413	+	NLP/P60	1013	7	0.006910168
Cthe_2300	-	CRISPR-associated protein Cas5	725	5	0.006896552
Cthe_2541	+	nitrite and sulphite reductase 4Fe-4S region	872	6	0.006880734
Cthe_2814	+	periplasmic sensor signal transduction histidine kinase	2039	14	0.006866111
Cthe_2337	-	hypothetical protein	1313	9	0.006854532
Cthe_1515	-	GCN5-related N-acetyltransferase	584	4	0.006849315
Cthe_2113	+	periplasmic sensor signal transduction histidine kinase	1025	7	0.006829268
Cthe_0885	-	dephospho-CoA kinase(EC:2.7.1.24)	587	4	0.006814310

المنارة للاستشارات

www.manaraa.com

Cthe_2939	+	tRNA pseudouridine synthase A(EC:4.2.1.70)	734	5	0.006811989
Cthe_1192	-	hypothetical protein	1031	7	0.006789525
Cthe_2049	-	protein of unknown function DUF324	884	6	0.006787330
Cthe_0740	-	transcriptional regulator, GntR family	737	5	0.006784261
Cthe_2192	+	hypothetical protein	590	4	0.006779661
Cthe_2935	-	methyltransferase small	590	4	0.006779661
Cthe_0293	+	hypothetical protein	443	3	0.006772009
Cthe_0463	+	flagellar basal-body rod protein FlgC	443	3	0.006772009
Cthe_1311	-	signal transduction histidine kinase regulating citrate/malate metabolism	1331	9	0.006761833
Cthe_0120	+	RNA polymerase, sigma 28 subunit	740	5	0.006756757
Cthe_1088	-	hypothetical protein	740	5	0.006756757
Cthe_0132	+	3-oxoacyl-(acyl-carrier-protein) synthase(EC:2.3.1.41)	1037	7	0.006750241
Cthe_0860	-	Rubrerythrin	593	4	0.006745363
Cthe_1230	-	competence protein ComEA helix-hairpin-helix repeat protein	890	6	0.006741573
Cthe_0755	-	aminotransferase, class I and II(EC:2.6.1.1)	1187	8	0.006739680
Cthe_2141	-	metal-dependent phosphohydrolase, HD subdomain	1187	8	0.006739680
Cthe_RF0075	+	(miscRNA)	297	2	0.006734007
Cthe_1524	-	hypothetical protein	1040	7	0.006730769
Cthe_0521	-	helicase-like protein	3269	22	0.006729887
Cthe_0830	-	acid phosphatase/vanadium-dependent haloperoxidase related	446	3	0.006726457
Cthe_2024	-	transposase IS116/IS110/IS902	1190	8	0.006722689
Cthe_2097	+	hypothetical protein	1043	7	0.006711409
Cthe_2717	+	ribosomal protein L33	149	1	0.006711409
Cthe_0439	-	CobB/CobQ-like glutamine amidotransferase	746	5	0.006702413
Cthe_1548	-	Enoyl-CoA hydratase/isomerase	746	5	0.006702413
Cthe_2462	+	BRO-like protein	746	5	0.006702413
Cthe_1220	-	hypothetical protein	896	6	0.006696429
Cthe_2570	+	hypothetical protein	2390	16	0.006694561
Cthe_0004	+	YD repeat containing protein	5828	39	0.006691833
Cthe_1696	-	hypothetical protein	599	4	0.006677796
Cthe_2701	-	glycosyl transferase, WecB/TagA/CpsF family(EC:2.4.1.187)	749	5	0.006675567
Cthe_0242	-	hypothetical protein	899	6	0.006674082
Cthe_1635	-	hypothetical protein	899	6	0.006674082
Cthe_0779	-	copper amine oxidase-like protein	1352	9	0.006656805
Cthe_1658	+	potassium uptake protein, TrkH family	1352	9	0.006656805
Cthe_1749	-	DNA-cytosine methyltransferase(EC:2.1.1.37)	1052	7	0.006653992
Cthe_1437	-	hypothetical protein	902	6	0.006651885
Cthe_3067	+	membrane spanning protein	752	5	0.006648936
Cthe_0247	+	DNA mismatch repair protein MutS-like protein	1805	12	0.006648199
Cthe_2983	+	methyl-accepting chemotaxis sensory transducer	1655	11	0.006646526
Cthe_2732	+	hypothetical protein	602	4	0.006644518
Cthe_2314	-	hypothetical protein	755	5	0.006622517
Cthe_2769	-	protein of unknown function DUF542, ScdA-like	755	5	0.006622517

للاستشارات المنارة
www.manaraa.com

		protein			
Cthe_2942	+	ABC transporter related protein	755	5	0.006622517
Cthe_1684	-	transposase IS3/IS911	605	4	0.006611570
Cthe_0909	-	lipoprotein signal peptidase	455	3	0.006593407
Cthe_1645	-	conserved hypothetical phage-associated protein	455	3	0.006593407
Cthe_0679	+	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	1214	8	0.006589786
Cthe_0908	-	pseudouridine synthase, RluA family(EC:4.2.1.70)	911	6	0.006586169
Cthe_1908	-	copper amine oxidase-like protein	2888	19	0.006578947
Cthe_2648	-	sugar transferase	608	4	0.006578947
Cthe_2343	-	GCN5-related N-acetyltransferase	458	3	0.006550218
Cthe_0256	+	histidine kinase	2138	14	0.006548176
Cthe_1475	-	hypothetical protein	917	6	0.006543075
Cthe_1536	-	ABC transporter related protein	917	6	0.006543075
Cthe_3062	+	signal transduction histidine kinase regulating citrate/malate metabolism	1382	9	0.006512301
Cthe_0708	+	FAD dependent oxidoreductase	1229	8	0.006509357
Cthe_1243	-	GCN5-related N-acetyltransferase	461	3	0.006507592
Cthe_1370	-	hypothetical protein	461	3	0.006507592
Cthe_1986	-	phage / plasmid primase, P4 family	2153	14	0.006502555
Cthe_1079	-	Nucleotidyl transferase	2462	16	0.006498781
Cthe_3050	+	Fibronectin, type III	2003	13	0.006490265
Cthe_0509	-	sodium ion-translocating decarboxylase, beta subunit(EC:4.1.1.70)	1079	7	0.006487488
Cthe_1168	-	Extradiol ring-cleavage dioxygenase, class III enzyme, subunit B	1388	9	0.006484150
Cthe_1595	-	hypothetical protein	617	4	0.006482982
Cthe_2647	-	protein of unknown function DUF201	926	6	0.006479482
Cthe_2309	-	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	773	5	0.006468305
Cthe_2458	-	hypothetical protein	464	3	0.006465517
Cthe_1367	-	PHP-like protein	776	5	0.006443299
Cthe_0548	+	ABC transporter related protein	779	5	0.006418485
Cthe_1124	-	phenazine biosynthesis protein PhzF family	779	5	0.006418485
Cthe_2620	+	protein of unknown function DUF1078-like protein	779	5	0.006418485
Cthe_2997	+	hypothetical protein	779	5	0.006418485
Cthe_0197	+	glutamine amidotransferase, class-II	1091	7	0.006416132
Cthe_1826	+	response regulator receiver sensor signal transduction histidine kinase	1091	7	0.006416132
Cthe_2828	+	hypothetical protein	470	3	0.006382979
Cthe_2151	-	hypothetical protein	2195	14	0.006378132
Cthe_2233	-	protein of unknown function DUF115	1412	9	0.006373938
Cthe_1450	-	hypothetical protein	785	5	0.006369427
Cthe_1456	-	ABC transporter related protein	785	5	0.006369427
Cthe_0018	+	hypothetical protein	314	2	0.006369427
Cthe_0464	+	flagellar hook-basal body complex subunit FliE	314	2	0.006369427
Cthe_0062	-	sulfatase	1889	12	0.006352567

المنارة للاستشارات

www.manaraa.com

Cthe_0219	-	5-nitroimidazole antibiotic resistance protein	473	3	0.006342495
Cthe_0956	+	hypothetical protein	473	3	0.006342495
Cthe_1502	-	transcriptional regulator, MarR family	473	3	0.006342495
Cthe_1606	-		1579	10	0.006333122
Cthe_0085	-	Negative regulator of genetic competence	632	4	0.006329114
Cthe_1277	-	putative methyltransferase	632	4	0.006329114
Cthe_2164	-	Rubredoxin-type Fe(Cys)4 protein	158	1	0.006329114
Cthe_2170	+	hypothetical protein	1109	7	0.006311993
Cthe_1073	-	protein of unknown function DUF1429	317	2	0.006309148
Cthe_1870	+	GCN5-related N-acetyltransferase	476	3	0.006302521
Cthe_2500	+	hypothetical protein	635	4	0.006299213
Cthe_1499	+	ammonium transporter	1754	11	0.006271380
Cthe_2610	+	hypothetical protein	1754	11	0.006271380
Cthe_1726	-	hypothetical protein	479	3	0.006263048
Cthe_2439	+	GCN5-related N-acetyltransferase	479	3	0.006263048
Cthe_2596	+	protein tyrosine phosphatase	479	3	0.006263048
Cthe_1647	-	hypothetical protein	2555	16	0.006262231
Cthe_2785	+	methyltransferase MtaA/CmuA family	1118	7	0.006261181
Cthe_2426	+	OmpA/MotB	800	5	0.006250000
Cthe_0379	+	hypothetical protein	962	6	0.006237006
Cthe_0984	-	prolipoprotein diacylglyceryl transferase	962	6	0.006237006
Cthe_0073	+	hypothetical protein	965	6	0.006217617
Cthe_2299	-	CRISPR-associated helicase Cas3	2252	14	0.006216696
Cthe_1340	-	hypothetical protein	1127	7	0.006211180
Cthe_2059	-	RNA polymerase, sigma-24 subunit, ECF subfamily	644	4	0.006211180
Cthe_0218	-	metallophosphoesterase	806	5	0.006203474
Cthe_0735	-	cellulosome anchoring protein, cohesin region	806	5	0.006203474
Cthe_2045	-	intein	806	5	0.006203474
Cthe_1085	+	glycosyl transferase, group 1	1292	8	0.006191950
Cthe_2771	-	beta-lactamase-like protein	1136	7	0.006161972
Cthe_1488	-	methyl-accepting chemotaxis sensory transducer	2276	14	0.006151142
Cthe_0781	-	GCN5-related N-acetyltransferase	488	3	0.006147541
Cthe_2261	+	hypothetical protein	326	2	0.006134969
Cthe_3017	-	hydrogenase accessory protein HypB	653	4	0.006125574
Cthe_1421	+	signal peptide peptidase SppA, 36K type	980	6	0.006122449
Cthe_1850	-	hypothetical protein	491	3	0.006109980
Cthe_1695	-	Radical SAM	1475	9	0.006101695
Cthe_2866	+	phage major capsid protein, HK97	1313	8	0.006092917
Cthe_1169	-	Beta-lactamase class A-like protein	1151	7	0.006081668
Cthe_2414	-	Monogalactosyldiacylglycerol synthase	1151	7	0.006081668
Cthe_0336	+	DRTGG domain containing protein	329	2	0.006079027
Cthe_0034	-	cyclic nucleotide-binding protein	659	4	0.006069803
Cthe_2297	-	CRISPR-associated protein Cas1	992	6	0.006048387
Cthe_1065	-	type IV pilus assembly PilZ	662	4	0.006042296
Cthe_2700	+	hypothetical protein	497	3	0.006036217

المنارة للاستشارات

www.manaraa.com

Cthe_0635	-	protein serine/threonine phosphatases	830	5	0.006024096
Cthe_1227	-	Thymidylate synthase(EC:2.1.1.45)	830	5	0.006024096
Cthe_1583	+	deoxyribose-phosphate aldolase(EC:4.1.2.4)	668	4	0.005988024
Cthe_0556	-	asparagine synthase (glutamine-hydrolyzing)(EC:6.3.5.4)	1841	11	0.005975014
Cthe_0363	+	aminotransferase, class I and II(EC:2.6.1.1)	1172	7	0.005972696
Cthe_1244	+	glycosyl transferase, family 2	1172	7	0.005972696
Cthe_1567	+	Dinitrogenase iron-molybdenum cofactor biosynthesis	335	2	0.005970149
Cthe_2302	-	hypothetical protein	1676	10	0.005966587
Cthe_2787	+	ferredoxin	1676	10	0.005966587
Cthe_0484	+	flagellar biosynthetic protein FlhB	1178	7	0.005942275
Cthe_1578	+	Radical SAM	842	5	0.005938242
Cthe_2948	+	Allergen V5/Tpx-1 related	842	5	0.005938242
Cthe_1471	+	glycoside hydrolase, family 5	1685	10	0.005934718
Cthe_R0066	+	(rRNA)	1520	9	0.005921053
Cthe_0852	-	peptidase A24A-like protein	845	5	0.005917160
Cthe_1680	-	DNA polymerase, beta-like region	845	5	0.005917160
Cthe_1837	-	hypothetical protein	509	3	0.005893910
Cthe_1238	-	iojap-like protein	341	2	0.005865103
Cthe_2259	-	hypothetical protein	341	2	0.005865103
Cthe_3175	-	protein of unknown function DUF421	683	4	0.005856515
Cthe_0960	-	SpoIID/LytB domain containing protein	1712	10	0.005841121
Cthe_2070	-	hypothetical protein	857	5	0.005834306
Cthe_1288	-	two component transcriptional regulator, winged helix family	686	4	0.005830904
Cthe_0025	+	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	1373	8	0.005826657
Cthe_1133	-	hypothetical protein	515	3	0.005825243
Cthe_1516	-	ribosome small subunit-dependent GTPase A	1031	6	0.005819593
Cthe_0819	-	ABC transporter related protein	863	5	0.005793743
Cthe_1954	-	hypothetical protein	863	5	0.005793743
Cthe_2085	-	deoxyuridine 5'-triphosphate nucleotidohydrolase Dut(EC:3.6.1.23)	518	3	0.005791506
Cthe_1895	-	transcriptional regulator, PadR-like family	347	2	0.005763689
Cthe_1100	-	prepilin-type cleavage/methylation	521	3	0.005758157
Cthe_1609	-	Recombinase	1568	9	0.005739796
Cthe_2317	-	response regulator receiver protein	1394	8	0.005738881
Cthe_1920	-	stage II sporulation protein R	698	4	0.005730659
Cthe_0797	-	glycoside hydrolase, family 5	2444	14	0.005728314
Cthe_1285	-	metal dependent phosphohydrolase	1397	8	0.005726557
Cthe_2763	+	hypothetical protein	1052	6	0.005703422
Cthe_0684	+	Polynucleotide adenylyltransferase region	2633	15	0.005696924
Cthe_1166	-	protein of unknown function DUF147	878	5	0.005694761
Cthe_2759	+	transcriptional regulator, AraC family	878	5	0.005694761
Cthe_1143	-	hypothetical protein	1055	6	0.005687204
Cthe_0397	+	ABC transporter related protein	1760	10	0.005681818
Cthe_1708	-	phage putative tail component	704	4	0.005681818
Cthe_1819	-	ABC transporter related protein	704	4	0.005681818

المنارة للاستشارات
www.manaraa.com

Cthe_2306	-	MATE efflux family protein	1409	8	0.005677786
Cthe_2161	-	hypothetical protein	881	5	0.005675369
Cthe_0582	-	PAS/PAC sensor signal transduction histidine kinase	1763	10	0.005672150
Cthe_1016	-	cell wall hydrolase/autolysin	1058	6	0.005671078
Cthe_0333	+	hypothetical protein	353	2	0.005665722
Cthe_2978	+	transcriptional regulator, Crp/Fnr family	707	4	0.005657709
Cthe_2533	+	sulfate ABC transporter, inner membrane subunit CysW	884	5	0.005656109
Cthe_0667	+	GerA spore germination protein	1769	10	0.005652911
Cthe_0358	-	hypothetical protein	533	3	0.005628518
Cthe_1925	-	hypothetical protein	533	3	0.005628518
Cthe_2385	+	hypothetical protein	533	3	0.005628518
Cthe_0676	+	tyrosine recombinase XerD	890	5	0.005617978
Cthe_0882	-	protein of unknown function DUF6, transmembrane	890	5	0.005617978
Cthe_2531	+	sulfate ABC transporter, periplasmic sulfate-binding protein	1070	6	0.005607477
Cthe_3215	+	CRISPR-associated protein, TM1812 family	1250	7	0.005600000
Cthe_1477	-	deoxycytidine triphosphate deaminase	536	3	0.005597015
Cthe_2271	-	cellulosome enzyme, dockerin type I	536	3	0.005597015
Cthe_3179	+	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	1256	7	0.005573248
Cthe_0002	+	hypothetical protein	719	4	0.005563282
Cthe_2395	-	transcriptional regulator, XRE family	719	4	0.005563282
Cthe_1823	-	Extracellular ligand-binding receptor	1259	7	0.005559968
Cthe_1209	+	hypothetical protein	1799	10	0.005558644
Cthe_2466	+	hypothetical protein	1082	6	0.005545287
Cthe_1252	-	Auxin Efflux Carrier	1085	6	0.005529954
Cthe_2281	-	response regulator receiver modulated CheB methylesterase(EC:3.1.1.61)	1085	6	0.005529954
Cthe_0036	+	hybrid cluster protein	1628	9	0.005528256
Cthe_0479	+	response regulator receiver protein	362	2	0.005524862
Cthe_2750	+	IstB-like ATP-binding protein	725	4	0.005517241
Cthe_0859	+	hypothetical protein	545	3	0.005504587
Cthe_1047	-	hypothetical protein	545	3	0.005504587
Cthe_1132	-	hypothetical protein	728	4	0.005494505
Cthe_2310	+	hypothetical protein	728	4	0.005494505
Cthe_2125	-	binding-protein-dependent transport systems inner membrane component	911	5	0.005488474
Cthe_2137	+	cellulosome enzyme, dockerin type I	2372	13	0.005480607
Cthe_0061	-	BioY protein	548	3	0.005474453
Cthe_2226	-	FkbH like protein	1646	9	0.005467801
Cthe_0021	+	dethiobiotin synthase(EC:6.3.3.3)	734	4	0.005449591
Cthe_1959	-	Methyltransferase type 11	734	4	0.005449591
Cthe_0503	+	2'-5' RNA ligase	551	3	0.005444646
Cthe_0552	-	transcriptional regulator, XRE family with cupin sensor	551	3	0.005444646
Cthe_1640	-	hypothetical protein	551	3	0.005444646
Cthe_2655	+	transcriptional regulator, AbrB family	551	3	0.005444646

المنارة للاستشارات

www.manaraa.com

Cthe_3076	+	Radical SAM	1292	7	0.005417957
Cthe_2731	+	RNA polymerase, sigma-24 subunit, ECF subfamily	554	3	0.005415162
Cthe_2950	-	Pectate lyase/Amb allergen	1664	9	0.005408654
Cthe_0727	-	cell envelope-related transcriptional attenuator	926	5	0.005399568
Cthe_2091	-	hypothetical protein	1298	7	0.005392912
Cthe_0875	-	anthranilate synthase component I(EC:4.1.3.27)	1484	8	0.005390836
Cthe_0990	-	ribosome-binding factor A	371	2	0.005390836
Cthe_0291	-	protein of unknown function DUF255	2042	11	0.005386876
Cthe_1034	-	hypothetical protein	557	3	0.005385996
Cthe_0075	+	hypothetical protein	743	4	0.005383580
Cthe_0353	+	hypothetical protein	1115	6	0.005381166
Cthe_2287	+	transcriptional regulator, AraC family	932	5	0.005364807
Cthe_3152	+	cobalamin 5'-phosphate synthase	746	4	0.005361930
Cthe_1525	-	protein of unknown function DUF214	1868	10	0.005353319
Cthe_2653	+	hypothetical protein	749	4	0.005340454
Cthe_3044	+	protein of unknown function DUF975	749	4	0.005340454
Cthe_1739	-	SNF2-related protein	1127	6	0.005323869
Cthe_2336	-	glycosyl transferase, group 1	1127	6	0.005323869
Cthe_2072	-	hypothetical protein	752	4	0.005319149
Cthe_1279	+	ribosomal protein L28	188	1	0.005319149
Cthe_2490	+	hypothetical protein	944	5	0.005296610
Cthe_2548	+	Alpha-N-arabinofuranosidase(EC:3.2.1.55)	1511	8	0.005294507
Cthe_2155	-	S-layer-like domain containing protein	1895	10	0.005277045
Cthe_2782	+	N-acetylmuramyl-L-alanine amidase, negative regulator of AmpC, AmpD	758	4	0.005277045
Cthe_3169	-	short-chain dehydrogenase/reductase SDR(EC:1.3.1.9)	758	4	0.005277045
Cthe_3108	+	phage SPO1 DNA polymerase-related protein	569	3	0.005272408
Cthe_1266	-	methyl-accepting chemotaxis sensory transducer	2468	13	0.005267423
Cthe_0900	-	aspartate 1-decarboxylase(EC:4.1.1.11)	380	2	0.005263158
Cthe_2077	-	CoA-binding protein	380	2	0.005263158
Cthe_0531	+	hypothetical protein	2288	12	0.005244755
Cthe_1885	+	phage integrase-like SAM-like	572	3	0.005244755
Cthe_3011	+	hypothetical protein	383	2	0.005221932
Cthe_1654	-	hypothetical protein	1919	10	0.005211047
Cthe_3041	+	UbiA prenyltransferase	962	5	0.005197505
Cthe_2804	+	ABC transporter related protein	770	4	0.005194805
Cthe_1967	-	DNA segregation ATPase FtsK/SpoIIIE and related proteins	1736	9	0.005184332
Cthe_1097	-	peptidase S14, ClpP	773	4	0.005174644
Cthe_1829	-	Chromosome segregation ATPases-like protein	3296	17	0.005157767
Cthe_1653	-	SNF2-related protein	1358	7	0.005154639
Cthe_0023	+	pimeloyl-CoA synthesis (biotin biosynthesis)	776	4	0.005154639
Cthe_1289	-	hypothetical protein	194	1	0.005154639
Cthe_1827	-	copper amine oxidase-like protein	971	5	0.005149331
Cthe_0259	+	Rhomboid-like protein	584	3	0.005136986

المنارة للاستشارات

www.manaraa.com

Cthe_0810	-	CheA signal transduction histidine kinases	2147	11	0.005123428
Cthe_0620	+	iron (metal) dependent repressor, DtxR family	392	2	0.005102041
Cthe_1495	-	pyridoxamine 5'-phosphate oxidase-related, FMN-binding	392	2	0.005102041
Cthe_1552	-	MAEBL, putative	392	2	0.005102041
Cthe_0826	-	hypothetical protein	785	4	0.005095541
Cthe_2980	+	hypothetical protein	590	3	0.005084746
Cthe_0032	-	glycoside hydrolase, family 26	1772	9	0.005079007
Cthe_1746	-	hypothetical protein	788	4	0.005076142
Cthe_1772	-	protein of unknown function DUF218	788	4	0.005076142
Cthe_1812	-	Urease accessory protein UreD	788	4	0.005076142
Cthe_1262	-	hypothetical protein	197	1	0.005076142
Cthe_2219	-	flagellar protein FlaG protein	395	2	0.005063291
Cthe_0354	+	putative CoA-substrate-specific enzyme activase	989	5	0.005055612
Cthe_2959	+		991	5	0.005045409
Cthe_3127	+	peptidase M56, BlaR1	1388	7	0.005043228
Cthe_3112	+	glycosidase, PH1107-related	992	5	0.005040323
Cthe_2127	-	ATPase associated with various cellular activities, AAA_3	794	4	0.005037783
Cthe_2509	+	DNA topoisomerase(EC:5.99.1.2)	995	5	0.005025126
Cthe_0724	-	histidinol phosphate phosphatase HisJ family(EC:3.1.3.15)	797	4	0.005018821
Cthe_3173	+	sodium/calcium exchanger membrane region	998	5	0.005010020
Cthe_0189	+	non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family	599	3	0.005008347
Cthe_0597	+	thiamine biosynthesis protein ThiS	200	1	0.005000000
Cthe_1026	-	ribosomal protein L32	200	1	0.005000000
Cthe_3211	+	protein of unknown function DUF324	1601	8	0.004996877
Cthe_3204	+	CRISPR-associated helicase Cas3	2402	12	0.004995837
Cthe_2480	+	hypothetical protein	401	2	0.004987531
Cthe_0343	+	flavin reductase-like, FMN-binding	602	3	0.004983389
Cthe_2452	-	oligoendopeptidase F	1808	9	0.004977876
Cthe_0622	-	methylthioadenosine phosphorylase	806	4	0.004962779
Cthe_2409	+	hypothetical protein	806	4	0.004962779
Cthe_0028	-	uncharacterized protein, YcgL-like protein	1010	5	0.004950495
Cthe_2575	+	hypothetical protein	1010	5	0.004950495
Cthe_2252	-	thioesterase family protein	404	2	0.004950495
Cthe_0972	-	hypothetical protein	608	3	0.004934211
Cthe_1813	-	urease accessory protein UreG	608	3	0.004934211
Cthe_2228	-	MaoC-like dehydratase	611	3	0.004909984
Cthe_2507	+	peptidase U57, YabG	815	4	0.004907975
Cthe_0841	-	mutants block sporulation after engulfment	1229	6	0.004882018
Cthe_0508	+	hypothetical protein	410	2	0.004878049
Cthe_0742	-	hypothetical protein	410	2	0.004878049
Cthe_3042	+	hypothetical protein	821	4	0.004872107
Cthe_2213	-	hypothetical protein	617	3	0.004862237
Cthe_2683	+	hypothetical protein	617	3	0.004862237
Cthe_0867	-	4Fe-4S ferredoxin, iron-sulfur binding	206	1	0.004854369

www.manaraa.com

Cthe_0313	+	hypothetical protein	1031	5	0.004849661
Cthe_1196	-	hypothetical protein	1238	6	0.004846527
Cthe_1049	-	regulatory protein RecX	620	3	0.004838710
Cthe_0470	+	hypothetical protein	830	4	0.004819277
Cthe_2424	-	copper amine oxidase-like protein	830	4	0.004819277
Cthe_1301	+	protein of unknown function DUF795	1247	6	0.004811548
Cthe_3141	+	lipolytic enzyme, G-D-S-L	2495	12	0.004809619
Cthe_2820	+	MCP methyltransferase, CheR-type(EC:2.1.1.80)	833	4	0.004801921
Cthe_1353	-	polysaccharide biosynthesis protein	1460	7	0.004794521
Cthe_2554	-	glycosyl transferase, group 1	1253	6	0.004788508
Cthe_0399	-	protein of unknown function DUF477	836	4	0.004784689
Cthe_1000	-	phosphatidate cytidylyltransferase	836	4	0.004784689
Cthe_2643	-	Nucleotidyl transferase	1046	5	0.004780115
Cthe_0565	+	hypothetical protein	419	2	0.004773270
Cthe_2540	+	Mov34/MPN/PAD-1	419	2	0.004773270
Cthe_0853	-	type II secretion system protein E	1685	8	0.004747774
Cthe_0817	+	diguanylate cyclase	5480	26	0.004744526
Cthe_1689	-	hypothetical protein	845	4	0.004733728
Cthe_3046	+	cell cycle protein	1268	6	0.004731861
Cthe_2492	+	Baseplate J-like protein	1058	5	0.004725898
Cthe_1444	-	GCN5-related N-acetyltransferase	425	2	0.004705882
Cthe_1009	+	O-methyltransferase, family 3	638	3	0.004702194
Cthe_1522	-	NADPH-dependent FMN reductase	638	3	0.004702194
Cthe_3037	+	nitroreductase	638	3	0.004702194
Cthe_2786	+	methyltransferase MtaA/CmuA family	1067	5	0.004686036
Cthe_2126	-	binding-protein-dependent transport systems inner membrane component	854	4	0.004683841
Cthe_0773	+	SOS-response transcriptional repressor, LexA(EC:3.4.21.88)	641	3	0.004680187
Cthe_1137	-	hypothetical protein	641	3	0.004680187
Cthe_2689	-	transcriptional regulator, GntR family	1496	7	0.004679144
Cthe_0849	-	3-dehydroquinate dehydratase, type II(EC:4.2.1.10)	428	2	0.004672897
Cthe_2352	-	putative signal-transduction protein with CBS domains	428	2	0.004672897
Cthe_2510	+	protein of unknown function DUF368	857	4	0.004667445
Cthe_0644	+	Vitamin B12 dependent methionine synthase, activation region	644	3	0.004658385
Cthe_2642	-	transferase hexapeptide repeat	644	3	0.004658385
Cthe_0876	+	glycosyl transferase, family 39	1292	6	0.004643963
Cthe_0739	+	hypothetical protein	1724	8	0.004640371
Cthe_0838	-	hypothetical protein	647	3	0.004636785
Cthe_1601	-	phosphate uptake regulator, PhoU	650	3	0.004615385
Cthe_2841	+	hypothetical protein	434	2	0.004608295
Cthe_2578	-	Ppx/GppA phosphatase	1520	7	0.004605263
Cthe_R0054	+	(rRNA)	1520	7	0.004605263
Cthe_0666	+	HflC protein	869	4	0.004602992
Cthe_1973	-	Tetratricopeptide TPR_2	1523	7	0.004596192

المنارة للاستشارات
www.manaraa.com

Cthe_1659	+	TrkA-N	653	3	0.004594181
Cthe_0407	-	Radical SAM	1307	6	0.004590666
Cthe_0669	+	spore germination B3 GerAC like	1310	6	0.004580153
Cthe_1561	+	hypothetical protein	437	2	0.004576659
Cthe_3022	-	NADH ubiquinone oxidoreductase, 20 kDa subunit	437	2	0.004576659
Cthe_0733	-	type IV pilus assembly PilZ	656	3	0.004573171
Cthe_2036	-	amidohydrolase 2	656	3	0.004573171
Cthe_0097	+	peptidase M23B	875	4	0.004571429
Cthe_0099	-	5,10-methylenetetrahydrofolate reductase(EC:1.5.1.20)	875	4	0.004571429
Cthe_1603	-	phosphate ABC transporter, inner membrane subunit PstA	875	4	0.004571429
Cthe_2848	+	phage major capsid protein, HK97	1313	6	0.004569688
Cthe_1514	+	protein of unknown function DUF125, transmembrane	878	4	0.004555809
Cthe_2527	+	porphobilinogen deaminase(EC:2.5.1.61)	878	4	0.004555809
Cthe_3168	+	hypothetical protein	878	4	0.004555809
Cthe_2625	+	beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ	440	2	0.004545455
Cthe_2536	+	phosphoadenosine phosphosulfate reductase	881	4	0.004540295
Cthe_1396	-	phospholipase D/Transphosphatidylase	1544	7	0.004533679
Cthe_2248	-	phosphoribosyltransferase	662	3	0.004531722
Cthe_2603	+	ATP synthase F0, C subunit	221	1	0.004524887
Cthe_0891	+	response regulator receiver protein	443	2	0.004514673
Cthe_1894	-	hypothetical protein	443	2	0.004514673
Cthe_0592	+	Methyltransferase type 11	665	3	0.004511278
Cthe_3048	+	DNA repair protein RadC	665	3	0.004511278
Cthe_2688	+	polysaccharide biosynthesis protein	1553	7	0.004507405
Cthe_2455	-	phage integrase	1112	5	0.004496403
Cthe_2175	-	hypothetical protein	668	3	0.004491018
Cthe_0317	-	hypothetical protein	671	3	0.004470939
Cthe_1549	-	Methyltransferase type 12	671	3	0.004470939
Cthe_3008	-	Superoxide dismutase(EC:1.15.1.1)	671	3	0.004470939
Cthe_2473	+	Terminase small subunit	449	2	0.004454343
Cthe_1598	+	nitroreductase	674	3	0.004451039
Cthe_2568	-	hypothetical protein	674	3	0.004451039
Cthe_0191	+	proteinase inhibitor I4, serpin	1799	8	0.004446915
Cthe_1934	-	Abortive infection protein	677	3	0.004431315
Cthe_1323	-	GrpE protein	680	3	0.004411765
Cthe_2290	-	ABC transporter related protein	680	3	0.004411765
Cthe_0377	+	peptidase M56, BlaR1	908	4	0.004405286
Cthe_1558	-	transcriptional regulator, BadM/Rrf2 family	455	2	0.004395604
Cthe_0044	-	cellulosome enzyme, dockerin type I	1601	7	0.004372267
Cthe_1449	-	transcriptional regulator, MarR family	458	2	0.004366812
Cthe_1507	-	protein of unknown function UPF0131	458	2	0.004366812
Cthe_0273	-	metal dependent phosphohydrolase	1604	7	0.004364090
Cthe_1904	-	amino acid adenylation domain	5501	24	0.004362843
Cthe_2617	+	peptidase M23B	920	4	0.004347826

المنارة للاستشارات
www.manaraa.com

Cthe_1727	-	hypothetical protein	923	4	0.004333694
Cthe_2479	+	Lj928 prophage protein	923	4	0.004333694
Cthe_2546	-		693	3	0.004329004
Cthe_0126	+	CheC-like protein	464	2	0.004310345
Cthe_2308	+	CheC-like protein	464	2	0.004310345
Cthe_2748	+	SsrA-binding protein	464	2	0.004310345
Cthe_0107	+	Riboflavin synthase(EC:2.5.1.9)	467	2	0.004282655
Cthe_1081	+	hypothetical protein	467	2	0.004282655
Cthe_2112	+	two component transcriptional regulator, winged helix family	701	3	0.004279601
Cthe_0440	-	Domain of unknown function DUF1727	1403	6	0.004276550
Cthe_1349	+	Undecaprenyl-phosphate galactose phosphotransferase(EC:2.7.8.6)	1403	6	0.004276550
Cthe_1534	-	hypothetical protein	1637	7	0.004276115
Cthe_2357	+	hypothetical protein	1172	5	0.004266212
Cthe_1793	-	protein of unknown function DUF1624	704	3	0.004261364
Cthe_2806	-	hypothetical protein	704	3	0.004261364
Cthe_1114	-	Tn7-like transposition protein D	1886	8	0.004241782
Cthe_0705	+	hypothetical protein	236	1	0.004237288
Cthe_1314	-	DNA-directed RNA polymerase, omega subunit	236	1	0.004237288
Cthe_0193	+	sodium/hydrogen exchanger	1184	5	0.004222973
Cthe_0924	-	hypothetical protein	713	3	0.004207574
Cthe_0066	+	hypothetical protein	476	2	0.004201681
Cthe_1667	-	ABC-2 type transporter	716	3	0.004189944
Cthe_0171	-	hypothetical protein	239	1	0.004184100
Cthe_0228	+	hypothetical protein	479	2	0.004175365
Cthe_1276	-	pantetheine-phosphate adenylyltransferase(EC:2.7.7.3)	479	2	0.004175365
Cthe_3111	+	hypothetical protein	479	2	0.004175365
Cthe_3231	+	hypothetical protein	479	2	0.004175365
Cthe_1650	-	DNA-directed DNA polymerase(EC:2.7.7.7)	1922	8	0.004162331
Cthe_3026	-	GreA/GreB family elongation factor	482	2	0.004149378
Cthe_0136	+	4'-phosphopantetheinyl transferase	725	3	0.004137931
Cthe_1178	-	isochorismatase hydrolase	725	3	0.004137931
Cthe_2951	+	hypothetical protein	725	3	0.004137931
Cthe_0832	-	exodeoxyribonuclease VII, small subunit	242	1	0.004132231
Cthe_1830	-	hypothetical protein	1214	5	0.004118616
Cthe_3146	+	peptidase M56, BlaR1	2186	9	0.004117109
Cthe_1679	-	Methyltransferase type 11	731	3	0.004103967
Cthe_0180	+	crossover junction endodeoxyribonuclease RuvC	488	2	0.004098361
Cthe_1069	-	protein of unknown function UPF0054	488	2	0.004098361
Cthe_2216	-	hypothetical protein	488	2	0.004098361
Cthe_2671	-	transposase, mutator type	1223	5	0.004088307
Cthe_0131	+	hypothetical protein	245	1	0.004081633
Cthe_RF0079	-	(miscRNA)	245	1	0.004081633
Cthe_2335	-	polysaccharide biosynthesis protein	1472	6	0.004076087
Cthe_0857	+	protein of unknown function DUF107	491	2	0.004073320

www.manaraa.com

Cthe_1903	-	polysaccharide deacetylase	737	3	0.004070556
Cthe_2202	-	hypothetical protein	737	3	0.004070556
Cthe_RF0091	-	(miscRNA)	246	1	0.004065041
Cthe_1086	+	dTDP-glucose 4,6-dehydratase(EC:4.2.1.46)	986	4	0.004056795
Cthe_3233	+	hypothetical protein	740	3	0.004054054
Cthe_0127	+	RNA methyltransferase, TrmH family, group 2	494	2	0.004048583
Cthe_2398	+	putative spore coat protein	989	4	0.004044489
Cthe_1339	-	type II secretion system protein E	1238	5	0.004038772
Cthe_1510	-	short-chain dehydrogenase/reductase SDR	743	3	0.004037685
Cthe_2641	-	N-acylneuraminate-9-phosphate synthase(EC:2.5.1.57)	992	4	0.004032258
Cthe_0774	-	RNA chaperone Hfq	248	1	0.004032258
Cthe_2565	-	hypothetical protein	497	2	0.004024145
Cthe_3223	+	hypothetical protein	497	2	0.004024145
Cthe_0897	-	putative deoxyguanosinetriphosphate triphosphohydrolase	995	4	0.004020101
Cthe_1171	-	Serine-type D-Ala-D-Ala carboxypeptidase(EC:3.4.16.4)	1247	5	0.004009623
Cthe_0315	+	DNA-directed RNA polymerase sigma factor	749	3	0.004005340
Cthe_RF0092	-	(miscRNA)	250	1	0.004000000
Cthe_1928	-	intein	1001	4	0.003996004
Cthe_1529	-		1252	5	0.003993610
Cthe_1638	-	ParB-like nuclease	1253	5	0.003990423
Cthe_0476	+	hypothetical protein	503	2	0.003976143
Cthe_1438	-	RNA polymerase, sigma-24 subunit, ECF subfamily	503	2	0.003976143
Cthe_2687	+	thioesterase superfamily	503	2	0.003976143
Cthe_1378	-	response regulator receiver modulated metal dependent phosphohydrolase	1520	6	0.003947368
Cthe_2563	-	hypothetical protein	1781	7	0.003930376
Cthe_2037	-		1528	6	0.003926702
Cthe_1535	-	hypothetical protein	764	3	0.003926702
Cthe_1328	-	Stage II sporulation P	1277	5	0.003915427
Cthe_1506	-	metal dependent phosphohydrolase	1022	4	0.003913894
Cthe_2298	-	CRISPR-associated protein Cas4	512	2	0.003906250
Cthe_3225	+	hypothetical protein	512	2	0.003906250
Cthe_1038	-	spore protease(EC:3.4.24.78)	1025	4	0.003902439
Cthe_2553	-	Radical SAM	1025	4	0.003902439
Cthe_2636	-	integral membrane protein MviN	1538	6	0.003901170
Cthe_0038	+	chemotaxis protein cheW	257	1	0.003891051
Cthe_2520	+	hypothetical protein	257	1	0.003891051
Cthe_0190	+	proteinase inhibitor I4, serpin	1802	7	0.003884573
Cthe_2583	+	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase(EC:2.7.6.3)	515	2	0.003883495
Cthe_0448	+	RNA polymerase, sigma 28 subunit	773	3	0.003880983
Cthe_0452	+	cellulosome anchoring protein, cohesin region	773	3	0.003880983
Cthe_0923	-	hypothetical protein	773	3	0.003880983
Cthe_2110	+	ABC transporter related protein	773	3	0.003880983

www.manaraa.com

Cthe_1701	-	Recombinase	1289	5	0.003878976
Cthe_2132	+	abortive infection protein, putative	1298	5	0.003852080
Cthe_2562	-	glucose-1-phosphate cytidylyltransferase(EC:2.7.7.33)	779	3	0.003851091
Cthe_2803	+	binding-protein-dependent transport systems inner membrane component	779	3	0.003851091
Cthe_0844	-	stage III sporulation protein spoIIIAB	521	2	0.003838772
Cthe_2048	-	hypothetical protein	521	2	0.003838772
Cthe_0150	+	RNA modification enzyme, MiaB family	1304	5	0.003834356
Cthe_1461	+	FAD dependent oxidoreductase	1565	6	0.003833866
Cthe_3234	-	hypothetical protein	785	3	0.003821656
Cthe_0603	+	GCN5-related N-acetyltransferase	524	2	0.003816794
Cthe_3045	+	FHA domain containing protein	524	2	0.003816794
Cthe_1083	+	Putative homoserine kinase type II (protein kinase fold)-like protein	1049	4	0.003813155
Cthe_3201	+	CRISPR-associated protein, Csh1 family	1841	7	0.003802281
Cthe_0096	-	protein of unknown function UPF0180	263	1	0.003802281
Cthe_2765	+	hypothetical protein	1580	6	0.003797468
Cthe_2893	+	response regulator receiver protein	1055	4	0.003791469
Cthe_0090	+	hypothetical protein	530	2	0.003773585
Cthe_0965	-	ATP:corrinoid adenosyltransferase BtuR/CobO/CobP	530	2	0.003773585
Cthe_2894	+	hypothetical protein	530	2	0.003773585
Cthe_0238	+	hypothetical protein	266	1	0.003759398
Cthe_0294	+		266	1	0.003759398
Cthe_0851	-	hypothetical protein	266	1	0.003759398
Cthe_2678	-	hypothetical protein	533	2	0.003752345
Cthe_1140	-	hypothetical protein	800	3	0.003750000
Cthe_2691	+	galactoside O-acetyltransferase	536	2	0.003731343
Cthe_1338	-	Flp pilus assembly protein TadB-like protein	806	3	0.003722084
Cthe_1042	-	hypothetical protein	539	2	0.003710575
Cthe_1013	-	protein of unknown function UPF0118	1082	4	0.003696858
Cthe_3171	+	S-layer domain-like protein	2447	9	0.003677973
Cthe_RF0081	+	(miscRNA)	272	1	0.003676471
Cthe_3116	+	mannose-6-phosphate isomerase, class I(EC:5.3.1.8)	1091	4	0.003666361
Cthe_2560	+	dTDP-4-dehydrorhamnose 3,5-epimerase(EC:5.1.3.13)	548	2	0.003649635
Cthe_0522	-	hypothetical protein	275	1	0.003636364
Cthe_2277	-	hypothetical protein	827	3	0.003627570
Cthe_2467	+	replicative DNA helicase	836	3	0.003588517
Cthe_0514	+	DNA methylase N-4/N-6	1115	4	0.003587444
Cthe_1473	-	NADPH-dependent FMN reductase	563	2	0.003552398
Cthe_1974	-	hypothetical protein	563	2	0.003552398
Cthe_0983	-	cell cycle protein	1130	4	0.003539823
Cthe_0696	+	putative rRNA methylase	566	2	0.003533569
Cthe_2286	+	phage integrase	566	2	0.003533569
Cthe_2970	+	ABC transporter related protein	851	3	0.003525264
Cthe_2494	+	hypothetical protein	284	1	0.003521127

المنارة للاستشارات

www.manaraa.com

Cthe_3055	+	hypothetical protein	284	1	0.003521127
Cthe_2384	-	S-layer-like domain containing protein	569	2	0.003514938
Cthe_1604	-	phosphate ABC transporter, inner membrane subunit PstC	854	3	0.003512881
Cthe_1876	-	Tn7-like transposition protein A	854	3	0.003512881
Cthe_2167	-	glycoside hydrolase, family 8	1142	4	0.003502627
Cthe_3056	+	transposase, IS204/IS1001/IS1096/IS1165	1430	5	0.003496503
Cthe_0204	+	protein of unknown function DUF1294	287	1	0.003484321
Cthe_1891	-	hypothetical protein	287	1	0.003484321
Cthe_1585	-	signal transduction histidine kinase, LytS	1724	6	0.003480278
Cthe_1685	-	ABC transporter related protein	1724	6	0.003480278
Cthe_1291	-	amidohydrolase 2	866	3	0.003464203
Cthe_0367	+	NADP oxidoreductase, coenzyme F420-dependent	869	3	0.003452244
Cthe_0164	+	protein of unknown function UPF0044	290	1	0.003448276
Cthe_3219	+	CRISPR-associated protein Cas2	290	1	0.003448276
Cthe_1435	-	Phosphoglycerate mutase	581	2	0.003442341
Cthe_2316	-	hypothetical protein	581	2	0.003442341
Cthe_2982	+	hypothetical protein	872	3	0.003440367
Cthe_0022	+	8-amino-7-oxononanoate synthase(EC:2.3.1.47)	1166	4	0.003430532
Cthe_0113	+	Uncharacterized P-loop ATPase protein UPF0042	875	3	0.003428571
Cthe_2764	+	TROVE domain containing protein	1463	5	0.003417635
Cthe_0590	+	transposase IS3/IS911	293	1	0.003412969
Cthe_2514	+	hypothetical protein	293	1	0.003412969
Cthe_1441	-	hypothetical protein	587	2	0.003407155
Cthe_0314	+	glycosyltransferase 28-like protein	1178	4	0.003395586
Cthe_1268	-	histidine kinase	1178	4	0.003395586
Cthe_3153	+	Phosphoglycerate mutase	590	2	0.003389831
Cthe_1883	-	SMC protein-like protein	2066	7	0.003388190
Cthe_1021	-	stage IV sporulation protein A	1478	5	0.003382950
Cthe_1851	-	protein of unknown function DUF1113	593	2	0.003372681
Cthe_2420	+	HD superfamily phosphohydrolases-like protein	1781	6	0.003368894
Cthe_0446	+	peptidase U4, sporulation factor SpoIIGA	893	3	0.003359462
Cthe_1623	-	phage major tail protein, phi13 family	596	2	0.003355705
Cthe_1216	+	ATP-dependent Clp protease adaptor protein ClpS	299	1	0.003344482
Cthe_0657	-	hypothetical protein	599	2	0.003338898
Cthe_1310	+	Accessory gene regulator B	599	2	0.003338898
Cthe_1153	-	diguanylate cyclase	899	3	0.003337041
Cthe_1161	-	hypothetical protein	2402	8	0.003330558
Cthe_0639	-	glycoside hydrolase, family 13-like protein	1202	4	0.003327787
Cthe_2044	-	hypothetical protein	602	2	0.003322259
Cthe_2493	+	phage-like element PBSX protein	602	2	0.003322259
Cthe_2791	+	binding-protein-dependent transport systems inner membrane component	908	3	0.003303965
Cthe_2038	-	cellulosome enzyme, dockerin type I	2423	8	0.003301692
Cthe_3126	+	NUDIX hydrolase	911	3	0.003293085

المنارة للاستشارات
www.manaraa.com

Cthe_0746	-	conserved hypothetical protein-signal peptide prediction	608	2	0.003289474
Cthe_2503	+	Cupin 2, conserved barrel	608	2	0.003289474
Cthe_1924	-		1218	4	0.003284072
Cthe_1970	-	hypothetical protein	305	1	0.003278689
Cthe_1681	-		611	2	0.003273322
Cthe_0801	-	transposase, mutator type	1223	4	0.003270646
Cthe_3216	+	transposase, mutator type	1223	4	0.003270646
Cthe_1493	-	transcriptional regulator, PadR-like family	614	2	0.003257329
Cthe_3220	+	CRISPR-associated protein Cas4	614	2	0.003257329
Cthe_0280	+	Coat F	308	1	0.003246753
Cthe_1470	+	RNA polymerase, sigma-24 subunit, ECF subfamily	617	2	0.003241491
Cthe_2274	-	hypothetical protein	620	2	0.003225806
Cthe_2572	-	hypothetical protein	620	2	0.003225806
Cthe_1059	-	hypothetical protein	311	1	0.003215434
Cthe_1698	-	hypothetical protein	311	1	0.003215434
Cthe_1725	-	hypothetical protein	311	1	0.003215434
Cthe_2056	-	hypothetical protein	941	3	0.003188098
Cthe_2034	-	serine/threonine protein kinase	1883	6	0.003186405
Cthe_2375	+	hypothetical protein	314	1	0.003184713
Cthe_2898	+	anti-sigma-factor antagonist	314	1	0.003184713
Cthe_0737	-	Phosphopantethiene-protein transferase	629	2	0.003179650
Cthe_1077	+	hypothetical protein	629	2	0.003179650
Cthe_1335	-	peptidase A24A, prepilin type IV	629	2	0.003179650
Cthe_0840	-	hypothetical protein	632	2	0.003164557
Cthe_2099	-	nucleoside recognition	632	2	0.003164557
Cthe_3235	-	hypothetical protein	950	3	0.003157895
Cthe_1781	-	hypothetical protein	317	1	0.003154574
Cthe_0798	-	lipolytic enzyme, G-D-S-L	1586	5	0.003152585
Cthe_1053	+	L-lactate dehydrogenase(EC:1.1.1.27)	953	3	0.003147954
Cthe_0029	+	hypothetical protein	638	2	0.003134796
Cthe_1497	-	hypothetical protein	638	2	0.003134796
Cthe_3140	-	transcriptional regulator, TetR family	638	2	0.003134796
Cthe_0398	+	protein of unknown function DUF74	320	1	0.003125000
Cthe_1854	-	methylated-DNA--protein-cysteine methyltransferase	320	1	0.003125000
Cthe_2266	+	Vacuolar H+-transporting two-sector ATPase, F subunit	320	1	0.003125000
Cthe_0108	-	type IV pilus assembly PilZ	641	2	0.003120125
Cthe_1148	-	hypothetical protein	1283	4	0.003117693
Cthe_2638	-	O-antigen polymerase	1286	4	0.003110420
Cthe_1958	-	UV-endonuclease UvdE	965	3	0.003108808
Cthe_2231	-	Methyltransferase type 11	644	2	0.003105590
Cthe_2145	+		967	3	0.003102378
Cthe_1922	+	hypothetical protein	968	3	0.003099174
Cthe_1173	-	type IV pilus assembly PilZ	647	2	0.003091190
Cthe_3221	+	hypothetical protein	647	2	0.003091190

المنارة للاستشارات
www.manaraa.com

Cthe_0558	+	hypothetical protein	650	2	0.003076923
Cthe_1897	-	hypothetical protein	326	1	0.003067485
Cthe_1966	-	hypothetical protein	326	1	0.003067485
Cthe_2358	-	Accessory gene regulator B	653	2	0.003062787
Cthe_2645	-	NAD-dependent epimerase/dehydratase	980	3	0.003061224
Cthe_0648	-	glutamyl-tRNA synthetase	1640	5	0.003048780
Cthe_0123	+	Stage V sporulation AD	989	3	0.003033367
Cthe_0049	+	transcriptional regulator, XRE family	662	2	0.003021148
Cthe_2566	-	Radical SAM	1325	4	0.003018868
Cthe_1643	-	Uncharacterized phage-associated protein-like protein	998	3	0.003006012
Cthe_2652	+	lipopolysaccharide biosynthesis	668	2	0.002994012
Cthe_2675	+	hypothetical protein	668	2	0.002994012
Cthe_1046	-	extracellular solute-binding protein, family 1	1340	4	0.002985075
Cthe_0820	-	hypothetical protein	2012	6	0.002982107
Cthe_1505	-	S-layer-like domain containing protein	671	2	0.002980626
Cthe_1702	-	cell wall hydrolase/autolysin	671	2	0.002980626
Cthe_1759	-	hypothetical protein	671	2	0.002980626
Cthe_1705	-	hypothetical protein	1688	5	0.002962085
Cthe_0651	+	hypothetical protein	338	1	0.002958580
Cthe_1994	-	hypothetical protein	338	1	0.002958580
Cthe_1664	-	transcriptional regulator, MerR family	677	2	0.002954210
Cthe_2949	-	Pectinesterase	1703	5	0.002935995
Cthe_2780	+	hypothetical protein	341	1	0.002932551
Cthe_0780	-	HAD-superfamily hydrolase, subfamily IA, variant 3	683	2	0.002928258
Cthe_0400	-	hypothetical protein	1028	3	0.002918288
Cthe_2505	-	transcriptional regulator, LacI family	1028	3	0.002918288
Cthe_1677	-	Abortive infection protein	686	2	0.002915452
Cthe_2564	-	hypothetical protein	686	2	0.002915452
Cthe_2744	-	Lytic transglycosylase, catalytic	686	2	0.002915452
Cthe_0227	+	hypothetical protein	344	1	0.002906977
Cthe_2805	-	Carbohydrate-binding, CenC-like protein	1382	4	0.002894356
Cthe_1084	+	spore coat protein	1037	3	0.002892960
Cthe_2637	-	glycosyl transferase, group 1	1037	3	0.002892960
Cthe_1637	-	virulence-related protein	692	2	0.002890173
Cthe_1305	-	Cupin 2, conserved barrel	347	1	0.002881844
Cthe_1463	-	protein of unknown function DUF891	347	1	0.002881844
Cthe_2499	+	hypothetical protein	347	1	0.002881844
Cthe_2813	+	two component transcriptional regulator, winged helix family	695	2	0.002877698
Cthe_0807	-	response regulator receiver modulated CheB methylesterase(EC:3.1.1.61)	1043	3	0.002876318
Cthe_1408	+	two component transcriptional regulator, winged helix family	698	2	0.002865330
Cthe_3099	-	glycosyl transferase, family 2	698	2	0.002865330
Cthe_2985	-	hypothetical protein	350	1	0.002857143
Cthe_0257	+	twitching motility protein	1052	3	0.002851711
Cthe_1480	+	hypothetical protein	2105	6	0.002850356

www.manaraa.com

Cthe_2777	+	nicotinamide mononucleotide transporter PnuC	704	2	0.002840909
Cthe_0346	+	hypothetical protein	353	1	0.002832861
Cthe_1198	-	hypothetical protein	353	1	0.002832861
Cthe_1394	-	two component transcriptional regulator, winged helix family	710	2	0.002816901
Cthe_1641	-	HNH endonuclease	359	1	0.002785515
Cthe_0693	+	Pseudouridine synthase, Rsu	719	2	0.002781641
Cthe_2677	+	histidine kinase	719	2	0.002781641
Cthe_1814	-	Urease accessory protein UreF	722	2	0.002770083
Cthe_2303	-	CRISPR-associated protein Cas6	722	2	0.002770083
Cthe_0457	+	protein of unknown function DUF307	362	1	0.002762431
Cthe_1208	-	GCN5-related N-acetyltransferase	362	1	0.002762431
Cthe_3145	+	transcriptional repressor, CopY family	362	1	0.002762431
Cthe_1379	+	oxidoreductase-like protein	1088	3	0.002757353
Cthe_1824	-	two component transcriptional regulator, AraC family	1088	3	0.002757353
Cthe_1728	-	DNA-cytosine methyltransferase(EC:2.1.1.37)	1451	4	0.002756720
Cthe_3213	+	hypothetical protein	365	1	0.002739726
Cthe_3144	-	transcriptional regulator, AraC family	731	2	0.002735978
Cthe_0434	+	ATPase	1100	3	0.002727273
Cthe_3130	+	hypothetical protein	734	2	0.002724796
Cthe_0167	+	response regulator receiver protein	368	1	0.002717391
Cthe_0668	+	spore germination protein	1109	3	0.002705140
Cthe_2230	-	acylneuraminate cytidylyltransferase	740	2	0.002702703
Cthe_3129	+	Citrate transporter	1112	3	0.002697842
Cthe_1770	-	hypothetical protein	371	1	0.002695418
Cthe_2966	+	hypothetical protein	746	2	0.002680965
Cthe_1120	-	hypothetical protein	374	1	0.002673797
Cthe_1459	-	hypothetical protein	1127	3	0.002661934
Cthe_1651	-	hypothetical protein	752	2	0.002659574
Cthe_0010	+	hypothetical protein	377	1	0.002652520
Cthe_0301	+	protein of unknown function DUF1212	755	2	0.002649007
Cthe_0562	+	protein of unknown function DUF1312	380	1	0.002631579
Cthe_1214	-	hypothetical protein	380	1	0.002631579
Cthe_1745	-	transcriptional regulator, XRE family	380	1	0.002631579
Cthe_3072	-	acyl-ACP thioesterase	761	2	0.002628121
Cthe_2975	-	DNA-directed RNA polymerase sigma factor	764	2	0.002617801
Cthe_3110	+	UBA/THIF-type NAD/FAD binding fold	764	2	0.002617801
Cthe_0077	+	hypothetical protein	383	1	0.002610966
Cthe_0330	+	hypothetical protein	383	1	0.002610966
Cthe_2015	-	hypothetical protein	383	1	0.002610966
Cthe_1128	-	hypothetical protein	386	1	0.002590674
Cthe_3019	-	4Fe-4S ferredoxin, iron-sulfur binding	386	1	0.002590674
Cthe_1422	+	RDD domain containing protein	773	2	0.002587322
Cthe_2076	+	periplasmic sensor signal transduction histidine kinase	1163	3	0.002579536
Cthe_1875	-	HMG-I and HMG-Y, DNA-binding	776	2	0.002577320
Cthe_0473	+	flagellar operon protein	389	1	0.002570694

المنارة للاستشارات
www.manaraa.com

Cthe_2846	+	hypothetical protein	389	1	0.002570694
Cthe_2644	-	DegT/DnrJ/EryC1/StrS aminotransferase	1169	3	0.002566296
Cthe_0717	+	MCP methyltransferase, CheR-type(EC:2.1.1.80)	782	2	0.002557545
Cthe_2640	-	UDP-N-acetylglucosamine 2-epimerase(EC:5.1.3.14)	1175	3	0.002553191
Cthe_0006	+	hypothetical protein	392	1	0.002551020
Cthe_3089	-	UspA	392	1	0.002551020
Cthe_1364	-	hypothetical protein	1178	3	0.002546689
Cthe_2276	+	AAA ATPase, central region	1178	3	0.002546689
Cthe_2624	-		394	1	0.002538071
Cthe_1888	-	hypothetical protein	1583	4	0.002526848
Cthe_1737	-	phage NTP-binding protein	1592	4	0.002512563
Cthe_1112	-	transcriptional regulator, ArsR family	398	1	0.002512563
Cthe_1487	-	hypothetical protein	398	1	0.002512563
Cthe_0268	-	hypothetical protein	797	2	0.002509410
Cthe_1649	-	BRO-like protein	797	2	0.002509410
Cthe_1900	+		797	2	0.002509410
Cthe_2790	+	ABC transporter related protein(EC:3.6.3.25)	806	2	0.002481390
Cthe_1350	-	single-strand binding protein	404	1	0.002475248
Cthe_2665	-	Zn-finger containing protein	404	1	0.002475248
Cthe_0048	+	copper amine oxidase-like protein	809	2	0.002472188
Cthe_2158	+		1219	3	0.002461034
Cthe_2397	+	hypothetical protein	1220	3	0.002459016
Cthe_2609	+	ATP synthase F1, epsilon subunit	407	1	0.002457002
Cthe_2646	-	hypothetical protein	1223	3	0.002452984
Cthe_0808	-	MCP methyltransferase, CheR-type(EC:2.1.1.80)	818	2	0.002444988
Cthe_2627	+	purine operon repressor, PurR	818	2	0.002444988
Cthe_0716	+	Nucleoside-diphosphate kinase(EC:2.7.4.6)	410	1	0.002439024
Cthe_2169	-	metallophosphoesterase	821	2	0.002436054
Cthe_2237	-	flagellin-like protein	821	2	0.002436054
Cthe_1174	+	protein of unknown function DUF155	824	2	0.002427184
Cthe_1156	+	zinc finger, CHC2-type	2063	5	0.002423655
Cthe_2007	-		413	1	0.002421308
Cthe_1582	+	cytidine deaminase	416	1	0.002403846
Cthe_1652	-	hypothetical protein	416	1	0.002403846
Cthe_0754	-	hypothetical protein	833	2	0.002400960
Cthe_1123	-	hypothetical protein	836	2	0.002392344
Cthe_1788	+	glycosyl transferase, family 2	1256	3	0.002388535
Cthe_0546	+	ferric uptake regulator, Fur family	419	1	0.002386635
Cthe_1612	-	toxin secretion/phage lysis holin	419	1	0.002386635
Cthe_2469	+	phage transcriptional regulator, RinA family	422	1	0.002369668
Cthe_3187	-	putative anti-sigma regulatory factor, serine/threonine protein kinase	422	1	0.002369668
Cthe_1117	-	Tn7-like transposition protein A	845	2	0.002366864
Cthe_1735	-	phage DNA polymerase	1691	4	0.002365464
Cthe_2532	+	sulfate ABC transporter, inner membrane subunit CysT	848	2	0.002358491

المنارة للاستشارات
www.manaraa.com

Cthe_0264	+	hypothetical protein	425	1	0.002352941
Cthe_1648	-	hypothetical protein	425	1	0.002352941
Cthe_2440	+	hypothetical protein	425	1	0.002352941
Cthe_2817	+	CheW protein	425	1	0.002352941
Cthe_0024	+	biotin biosynthesis protein BioC	851	2	0.002350176
Cthe_0545	+	prepilin peptidase CpaA	428	1	0.002336449
Cthe_1663	-	Endoribonuclease L-PSP	428	1	0.002336449
Cthe_1719	-	phage major capsid protein, HK97 family	1289	3	0.002327386
Cthe_2613	+	S-layer-like domain containing protein	860	2	0.002325581
Cthe_3224	+	hypothetical protein	860	2	0.002325581
Cthe_3217	+	hypothetical protein	1292	3	0.002321981
Cthe_0080	-	CheW protein	431	1	0.002320186
Cthe_2315	+	DNA mismatch endonuclease vsr	434	1	0.002304147
Cthe_2901	+	putative anti-sigma regulatory factor, serine/threonine protein kinase	434	1	0.002304147
Cthe_1118	-	hypothetical protein	869	2	0.002301496
Cthe_1877	-	hypothetical protein	869	2	0.002301496
Cthe_0420	+	dipicolinate synthase	872	2	0.002293578
Cthe_1758	+	protein of unknown function DUF583	437	1	0.002288330
Cthe_2154	+	Methyltransferase type 11	875	2	0.002285714
Cthe_2400	+	peptidase U57, YabG	875	2	0.002285714
Cthe_1134	-	hypothetical protein	440	1	0.002272727
Cthe_1530	-	transposase, IS204/IS1001/IS1096/IS1165	440	1	0.002272727
Cthe_0086	+	hypothetical protein	884	2	0.002262443
Cthe_2524	+	transcriptional regulator, BadM/Rrf2 family	443	1	0.002257336
Cthe_1176	-	CheC-like protein	446	1	0.002242152
Cthe_2878	+	peptidase C39, bacteriocin processing	893	2	0.002239642
Cthe_0813	-	peptidase S55, SpoIVB	1340	3	0.002238806
Cthe_3125	+	heat shock protein Hsp20	449	1	0.002227171
Cthe_0553	-	transcriptional regulator, LysR family	899	2	0.002224694
Cthe_1429	-	hypothetical protein	899	2	0.002224694
Cthe_1465	-	Redoxin	452	1	0.002212389
Cthe_1822	-	inner-membrane translocator	908	2	0.002202643
Cthe_0395	+	RbsD or FucU transport	455	1	0.002197802
Cthe_1255	+	hypothetical protein	455	1	0.002197802
Cthe_3180	+	hypothetical protein	911	2	0.002195390
Cthe_2272	+	hypothetical protein	914	2	0.002188184
Cthe_0437	+	hypothetical protein	458	1	0.002183406
Cthe_0469	+	flagellar export protein FliJ	458	1	0.002183406
Cthe_0964	+	amino acid-binding ACT	458	1	0.002183406
Cthe_1792	-	Firmicute transcriptional repressor of class III stress genes	461	1	0.002169197
Cthe_1815	-	UreE urease accessory-like protein	461	1	0.002169197
Cthe_1951	+	hypothetical protein	461	1	0.002169197
Cthe_0788	-	DivIVA	464	1	0.002155172
Cthe_2577	-	adenylate cyclase	470	1	0.002127660
Cthe_2746	+	protein of unknown function DUF402	470	1	0.002127660

www.manaraa.com

Cthe_1596	+	transcriptional regulator, MarR family	473	1	0.002114165
Cthe_2275	-	hypothetical protein	947	2	0.002111932
Cthe_2486	+	hypothetical protein	476	1	0.002100840
Cthe_0671	-	hypothetical protein	479	1	0.002087683
Cthe_1402	-	iron (metal) dependent repressor, DtxR family	479	1	0.002087683
Cthe_0382	+	hypothetical protein	482	1	0.002074689
Cthe_0525	+	hypothetical protein	482	1	0.002074689
Cthe_1723	-	phage terminase, small subunit, putative, P27 family	482	1	0.002074689
Cthe_2954	-	hypothetical protein	482	1	0.002074689
Cthe_0072	+	phage shock protein C, PspC	485	1	0.002061856
Cthe_2502	+	CMP/dCMP deaminase, zinc-binding	485	1	0.002061856
Cthe_2580	+	phosphodiesterase, MJ0936 family	485	1	0.002061856
Cthe_2235	-	protein of unknown function DUF115	1946	4	0.002055498
Cthe_2549	+	cellulosome enzyme, dockerin type I	974	2	0.002053388
Cthe_0532	+	hydrolase (HAD superfamily)-like protein	491	1	0.002036660
Cthe_1359	-	glycosyltransferase 28-like protein	494	1	0.002024291
Cthe_1703	-	HNH endonuclease	494	1	0.002024291
Cthe_0728	-	phage integrase	992	2	0.002016129
Cthe_1669	+	DNA recombinase, putative	497	1	0.002012072
Cthe_2525	+	glutamyl-tRNA reductase	995	2	0.002010050
Cthe_0709	+	hypothetical protein	500	1	0.002000000
Cthe_0939	-	hypothetical protein	500	1	0.002000000
Cthe_2245	-	hypothetical protein	500	1	0.002000000
Cthe_1300	+	hypothetical protein	1004	2	0.001992032
Cthe_1611	-	Peptidoglycan-binding LysM	1004	2	0.001992032
Cthe_2616	+	SpoIID/LytB domain containing protein	1004	2	0.001992032
Cthe_1293	-	metal dependent phosphohydrolase	1007	2	0.001986097
Cthe_2556	-	Radical SAM	1013	2	0.001974334
Cthe_3115	+	hypothetical protein	1013	2	0.001974334
Cthe_1722	-	phage Terminase	2549	5	0.001961554
Cthe_2552	-	Radical SAM	1022	2	0.001956947
Cthe_0563	+	Heptaprenyl diphosphate synthase component I	512	1	0.001953125
Cthe_2321	-	hypothetical protein	515	1	0.001941748
Cthe_2551	-	Radical SAM	1034	2	0.001934236
Cthe_2152	-	hypothetical protein	518	1	0.001930502
Cthe_2957	+	hypothetical protein	521	1	0.001919386
Cthe_0702	+	hypothetical protein	527	1	0.001897533
Cthe_0750	-	spermidine/putrescine ABC transporter ATPase subunit	1058	2	0.001890359
Cthe_0968	-	UvrD/REP helicase	2120	4	0.001886792
Cthe_0007	+	hypothetical protein	533	1	0.001876173
Cthe_2098	-	nucleoside recognition	533	1	0.001876173
Cthe_2875	+	sigma 54 modulation protein/ribosomal protein S30EA	533	1	0.001876173
Cthe_0595	+	hypothetical protein	1067	2	0.001874414
Cthe_0589	+		1068	2	0.001872659
Cthe_1809	-	RNA polymerase, sigma-24 subunit, ECF	539	1	0.001855288

المنارة للاستشارات
www.manaraa.com

		subfamily			
Cthe_1520	-	hypothetical protein	542	1	0.001845018
Cthe_1150	-	hypothetical protein	1085	2	0.001843318
Cthe_2464	+	hypothetical protein	551	1	0.001814882
Cthe_1587	-	ABC transporter related protein	1103	2	0.001813237
Cthe_2288	-	PEBP	554	1	0.001805054
Cthe_2710	+	Uncharacterized membrane protein-like protein	1109	2	0.001803427
Cthe_2478	+	phage minor structural GP20	557	1	0.001795332
Cthe_2667	+	transcriptional regulator, TetR family	557	1	0.001795332
Cthe_1478	+	transcriptional regulator, TetR family	560	1	0.001785714
Cthe_0125	+	stage V sporulation AE	563	1	0.001776199
Cthe_2544	-	RNA polymerase, sigma-24 subunit, ECF subfamily	563	1	0.001776199
Cthe_3151	+	Adenosylcobinamide-phosphate guanylyltransferase(EC:2.7.7.62)	563	1	0.001776199
Cthe_2079	+	signal peptidase I	566	1	0.001766784
Cthe_2166	+	putative PAS/PAC sensor protein	1136	2	0.001760563
Cthe_0689	+	RDD domain containing protein	569	1	0.001757469
Cthe_1593	-	NUDIX hydrolase	569	1	0.001757469
Cthe_1553	-	hypothetical protein	572	1	0.001748252
Cthe_1717	-	hypothetical protein	572	1	0.001748252
Cthe_0753	-	hypothetical protein	575	1	0.001739130
Cthe_1295	-	5-formyltetrahydrofolate cyclo-ligase	575	1	0.001739130
Cthe_1614	-	hypothetical protein	575	1	0.001739130
Cthe_2555	-	methyltransferase FkbM family	1151	2	0.001737619
Cthe_1210	+	hypothetical protein	578	1	0.001730104
Cthe_2241	-	hypothetical protein	578	1	0.001730104
Cthe_2835	+	hypothetical protein	578	1	0.001730104
Cthe_3137	-	hypothetical protein	584	1	0.001712329
Cthe_1828	-	hypothetical protein	593	1	0.001686341
Cthe_0383	+	protein of unknown function DUF214	596	1	0.001677852
Cthe_0523	+	Resolvase-like protein	596	1	0.001677852
Cthe_1672	-	hypothetical protein	596	1	0.001677852
Cthe_3028	-	Pyridoxal-dependent decarboxylase	1196	2	0.001672241
Cthe_0386	+	hypothetical protein	605	1	0.001652893
Cthe_0181	+	Holliday junction DNA helicase RuvA	608	1	0.001644737
Cthe_0378	+	hypothetical protein	608	1	0.001644737
Cthe_0690	+	hypothetical protein	608	1	0.001644737
Cthe_1479	+	hypothetical protein	608	1	0.001644737
Cthe_2133	+	hypothetical protein	611	1	0.001636661
Cthe_1533	+	transposase, mutator type	1223	2	0.001635323
Cthe_0941	-	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase(EC:2.7.8.5)	617	1	0.001620746
Cthe_1729	-	DNA methylase N-4/N-6	1235	2	0.001619433
Cthe_3132	+	cellulosome enzyme, dockerin type I	1235	2	0.001619433
Cthe_1113	-	SEC-C motif containing protein	1856	3	0.001616379
Cthe_1694	-	hypothetical protein	620	1	0.001612903
Cthe_0596	+	GTP-binding protein	623	1	0.001605136

المنارة للاستشارات

www.manaraa.com

Cthe_1734	-	phage / plasmid primase, P4 family	1874	3	0.001600854
Cthe_1414	-	hypothetical protein	1253	2	0.001596169
Cthe_2009	-	intein	629	1	0.001589825
Cthe_3192	-	hypothetical protein	629	1	0.001589825
Cthe_2988	-	hypothetical protein	632	1	0.001582278
Cthe_1354	-	O-antigen polymerase	1271	2	0.001573564
Cthe_0319	+	class II aldolase/adducin-like protein	644	1	0.001552795
Cthe_1490	-		646	1	0.001547988
Cthe_0704	-	hypothetical protein	647	1	0.001545595
Cthe_0680	+	Lysine exporter protein (LYSE/YGGA)	653	1	0.001531394
Cthe_3170	+	ABC transporter related protein	653	1	0.001531394
Cthe_0102	-	tRNA (guanine-N(7)-)-methyltransferase(EC:2.1.1.33)	656	1	0.001524390
Cthe_2498	+	N-acetylmuramoyl-L-alanine amidase(EC:3.5.1.28)	656	1	0.001524390
Cthe_2489	+	Peptidoglycan-binding LysM	659	1	0.001517451
Cthe_2752	+	beta-lactamase-like protein	1331	2	0.001502630
Cthe_1893	-	hypothetical protein	668	1	0.001497006
Cthe_3205	+	hypothetical protein	668	1	0.001497006
Cthe_3128	+	iron-sulfur flavoprotein	1337	2	0.001495886
Cthe_0685	+	peptidase M50	671	1	0.001490313
Cthe_0757	-	DNA repair protein RadC	671	1	0.001490313
Cthe_0818	-	hypothetical protein	674	1	0.001483680
Cthe_1528	-	two component transcriptional regulator, winged helix family	674	1	0.001483680
Cthe_0730	+	hypothetical protein	683	1	0.001464129
Cthe_2781	+	hypothetical protein	686	1	0.001457726
Cthe_2639	-	acylneuraminate cytidylyltransferase	692	1	0.001445087
Cthe_1281	+	hypothetical protein	695	1	0.001438849
Cthe_2775	-	transcriptional regulator, Crp/Fnr family	698	1	0.001432665
Cthe_1218	+	Leucyltransferase(EC:2.3.2.6)	701	1	0.001426534
Cthe_1336	-	hypothetical protein	701	1	0.001426534
Cthe_1668	-	ABC transporter related protein	704	1	0.001420455
Cthe_2832	-	transposase, mutator type	704	1	0.001420455
Cthe_1012	+	RNA polymerase, sigma 28 subunit	710	1	0.001408451
Cthe_2654	+	PHP-like protein	710	1	0.001408451
Cthe_3193	+	phospholipase C zinc-binding	710	1	0.001408451
Cthe_0384	+	ABC transporter related protein	713	1	0.001402525
Cthe_3203	+	CRISPR-associated protein Cas5, Hmari subtype	713	1	0.001402525
Cthe_2162	-	peptidase C26	716	1	0.001396648
Cthe_2754	+	beta-lactamase-like protein	716	1	0.001396648
Cthe_0381	+	hypothetical protein	722	1	0.001385042
Cthe_2952	-	IstB-like ATP-binding protein	725	1	0.001379310
Cthe_0697	-	type IV pilus assembly PilZ	734	1	0.001362398
Cthe_1426	+	AzlC-like protein	734	1	0.001362398
Cthe_0406	-	hypothetical protein	737	1	0.001356852
Cthe_1272	+	DNA-directed RNA polymerase sigma factor	737	1	0.001356852

المنارة للاستشارات

www.manaraa.com

Cthe_1720	-	peptidase S14, ClpP	743	1	0.001345895
Cthe_2318	-	hypothetical protein	749	1	0.001335113
Cthe_2557	-	glycosyl transferase, family 2	1499	2	0.001334223
Cthe_1998	-	intein	755	1	0.001324503
Cthe_1896	-	metallophosphoesterase	761	1	0.001314060
Cthe_2968	-	4Fe-4S ferredoxin, iron-sulfur binding	779	1	0.001283697
Cthe_1188	+	protein of unknown function DUF990	782	1	0.001278772
Cthe_2986	-	hypothetical protein	782	1	0.001278772
Cthe_1374	-	copper amine oxidase-like protein	788	1	0.001269036
Cthe_0748	-	binding-protein-dependent transport systems inner membrane component	794	1	0.001259446
Cthe_2853	+	prophage Lp4 protein 7, DNA replication	800	1	0.001250000
Cthe_3222	+	hypothetical protein	800	1	0.001250000
Cthe_1222	-	RNA methyltransferase, TrmH family, group 3	803	1	0.001245330
Cthe_0098	+	peptidase M50	809	1	0.001236094
Cthe_1175	+	hypothetical protein	809	1	0.001236094
Cthe_0664	+	hypothetical protein	812	1	0.001231527
Cthe_2282	-	MCP methyltransferase, CheR-type(EC:2.1.1.80)	812	1	0.001231527
Cthe_2574	+	binding-protein-dependent transport systems inner membrane component	812	1	0.001231527
Cthe_0403	+	DNA-directed RNA polymerase sigma factor	824	1	0.001213592
Cthe_0749	-	binding-protein-dependent transport systems inner membrane component	827	1	0.001209190
Cthe_0329	+	hypothetical protein	836	1	0.001196172
Cthe_2394	-	hypothetical protein	839	1	0.001191895
Cthe_2425	+	MotA/TolQ/ExbB proton channel	839	1	0.001191895
Cthe_0083	+	serine phosphatase	1679	2	0.001191185
Cthe_3131	+	von Willebrand factor, type A	1697	2	0.001178550
Cthe_1187	+	protein of unknown function DUF990	854	1	0.001170960
Cthe_1902	-	Pyridoxal-5'-phosphate-dependent enzyme, beta subunit	878	1	0.001138952
Cthe_2176	+	Abortive infection protein	878	1	0.001138952
Cthe_1167	-	Radical SAM	884	1	0.001131222
Cthe_1159	-	hypothetical protein	911	1	0.001097695
Cthe_2650	-	polysaccharide biosynthesis protein CapD	1823	2	0.001097093
Cthe_2558	-	NAD-dependent epimerase/dehydratase	917	1	0.001090513
Cthe_2071	-	hypothetical protein	923	1	0.001083424
Cthe_3085	-	response regulator receiver protein	926	1	0.001079914
Cthe_2550	-	glycosyltransferase sugar-binding region containing DXD motif	983	1	0.001017294
Cthe_0352	+	hypothetical protein	992	1	0.001008065
Cthe_0211	+	glycoside hydrolase, family 16(EC:3.2.1.73)	1004	1	0.000996016
Cthe_1430	-	hypothetical protein	1013	1	0.000987167
Cthe_2454	-	Fibronectin, type III	3128	3	0.000959079
Cthe_2485	+	hypothetical protein	1046	1	0.000956023
Cthe_3208	-	Integrase, catalytic region	1070	1	0.000934579
Cthe_1417	-	hypothetical protein	1082	1	0.000924214
Cthe_1588	-	extracellular solute-binding protein, family 1	1106	1	0.000904159

المنارة للاستشارات

www.manaraa.com

Cthe_2858	-	transposase, mutator type	1136	1	0.000880282
Cthe_1194	-	hypothetical protein	1148	1	0.000871080
Cthe_3209	+	hypothetical protein	1199	1	0.000834028
Cthe_2672	-	transposase, mutator type	1220	1	0.000819672
Cthe_2850	+	transposase IS116/IS110/IS902	1256	1	0.000796178
Cthe_0008	+	hypothetical protein	302	0	0.000000000
Cthe_0011	+	hypothetical protein	275	0	0.000000000
Cthe_0012	+	YD repeat protein	191	0	0.000000000
Cthe_0013	+	hypothetical protein	356	0	0.000000000
Cthe_0014	+	hypothetical protein	296	0	0.000000000
Cthe_0027	+	hypothetical protein	371	0	0.000000000
Cthe_0031	+	putative membrane transporter	623	0	0.000000000
Cthe_0037	-	heat shock protein Hsp20	428	0	0.000000000
Cthe_0041	-	hypothetical protein	434	0	0.000000000
Cthe_0074	+	RNA polymerase, sigma-24 subunit, ECF subfamily	191	0	0.000000000
Cthe_0078	-	hypothetical protein	206	0	0.000000000
Cthe_0079	+	hypothetical protein	515	0	0.000000000
Cthe_0103	+	protein of unknown function UPF0047	392	0	0.000000000
Cthe_0109	+	cellulosome enzyme, dockerin type I	317	0	0.000000000
Cthe_0117	+	HPrNtr domain containing protein	272	0	0.000000000
Cthe_0118	+	anti-sigma-factor antagonist	335	0	0.000000000
Cthe_0119	+	putative anti-sigma regulatory factor, serine/threonine protein kinase	449	0	0.000000000
Cthe_0121	+	hypothetical protein	257	0	0.000000000
Cthe_0122	+	SpoVA protein	443	0	0.000000000
Cthe_0124	+	SpoVA protein	356	0	0.000000000
Cthe_0128	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_0141	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_0144	+	preprotein translocase, SecG subunit	251	0	0.000000000
Cthe_0148	+	transposase, mutator type	1223	0	0.000000000
Cthe_0149	-	HPrNtr domain containing protein	257	0	0.000000000
Cthe_0159	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_0165	-	hypothetical protein	557	0	0.000000000
Cthe_0168	+	hypothetical protein	470	0	0.000000000
Cthe_0170	-	hypothetical protein	473	0	0.000000000
Cthe_0172	-	hypothetical protein	242	0	0.000000000
Cthe_0177	+	hypothetical protein	197	0	0.000000000
Cthe_0192	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_0194	+	hypothetical protein	383	0	0.000000000
Cthe_0195	+	RNA polymerase, sigma-24 subunit, ECF subfamily	293	0	0.000000000
Cthe_0199	+	4Fe-4S ferredoxin, iron-sulfur binding	455	0	0.000000000
Cthe_0209	+	glycosyltransferase	308	0	0.000000000
Cthe_0216	-	small acid-soluble spore protein, alpha/beta type	287	0	0.000000000
Cthe_0220	-	transposase, IS4	947	0	0.000000000
Cthe_0221	+	hypothetical protein	344	0	0.000000000

Cthe_0222	+	hypothetical protein	248	0	0.000000000
Cthe_0223	+	hypothetical protein	437	0	0.000000000
Cthe_0224	+		203	0	0.000000000
Cthe_0225	+		362	0	0.000000000
Cthe_0245	-	transcriptional regulator, ArsR family	377	0	0.000000000
Cthe_0248	-	transposase, IS4	947	0	0.000000000
Cthe_0255	+	hypothetical protein	413	0	0.000000000
Cthe_0263	+	hypothetical protein	371	0	0.000000000
Cthe_0279	+	hypothetical protein	191	0	0.000000000
Cthe_0292	+	transposase, mutator type	1220	0	0.000000000
Cthe_0302	+	hypothetical protein	440	0	0.000000000
Cthe_0303	-	hypothetical protein	446	0	0.000000000
Cthe_0304	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_0308	+	hypothetical protein	338	0	0.000000000
Cthe_0323	+	hypothetical protein	242	0	0.000000000
Cthe_0326	-	hypothetical protein	203	0	0.000000000
Cthe_0348	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_0356	+	transposase, mutator type	1223	0	0.000000000
Cthe_0359	+	hypothetical protein	401	0	0.000000000
Cthe_0364	+	transcriptional regulator, XRE family	695	0	0.000000000
Cthe_0371	-	transposase, mutator type	1223	0	0.000000000
Cthe_0380	+	hypothetical protein	239	0	0.000000000
Cthe_0385	-	hypothetical protein	257	0	0.000000000
Cthe_0387	+	transposase, IS4	947	0	0.000000000
Cthe_0409	+	ACT domain-containing protein	269	0	0.000000000
Cthe_0415	-	spore coat protein CotJB	245	0	0.000000000
Cthe_0416	-	hypothetical protein	257	0	0.000000000
Cthe_0417	+	ribosomal protein S15	263	0	0.000000000
Cthe_0421	+	flavoprotein	584	0	0.000000000
Cthe_0438	+	cellulosome enzyme, dockerin type I	386	0	0.000000000
Cthe_0447	+	RNA polymerase, sigma 28 subunit	734	0	0.000000000
Cthe_0449	+	hypothetical protein	260	0	0.000000000
Cthe_0450	+	Ribonucleotide reductase regulator NrdR-like protein	455	0	0.000000000
Cthe_0453	+	hypothetical protein	440	0	0.000000000
Cthe_0454	-	hypothetical protein	404	0	0.000000000
Cthe_0456	+	protein of unknown function UPF0047	416	0	0.000000000
Cthe_0475	+	flagellar protein	290	0	0.000000000
Cthe_0482	+	flagellar biosynthetic protein FliQ	269	0	0.000000000
Cthe_0496	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_0497	+		125	0	0.000000000
Cthe_0510	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_0513	-	DNA binding domain, excisionase family	209	0	0.000000000
Cthe_0516	-	IS66 Orf2 like	356	0	0.000000000
Cthe_0517	-	hypothetical protein	368	0	0.000000000
Cthe_0520	-	hypothetical protein	725	0	0.000000000
Cthe_0524	+	hypothetical protein	473	0	0.000000000

www.manaraa.com

Cthe_0526	+	hypothetical protein	443	0	0.000000000
Cthe_0527	-	hypothetical protein	245	0	0.000000000
Cthe_0528	+	hypothetical protein	395	0	0.000000000
Cthe_0529	+		158	0	0.000000000
Cthe_0530	+	hypothetical protein	614	0	0.000000000
Cthe_0535	+	hypothetical protein	272	0	0.000000000
Cthe_0584	+	protein of unknown function DUF503	290	0	0.000000000
Cthe_0585	-	Integrase, catalytic region	836	0	0.000000000
Cthe_0586	-	transposase IS3/IS911	293	0	0.000000000
Cthe_0587	-	transposase, mutator type	1220	0	0.000000000
Cthe_0588	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_0591	+	Integrase, catalytic region	836	0	0.000000000
Cthe_0594	+	transposase, mutator type	1220	0	0.000000000
Cthe_0604	+	hypothetical protein	257	0	0.000000000
Cthe_0619	-	FeoA	221	0	0.000000000
Cthe_0636	-	hypothetical protein	248	0	0.000000000
Cthe_0637	-	protein of unknown function DUF909	302	0	0.000000000
Cthe_0641	-	Coat F	242	0	0.000000000
Cthe_0642	-	hypothetical protein	185	0	0.000000000
Cthe_0649	+	hypothetical protein	329	0	0.000000000
Cthe_0650	+	hypothetical protein	299	0	0.000000000
Cthe_0652	+	4-oxalocrotonate tautomerase	236	0	0.000000000
Cthe_0658	+	hypothetical protein	296	0	0.000000000
Cthe_0659	-	hypothetical protein	389	0	0.000000000
Cthe_0662	+	hypothetical protein	299	0	0.000000000
Cthe_0663	-	hypothetical protein	569	0	0.000000000
Cthe_0675	+	hypothetical protein	554	0	0.000000000
Cthe_0682	-	hypothetical protein	977	0	0.000000000
Cthe_0692	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_0698	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_0703	+	hypothetical protein	242	0	0.000000000
Cthe_0715	+	Adenosylmethionine decarboxylase(EC:4.1.1.50)	374	0	0.000000000
Cthe_0718	+	hypothetical protein	287	0	0.000000000
Cthe_0719	+	transcriptional regulator, BadM/Rrf2 family	431	0	0.000000000
Cthe_0738	-	copper ion binding protein	212	0	0.000000000
Cthe_0751	-	transcriptional regulator, XRE family with cupin sensor	539	0	0.000000000
Cthe_0752	-	hypothetical protein	239	0	0.000000000
Cthe_0767	-	16S rRNA processing protein RimM	524	0	0.000000000
Cthe_0772	-	Peptidoglycan-binding LysM	287	0	0.000000000
Cthe_0782	-	putative methyl-accepting chemotaxis sensory transducer	521	0	0.000000000
Cthe_0785	+	hypothetical protein	302	0	0.000000000
Cthe_0790	-	conserved hythetical protein	284	0	0.000000000
Cthe_0796	-	hypothetical protein	266	0	0.000000000
Cthe_0802	-	hypothetical protein	287	0	0.000000000
Cthe_0805	-	response regulator receiver protein	410	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_0809	-	CheW protein	491	0	0.000000000
Cthe_0811	-	response regulator receiver protein	353	0	0.000000000
Cthe_0822	-	IstB-like ATP-binding protein	725	0	0.000000000
Cthe_0823	-	hypothetical protein	1514	0	0.000000000
Cthe_0842	-	stage III sporulation protein AD	389	0	0.000000000
Cthe_0843	-	Stage III sporulation AC	194	0	0.000000000
Cthe_0845	-	stage III sporulation protein spoIIIAA	1010	0	0.000000000
Cthe_0862	+	hypothetical protein	320	0	0.000000000
Cthe_0879	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_0883	+	flagellar protein	419	0	0.000000000
Cthe_0887	-	hypothetical protein	212	0	0.000000000
Cthe_0894	-	hypothetical protein	218	0	0.000000000
Cthe_0899	+	hypothetical protein	272	0	0.000000000
Cthe_0913	+	transposase IS116/IS110/IS902	1220	0	0.000000000
Cthe_0919	-	Integrase, catalytic region	836	0	0.000000000
Cthe_0920	-	transposase IS3/IS911	293	0	0.000000000
Cthe_0921	+	hypothetical protein	188	0	0.000000000
Cthe_0940	-	hypothetical protein	440	0	0.000000000
Cthe_0943	-	hypothetical protein	251	0	0.000000000
Cthe_0955	-	hypothetical protein	410	0	0.000000000
Cthe_0966	+	hypothetical protein	434	0	0.000000000
Cthe_0967	-	transposase, IS4	947	0	0.000000000
Cthe_0969	+		383	0	0.000000000
Cthe_0970	+	transposase, mutator type	1223	0	0.000000000
Cthe_0971	+		239	0	0.000000000
Cthe_0973	-	UDP-N-acetylglucosamine 1-carboxyvinyltransferase(EC:2.5.1.7)	374	0	0.000000000
Cthe_1002	-	hypothetical protein	200	0	0.000000000
Cthe_1015	-	hypothetical protein	422	0	0.000000000
Cthe_1017	-	hypothetical protein	302	0	0.000000000
Cthe_1033	-	glutamyl-tRNA(Gln) amidotransferase, C subunit	287	0	0.000000000
Cthe_1043	+	hypothetical protein	245	0	0.000000000
Cthe_1044	+	hypothetical protein	329	0	0.000000000
Cthe_1045	+	hypothetical protein	326	0	0.000000000
Cthe_1057	-	phage shock protein C, PspC	194	0	0.000000000
Cthe_1078	+	hypothetical protein	317	0	0.000000000
Cthe_1080	+	carbohydrate binding family 25	323	0	0.000000000
Cthe_1082	+	hypothetical protein	257	0	0.000000000
Cthe_1087	-	hypothetical protein	263	0	0.000000000
Cthe_1096	-	hypothetical protein	224	0	0.000000000
Cthe_1108	-	hypothetical protein	218	0	0.000000000
Cthe_1121	-	hypothetical protein	281	0	0.000000000
Cthe_1122	-	hypothetical protein	197	0	0.000000000
Cthe_1125	-	hypothetical protein	218	0	0.000000000
Cthe_1126	-	hypothetical protein	473	0	0.000000000
Cthe_1127	+	phage integrase	905	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_1129	+	hypothetical protein	335	0	0.000000000
Cthe_1130	+	hypothetical protein	305	0	0.000000000
Cthe_1131	-	hypothetical protein	287	0	0.000000000
Cthe_1146	-	hypothetical protein	362	0	0.000000000
Cthe_1147	-	hypothetical protein	239	0	0.000000000
Cthe_1149	-	hypothetical protein	344	0	0.000000000
Cthe_1152	+	hypothetical protein	638	0	0.000000000
Cthe_1154	+	transcriptional regulator, XRE family	254	0	0.000000000
Cthe_1155	-	hypothetical protein	290	0	0.000000000
Cthe_1157	+	hypothetical protein	287	0	0.000000000
Cthe_1158	-	hypothetical protein	290	0	0.000000000
Cthe_1160	+	S23 ribosomal protein	416	0	0.000000000
Cthe_1179	-	hypothetical protein	194	0	0.000000000
Cthe_1180	-	uncharacterized predicted metal-binding protein	386	0	0.000000000
Cthe_1190	+	hypothetical protein	491	0	0.000000000
Cthe_1191	-	HAD-superfamily hydrolase, subfamily IA, variant 1	644	0	0.000000000
Cthe_1193	+	transposase, mutator type	1223	0	0.000000000
Cthe_1195	-	hypothetical protein	344	0	0.000000000
Cthe_1206	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_1213	+	hypothetical protein	272	0	0.000000000
Cthe_1215	-	permease	371	0	0.000000000
Cthe_1224	-	ribosomal protein L35	197	0	0.000000000
Cthe_1226	-	dihydrofolate reductase region	503	0	0.000000000
Cthe_1233	+	hypothetical protein	539	0	0.000000000
Cthe_1242	-	Integrase, catalytic region	1112	0	0.000000000
Cthe_1254	+	hypothetical protein	368	0	0.000000000
Cthe_1263	-	tryptophan RNA-binding attenuator protein	233	0	0.000000000
Cthe_1267	-	two component transcriptional regulator, LuxR family	650	0	0.000000000
Cthe_1280	-	hypothetical protein	266	0	0.000000000
Cthe_1290	+	hypothetical protein	341	0	0.000000000
Cthe_1298	-	ribosomal protein L25	317	0	0.000000000
Cthe_1299	-	hypothetical protein	314	0	0.000000000
Cthe_1306	+	hypothetical protein	401	0	0.000000000
Cthe_1318	-	response regulator receiver protein	371	0	0.000000000
Cthe_1327	-	hypothetical protein	281	0	0.000000000
Cthe_1366	-	dTDP-4-dehydrorhamnose 3,5-epimerase(EC:5.1.3.13)	551	0	0.000000000
Cthe_1369	-	histidine triad (HIT) protein	344	0	0.000000000
Cthe_1377	-	amino acid-binding ACT	443	0	0.000000000
Cthe_1380	-	response regulator receiver protein	374	0	0.000000000
Cthe_1387	-	hypothetical protein	395	0	0.000000000
Cthe_1388	-	TrpR like protein, YerC/YecD	293	0	0.000000000
Cthe_1392	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_1395	-		191	0	0.000000000
Cthe_1401	+	hypothetical protein	332	0	0.000000000
Cthe_1403	+	FeoA	239	0	0.000000000

المنارة للاستشارات
www.manaraa.com

Cthe_1407	-	hypothetical protein	686	0	0.000000000
Cthe_1416	-	hypothetical protein	329	0	0.000000000
Cthe_1419	-	hypothetical protein	422	0	0.000000000
Cthe_1420	-	protein of unknown function DUF204	596	0	0.000000000
Cthe_1427	+	branched-chain amino acid transport	302	0	0.000000000
Cthe_1436	-		374	0	0.000000000
Cthe_1439	-	hypothetical protein	344	0	0.000000000
Cthe_1440	-	hypothetical protein	203	0	0.000000000
Cthe_1445	-	uncharacterized Zn-finger protein	245	0	0.000000000
Cthe_1447	-	hypothetical protein	536	0	0.000000000
Cthe_1451	-	GCN5-related N-acetyltransferase	473	0	0.000000000
Cthe_1454	-	hypothetical protein	425	0	0.000000000
Cthe_1460	-	hypothetical protein	359	0	0.000000000
Cthe_1462	-	transcriptional regulator, XRE family	314	0	0.000000000
Cthe_1464	-	Endoribonuclease L-PSP	281	0	0.000000000
Cthe_1466	-	protein of unknown function DUF1706	545	0	0.000000000
Cthe_1467	-	hypothetical protein	227	0	0.000000000
Cthe_1468	+	hypothetical protein	260	0	0.000000000
Cthe_1469	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_1476	-	hypothetical protein	203	0	0.000000000
Cthe_1489	-	hypothetical protein	266	0	0.000000000
Cthe_1491	-	transcriptional regulator, PadR-like family	317	0	0.000000000
Cthe_1494	-	protein of unknown function DUF6, transmembrane	413	0	0.000000000
Cthe_1496	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_1508	-	hypothetical protein	356	0	0.000000000
Cthe_1526	-	ABC transporter related protein	764	0	0.000000000
Cthe_1531	-	hypothetical protein	488	0	0.000000000
Cthe_1532	-	flagellin-like protein	386	0	0.000000000
Cthe_1537	-	hypothetical protein	206	0	0.000000000
Cthe_1542	-	glutamyl-tRNA(Gln) amidotransferase, C subunit	293	0	0.000000000
Cthe_1544	-	hypothetical protein	200	0	0.000000000
Cthe_1545	-	hypothetical protein	245	0	0.000000000
Cthe_1547	+	hypothetical protein	434	0	0.000000000
Cthe_1550	+	transposase IS116/IS110/IS902	1289	0	0.000000000
Cthe_1589	+	transcriptional repressor, CopY family	362	0	0.000000000
Cthe_1594	-	hypothetical protein	317	0	0.000000000
Cthe_1597	+	flavodoxin	443	0	0.000000000
Cthe_1608	-	Recombinase	455	0	0.000000000
Cthe_1610	-	hypothetical protein	218	0	0.000000000
Cthe_1615	-	hypothetical protein	194	0	0.000000000
Cthe_1619	+	transcriptional regulator, AbrB family	272	0	0.000000000
Cthe_1620	+	Protein of unknown function DUF132	404	0	0.000000000
Cthe_1621	-	hypothetical protein	191	0	0.000000000
Cthe_1622	-	hypothetical protein	383	0	0.000000000
Cthe_1624	-	hypothetical protein	344	0	0.000000000

المنارة للاستشارات
www.manaraa.com

Cthe_1625	-	phage protein, HK97 gp10 family	410	0	0.000000000
Cthe_1626	-	phage head-tail adaptor, putative	335	0	0.000000000
Cthe_1627	-	uncharacterized phage protein	308	0	0.000000000
Cthe_1633	-	hypothetical protein	194	0	0.000000000
Cthe_1634	-	hypothetical protein	467	0	0.000000000
Cthe_1636	-	hypothetical protein	263	0	0.000000000
Cthe_1642	-	hypothetical protein	242	0	0.000000000
Cthe_1644	-	hypothetical protein	434	0	0.000000000
Cthe_1646	-	hypothetical protein	302	0	0.000000000
Cthe_1655	+	hypothetical protein	206	0	0.000000000
Cthe_1656	-	Integrase, catalytic region	836	0	0.000000000
Cthe_1657	-	transposase IS3/IS911	293	0	0.000000000
Cthe_1660	-	transposase, IS4	947	0	0.000000000
Cthe_1661	-	Integrase, catalytic region	836	0	0.000000000
Cthe_1662	-	transposase IS3/IS911	293	0	0.000000000
Cthe_1666	+	transposase, IS4	947	0	0.000000000
Cthe_1670	+	Recombinase	431	0	0.000000000
Cthe_1673	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1674	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1675	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1676	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1682	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1683	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1686	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1687	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1691	-	Integrase, catalytic region	836	0	0.000000000
Cthe_1692	-	transposase IS3/IS911	293	0	0.000000000
Cthe_1697	-	transcriptional regulator, ArsR family	260	0	0.000000000
Cthe_1699	-		134	0	0.000000000
Cthe_1704	-	toxin secretion/phage lysis holin	410	0	0.000000000
Cthe_1710	-	hypothetical protein	323	0	0.000000000
Cthe_1711	-	hypothetical protein	452	0	0.000000000
Cthe_1712	-	Integrase, catalytic region	1112	0	0.000000000
Cthe_1713	-	phage major tail protein, phi13 family	572	0	0.000000000
Cthe_1714	-	hypothetical protein	329	0	0.000000000
Cthe_1715	-	phage protein, HK97 gp10 family	392	0	0.000000000
Cthe_1716	-	phage head-tail adaptor, putative	326	0	0.000000000
Cthe_1718	-	hypothetical protein	440	0	0.000000000
Cthe_1730	-	hypothetical protein	425	0	0.000000000
Cthe_1731	-	HNH endonuclease	380	0	0.000000000
Cthe_1732	-	hypothetical protein	404	0	0.000000000
Cthe_1733	-	hypothetical protein	302	0	0.000000000
Cthe_1736	-	hypothetical protein	458	0	0.000000000
Cthe_1738	-	hypothetical protein	365	0	0.000000000
Cthe_1740	-	BRO-like protein	764	0	0.000000000
Cthe_1741	-	hypothetical protein	251	0	0.000000000
Cthe_1742	-	Helix-turn-helix, type 11 containing protein	323	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_1744	-	hypothetical protein	368	0	0.000000000
Cthe_1747	-	DNA mismatch endonuclease vsr	416	0	0.000000000
Cthe_1750	-	hypothetical protein	209	0	0.000000000
Cthe_1769	+	hypothetical protein	392	0	0.000000000
Cthe_1779	-	hypothetical protein	392	0	0.000000000
Cthe_1784	+	hypothetical protein	395	0	0.000000000
Cthe_1808	+	transposase IS116/IS110/IS902	1220	0	0.000000000
Cthe_1811	-	transcriptional repressor, CopY family	374	0	0.000000000
Cthe_1817	-	urease, beta subunit(EC:3.5.1.5)	368	0	0.000000000
Cthe_1818	-	urease, gamma subunit(EC:3.5.1.5)	302	0	0.000000000
Cthe_1843	-	transcriptional regulator, BadM/Rrf2 family	449	0	0.000000000
Cthe_1856	-	small acid-soluble spore protein beta	173	0	0.000000000
Cthe_1874	-	transposase, mutator type	1223	0	0.000000000
Cthe_1879	-	hypothetical protein	374	0	0.000000000
Cthe_1880	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1881	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1882	-	hypothetical protein	461	0	0.000000000
Cthe_1884	-	transposase IS116/IS110/IS902	1289	0	0.000000000
Cthe_1886	-	Integrase, catalytic region	836	0	0.000000000
Cthe_1887	-	transposase IS3/IS911	293	0	0.000000000
Cthe_1889	+	transposase, mutator type	1220	0	0.000000000
Cthe_1892	-		225	0	0.000000000
Cthe_1921	+	transcriptional regulator, PadR-like family	344	0	0.000000000
Cthe_1941	+	putative regulatory protein, FmdB family	248	0	0.000000000
Cthe_1949	-	hypothetical protein	281	0	0.000000000
Cthe_1950	+	hypothetical protein	323	0	0.000000000
Cthe_1952	+	hypothetical protein	374	0	0.000000000
Cthe_1971	-	hypothetical protein	329	0	0.000000000
Cthe_1972	-	hypothetical protein	320	0	0.000000000
Cthe_1975	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1976	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1977	-	hypothetical protein	326	0	0.000000000
Cthe_1978	-	hypothetical protein	245	0	0.000000000
Cthe_1981	-	hypothetical protein	389	0	0.000000000
Cthe_1982	-	hypothetical protein	242	0	0.000000000
Cthe_1984	-	hypothetical protein	347	0	0.000000000
Cthe_1987	-	hypothetical protein	392	0	0.000000000
Cthe_1988	-	hypothetical protein	278	0	0.000000000
Cthe_1990	-	hypothetical protein	347	0	0.000000000
Cthe_1992	+	transposase IS3/IS911	293	0	0.000000000
Cthe_1993	+	Integrase, catalytic region	836	0	0.000000000
Cthe_1995	+	hypothetical protein	1514	0	0.000000000
Cthe_1996	+	IstB-like ATP-binding protein	725	0	0.000000000
Cthe_1997	-	hypothetical protein	314	0	0.000000000
Cthe_1999	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_2000	-	Integrase, catalytic region	836	0	0.000000000
Cthe_2001	-	transposase IS3/IS911	293	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_2002	-	hypothetical protein	476	0	0.000000000
Cthe_2003	-		593	0	0.000000000
Cthe_2004	+	transposase	1484	0	0.000000000
Cthe_2005	+	IstB-like ATP-binding protein	758	0	0.000000000
Cthe_2008	-	conserved hypothetical protein, CF-22 family	518	0	0.000000000
Cthe_2012	-	Integrase, catalytic region	836	0	0.000000000
Cthe_2013	-	transposase IS3/IS911	293	0	0.000000000
Cthe_2017	+	transposase, mutator type	1220	0	0.000000000
Cthe_2020	-	hypothetical protein	791	0	0.000000000
Cthe_2021	-	hypothetical protein	392	0	0.000000000
Cthe_2022	-	hypothetical protein	278	0	0.000000000
Cthe_2023	-		335	0	0.000000000
Cthe_2026	-	hypothetical protein	350	0	0.000000000
Cthe_2027	-	hypothetical protein	239	0	0.000000000
Cthe_2028	-	hypothetical protein	263	0	0.000000000
Cthe_2033	-	hypothetical protein	314	0	0.000000000
Cthe_2035	-		520	0	0.000000000
Cthe_2041	-	hypothetical protein	449	0	0.000000000
Cthe_2054	-	protein of unknown function DUF324	545	0	0.000000000
Cthe_2067	+	hypothetical protein	407	0	0.000000000
Cthe_2074	+	protein of unknown function DUF378	218	0	0.000000000
Cthe_2080	-	NUDIX hydrolase	572	0	0.000000000
Cthe_2082	-	hypothetical protein	239	0	0.000000000
Cthe_2084	-	hypothetical protein	293	0	0.000000000
Cthe_2100	+	transcriptional regulator, AbrB family	239	0	0.000000000
Cthe_2103	-	4Fe-4S ferredoxin, iron-sulfur binding	182	0	0.000000000
Cthe_2114	+	Integrase, catalytic region	1070	0	0.000000000
Cthe_2115	+	protein of unknown function DUF77	287	0	0.000000000
Cthe_2122	+	hypothetical protein	398	0	0.000000000
Cthe_2129	-		200	0	0.000000000
Cthe_2134	-		533	0	0.000000000
Cthe_2135	-	Integrase, catalytic region	836	0	0.000000000
Cthe_2136	-	transposase IS3/IS911	293	0	0.000000000
Cthe_2146	+	copper amine oxidase-like protein	341	0	0.000000000
Cthe_2153	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_2156	-	hypothetical protein	278	0	0.000000000
Cthe_2165	-	protein of unknown function DUF1256	593	0	0.000000000
Cthe_2173	-	Integrase, catalytic region	1112	0	0.000000000
Cthe_2177	+	hypothetical protein	197	0	0.000000000
Cthe_2178	+	hypothetical protein	461	0	0.000000000
Cthe_2187	-	ribosomal protein S6	287	0	0.000000000
Cthe_2188	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_2198	+	transposase	1484	0	0.000000000
Cthe_2199	+	IstB-like ATP-binding protein	758	0	0.000000000
Cthe_2201	-	transposase, mutator type	1220	0	0.000000000
Cthe_2206	-	hypothetical protein	296	0	0.000000000
Cthe_2225	-	Glyoxalase/bleomycin resistance	410	0	0.000000000

المنارة للاستشارات
www.manaraa.com

		protein/dioxygenase			
Cthe_2227	-	phosphopantetheine-binding	215	0	0.000000000
Cthe_2232	-	polysaccharide biosynthesis protein CapD	998	0	0.000000000
Cthe_2234	-	hypothetical protein	383	0	0.000000000
Cthe_2239	-	carbon storage regulator, CsrA	227	0	0.000000000
Cthe_2240	-	protein of unknown function DUF180	479	0	0.000000000
Cthe_2246	-	anti-sigma-28 factor, FlgM	293	0	0.000000000
Cthe_2247	-	regulatory protein, MerR	443	0	0.000000000
Cthe_2250	-	hypothetical protein	404	0	0.000000000
Cthe_2273	+	hypothetical protein	1121	0	0.000000000
Cthe_2285	-	CheW protein	425	0	0.000000000
Cthe_2291	-	hypothetical protein	290	0	0.000000000
Cthe_2292	-	hypothetical protein	461	0	0.000000000
Cthe_2293	-	hypothetical protein	338	0	0.000000000
Cthe_2294	+	transposase	1484	0	0.000000000
Cthe_2295	+	IstB-like ATP-binding protein	758	0	0.000000000
Cthe_2296	-	CRISPR-associated protein Cas2	263	0	0.000000000
Cthe_2307	+	hemerythrin-like metal-binding protein	422	0	0.000000000
Cthe_2322	-	DNA recombinase	332	0	0.000000000
Cthe_2323	-	transposase, IS4	947	0	0.000000000
Cthe_2324	-	Resolvase-like protein	287	0	0.000000000
Cthe_2349	-	small, acid-soluble spore proteins, alpha/be	170	0	0.000000000
Cthe_2350	+	cell wall hydrolase, SleB	443	0	0.000000000
Cthe_2351	+	hypothetical protein	449	0	0.000000000
Cthe_2359	+	metal dependent phosphohydrolase	617	0	0.000000000
Cthe_2368	-	protein of unknown function DUF37	215	0	0.000000000
Cthe_2370	-	ribosomal protein L34	134	0	0.000000000
Cthe_2373	+	RNA-binding S4	206	0	0.000000000
Cthe_2388	+	hypothetical protein	323	0	0.000000000
Cthe_2389	+	sigmaK-factor processing regulatory BofA	278	0	0.000000000
Cthe_2396	+	hypothetical protein	275	0	0.000000000
Cthe_2405	-	Heavy metal transport/detoxification protein	206	0	0.000000000
Cthe_2421	+	hypothetical protein	248	0	0.000000000
Cthe_2436	-	hypothetical protein	236	0	0.000000000
Cthe_2438	+	hypothetical protein	182	0	0.000000000
Cthe_2453	-	hypothetical protein	356	0	0.000000000
Cthe_2456	-	nuclease	422	0	0.000000000
Cthe_2457	+	hypothetical protein	206	0	0.000000000
Cthe_2459	-	protein of unknown function DUF955	440	0	0.000000000
Cthe_2460	-	transcriptional regulator, XRE family	428	0	0.000000000
Cthe_2461	+	transcriptional regulator, XRE family	239	0	0.000000000
Cthe_2463	+	DNA binding domain, excisionase family	224	0	0.000000000
Cthe_2468	+	hypothetical protein	374	0	0.000000000
Cthe_2477	+	hypothetical protein	206	0	0.000000000
Cthe_2481	+	hypothetical protein	335	0	0.000000000
Cthe_2482	+	hypothetical protein	410	0	0.000000000
Cthe_2483	+	hypothetical protein	419	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_2484	-	transposase IS116/IS110/IS902	1289	0	0.000000000
Cthe_2487	+	hypothetical protein	413	0	0.000000000
Cthe_2491	+	hypothetical protein	395	0	0.000000000
Cthe_2496	+	hypothetical protein	320	0	0.000000000
Cthe_2497	+	hypothetical protein	290	0	0.000000000
Cthe_2501	+	hypothetical protein	431	0	0.000000000
Cthe_2512	+	protein of unknown function DUF951	200	0	0.000000000
Cthe_2523	+	hypothetical protein	380	0	0.000000000
Cthe_2526	+	precorrin-2 oxidase / ferrochelatase	428	0	0.000000000
Cthe_2538	+	thiamine biosynthesis protein ThiS	209	0	0.000000000
Cthe_2542	+	SirA-like protein	245	0	0.000000000
Cthe_2545	-	signal transduction histidine kinase regulating citrate/malate metabolism	1085	0	0.000000000
Cthe_2547	+	hypothetical protein	326	0	0.000000000
Cthe_2567	-	hypothetical protein	194	0	0.000000000
Cthe_2569	-	hypothetical protein	359	0	0.000000000
Cthe_2571	-		254	0	0.000000000
Cthe_2592	+	hypothetical protein	467	0	0.000000000
Cthe_2618	+	stage III sporulation protein D, SpoIIID	263	0	0.000000000
Cthe_2621	+	protein of unknown function DUF1078-like protein	800	0	0.000000000
Cthe_2628	+	SpoVG	284	0	0.000000000
Cthe_2631	+	Aminoacyl-tRNA hydrolase(EC:3.1.1.29)	590	0	0.000000000
Cthe_2634	+	transcriptional regulator, AraC family	614	0	0.000000000
Cthe_2635	+	pyridoxamine 5'-phosphate oxidase-related, FMN-binding	434	0	0.000000000
Cthe_2649	-	HpcH/HpaI aldolase	827	0	0.000000000
Cthe_2651	+	hypothetical protein	332	0	0.000000000
Cthe_2657	+	histone-like DNA-binding protein	275	0	0.000000000
Cthe_2658	+	RNA-binding S4	239	0	0.000000000
Cthe_2659	+	YabP-like protein	290	0	0.000000000
Cthe_2660	+	hypothetical protein	509	0	0.000000000
Cthe_2661	+	Septum formation initiator	281	0	0.000000000
Cthe_2662	+	RNA binding S1	422	0	0.000000000
Cthe_2668	+	hypothetical protein	371	0	0.000000000
Cthe_2670	+	hypothetical protein	734	0	0.000000000
Cthe_2673	+	hypothetical protein	284	0	0.000000000
Cthe_2674	+	hypothetical protein	422	0	0.000000000
Cthe_2682	+	hypothetical protein	341	0	0.000000000
Cthe_2698	+	putative transcriptional regulator, CopG family	176	0	0.000000000
Cthe_2715	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_2716	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_2733	+	hypothetical protein	374	0	0.000000000
Cthe_2735	+	HPrNtr domain containing protein	269	0	0.000000000
Cthe_2745	-	hypothetical protein	323	0	0.000000000
Cthe_2751	+	hypothetical protein	404	0	0.000000000
Cthe_2753	+	hypothetical protein	212	0	0.000000000
Cthe_2756	-	transposase	260	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_2758	+	hypothetical protein	128	0	0.000000000
Cthe_2766	+	hypothetical protein	245	0	0.000000000
Cthe_2770	+	transposase IS116/IS110/IS902	1289	0	0.000000000
Cthe_2772	+	transposase IS3/IS911	293	0	0.000000000
Cthe_2773	+	Integrase, catalytic region	836	0	0.000000000
Cthe_2774	-	hypothetical protein	179	0	0.000000000
Cthe_2783	+	hypothetical protein	161	0	0.000000000
Cthe_2794	+	pyruvate/ketoisovalerate oxidoreductase, gamma subunit	545	0	0.000000000
Cthe_2795	+	pyruvate ferredoxin/flavodoxin oxidoreductase, delta subunit	311	0	0.000000000
Cthe_2816	-	transposase, mutator type	1220	0	0.000000000
Cthe_2823	+	hypothetical protein	296	0	0.000000000
Cthe_2827	+	hypothetical protein	311	0	0.000000000
Cthe_2829	-	ISDet4, transposase	443	0	0.000000000
Cthe_2830	-	IstB-like ATP-binding protein	758	0	0.000000000
Cthe_2831	-	transposase	1484	0	0.000000000
Cthe_2833	+	hypothetical protein	347	0	0.000000000
Cthe_2836	+	hypothetical protein	389	0	0.000000000
Cthe_2840	+	hypothetical protein	344	0	0.000000000
Cthe_2842	+	hypothetical protein	344	0	0.000000000
Cthe_2843	+	hypothetical protein	308	0	0.000000000
Cthe_2845	+	hypothetical protein	296	0	0.000000000
Cthe_2849	+	hypothetical protein	302	0	0.000000000
Cthe_2851	+	hypothetical protein	287	0	0.000000000
Cthe_2852	+	hypothetical protein	296	0	0.000000000
Cthe_2859	+	hypothetical protein	347	0	0.000000000
Cthe_2861	+	hypothetical protein	242	0	0.000000000
Cthe_2864	+	hypothetical protein	389	0	0.000000000
Cthe_2867	+	hypothetical protein	302	0	0.000000000
Cthe_2869	+	hypothetical protein	293	0	0.000000000
Cthe_2946	-	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase(EC:4.6.1.12)	470	0	0.000000000
Cthe_2953	-	hypothetical protein	335	0	0.000000000
Cthe_2955	-	hypothetical protein	239	0	0.000000000
Cthe_2956	+	hypothetical protein	470	0	0.000000000
Cthe_2958	+	transposase, mutator type	1220	0	0.000000000
Cthe_2960	+	hypothetical protein	332	0	0.000000000
Cthe_2969	+	transcriptional regulator, GntR family	374	0	0.000000000
Cthe_2971	+	hypothetical protein	611	0	0.000000000
Cthe_2976	+	hypothetical protein	239	0	0.000000000
Cthe_2979	+	4Fe-4S ferredoxin, iron-sulfur binding	197	0	0.000000000
Cthe_2984	-		624	0	0.000000000
Cthe_2987	-	hypothetical protein	392	0	0.000000000
Cthe_2999	-	hypothetical protein	572	0	0.000000000
Cthe_3009	-	hypothetical protein	308	0	0.000000000
Cthe_3010	+	hypothetical protein	542	0	0.000000000
Cthe_3015	-	hydrogenase assembly chaperone hypC/hupF	227	0	0.000000000

المنارة للاستشارات
www.manaraa.com

Cthe_3018	-	hydrogenase expression/synthesis, HypA	338	0	0.000000000
Cthe_3021	-	Ech hydrogenase, subunit EchD, putative	359	0	0.000000000
Cthe_3025	+	hypothetical protein	290	0	0.000000000
Cthe_3029	+	CheW protein	491	0	0.000000000
Cthe_3031	+	hypothetical protein	272	0	0.000000000
Cthe_3032	+	hypothetical protein	434	0	0.000000000
Cthe_3033	+	hypothetical protein	419	0	0.000000000
Cthe_3051	+	transposase, mutator type	1223	0	0.000000000
Cthe_3054	+	hypothetical protein	389	0	0.000000000
Cthe_3057	+	transposase	269	0	0.000000000
Cthe_3058	-	hypothetical protein	479	0	0.000000000
Cthe_3059	+	phosphotransferase KptA/Tpt1	548	0	0.000000000
Cthe_3060	+	hypothetical protein	254	0	0.000000000
Cthe_3061	+	transcriptional repressor, CopY family	446	0	0.000000000
Cthe_3070	+	protein of unknown function DUF1256	590	0	0.000000000
Cthe_3071	+	protein of unknown function DUF1540	176	0	0.000000000
Cthe_3082	+	hypothetical protein	245	0	0.000000000
Cthe_3083	+	hypothetical protein	281	0	0.000000000
Cthe_3088	-		224	0	0.000000000
Cthe_3098	-	hypothetical protein	350	0	0.000000000
Cthe_3121	-	hypothetical protein	167	0	0.000000000
Cthe_3134	+	hypothetical protein	200	0	0.000000000
Cthe_3135	+	hypothetical protein	332	0	0.000000000
Cthe_3138	-	hypothetical protein	275	0	0.000000000
Cthe_3139	-	hypothetical protein	560	0	0.000000000
Cthe_3142	+	hypothetical protein	488	0	0.000000000
Cthe_3143	-	hypothetical protein	350	0	0.000000000
Cthe_3174	-	hypothetical protein	389	0	0.000000000
Cthe_3176	-	small acid-soluble spore protein, alpha/beta type	209	0	0.000000000
Cthe_3181	-	Integrase, catalytic region	1070	0	0.000000000
Cthe_3182	+	Integrase, catalytic region	1112	0	0.000000000
Cthe_3197	+	PRC-barrel domain protein	554	0	0.000000000
Cthe_3198	+	hypothetical protein	185	0	0.000000000
Cthe_3206	+	hypothetical protein	377	0	0.000000000
Cthe_3207	+	transposase, mutator type	1223	0	0.000000000
Cthe_3210	+	protein of unknown function DUF324	674	0	0.000000000
Cthe_3218	+	CRISPR-associated protein Cas1	995	0	0.000000000
Cthe_3236	+	hypothetical protein	305	0	0.000000000
Cthe_3238	+	hypothetical protein	242	0	0.000000000
Cthe_R0001	+	(tRNA)	74	0	0.000000000
Cthe_R0002	+	(tRNA)	73	0	0.000000000
Cthe_R0003	+	(tRNA)	73	0	0.000000000
Cthe_R0004	+	(tRNA)	76	0	0.000000000
Cthe_R0005	+	(tRNA)	83	0	0.000000000
Cthe_R0006	+	(tRNA)	76	0	0.000000000
Cthe_R0007	+	(tRNA)	75	0	0.000000000

المنارة للاستشارات
www.manaraa.com

Cthe_R0008	+	(tRNA)	75	0	0.000000000
Cthe_R0009	-	(tRNA)	75	0	0.000000000
Cthe_R0010	+	(tRNA)	86	0	0.000000000
Cthe_R0011	+	(tRNA)	75	0	0.000000000
Cthe_R0012	-	(tRNA)	74	0	0.000000000
Cthe_R0013	-	(tRNA)	77	0	0.000000000
Cthe_R0014	-	(tRNA)	75	0	0.000000000
Cthe_R0015	-	(miscRNA)	312	0	0.000000000
Cthe_R0016	-	(tRNA)	76	0	0.000000000
Cthe_R0017	-	(tRNA)	75	0	0.000000000
Cthe_R0018	+	(tRNA)	82	0	0.000000000
Cthe_R0019	+	(tRNA)	76	0	0.000000000
Cthe_R0020	+	(tRNA)	75	0	0.000000000
Cthe_R0021	+	(tRNA)	85	0	0.000000000
Cthe_R0022	+	(tRNA)	86	0	0.000000000
Cthe_R0023	-	(tRNA)	75	0	0.000000000
Cthe_R0024	-	(tRNA)	73	0	0.000000000
Cthe_R0025	+	(tRNA)	76	0	0.000000000
Cthe_R0026	-	(tRNA)	74	0	0.000000000
Cthe_R0027	-	(tRNA)	74	0	0.000000000
Cthe_R0028	-	(tRNA)	75	0	0.000000000
Cthe_R0029	-	(tRNA)	76	0	0.000000000
Cthe_R0030	-	(tRNA)	75	0	0.000000000
Cthe_R0031	-	(tRNA)	75	0	0.000000000
Cthe_R0032	-	(tRNA)	73	0	0.000000000
Cthe_R0033	-	(miscRNA)	105	0	0.000000000
Cthe_R0034	-	(tRNA)	75	0	0.000000000
Cthe_R0035	-	(tRNA)	75	0	0.000000000
Cthe_R0036	-	(tRNA)	75	0	0.000000000
Cthe_R0037	-	(tRNA)	93	0	0.000000000
Cthe_R0038	-	(rRNA)	116	0	0.000000000
Cthe_R0039	-	(rRNA)	2907	0	0.000000000
Cthe_R0040	-	(tRNA)	75	0	0.000000000
Cthe_R0041	-	(rRNA)	1520	0	0.000000000
Cthe_R0042	+	(tRNA)	90	0	0.000000000
Cthe_R0043	+	(tRNA)	93	0	0.000000000
Cthe_R0044	+	(tRNA)	76	0	0.000000000
Cthe_R0045	+	(tRNA)	76	0	0.000000000
Cthe_R0046	+	(tRNA)	90	0	0.000000000
Cthe_R0047	+	(tRNA)	89	0	0.000000000
Cthe_R0048	+	(tRNA)	75	0	0.000000000
Cthe_R0049	+	(tRNA)	76	0	0.000000000
Cthe_R0050	+	(tRNA)	74	0	0.000000000
Cthe_R0051	+	(tRNA)	75	0	0.000000000
Cthe_R0052	+	(tRNA)	74	0	0.000000000
Cthe_R0053	+	(miscRNA)	386	0	0.000000000
Cthe_R0055	+	(tRNA)	76	0	0.000000000

المنارة للاستشارات

www.manaraa.com

Cthe_R0056	+	(rRNA)	2907	0	0.000000000
Cthe_R0057	+	(rRNA)	116	0	0.000000000
Cthe_R0058	+	(rRNA)	1520	0	0.000000000
Cthe_R0059	+	(rRNA)	2907	0	0.000000000
Cthe_R0060	+	(rRNA)	116	0	0.000000000
Cthe_R0061	+	(tRNA)	75	0	0.000000000
Cthe_R0062	+	(tRNA)	74	0	0.000000000
Cthe_R0063	+	(tRNA)	76	0	0.000000000
Cthe_R0064	+	(tRNA)	75	0	0.000000000
Cthe_R0065	+	(tRNA)	84	0	0.000000000
Cthe_R0067	+	(tRNA)	75	0	0.000000000
Cthe_R0068	+	(rRNA)	2907	0	0.000000000
Cthe_R0069	+	(rRNA)	116	0	0.000000000
Cthe_R0070	+	(tRNA)	75	0	0.000000000
Cthe_R0071	-	(tRNA)	76	0	0.000000000
Cthe_RF0072	+	(miscRNA)	128	0	0.000000000
Cthe_RF0074	-	(miscRNA)	206	0	0.000000000
Cthe_RF0076	+	(miscRNA)	105	0	0.000000000
Cthe_RF0077	+	(miscRNA)	201	0	0.000000000
Cthe_RF0082	-	(miscRNA)	122	0	0.000000000
Cthe_RF0083	+	(miscRNA)	177	0	0.000000000
Cthe_RF0084	-	(miscRNA)	96	0	0.000000000
Cthe_RF0085	-	(miscRNA)	82	0	0.000000000
Cthe_RF0086	+	(miscRNA)	97	0	0.000000000
Cthe_RF0087	-	(miscRNA)	130	0	0.000000000
Cthe_RF0088	-	(miscRNA)	181	0	0.000000000
Cthe_RF0089	-	(miscRNA)	230	0	0.000000000
Cthe_RF0093	+	(miscRNA)	83	0	0.000000000
Cthe_RF0094	-	(miscRNA)	185	0	0.000000000
Cthe_RF0095	+	(miscRNA)	107	0	0.000000000
Cthe_RF0096	-	(miscRNA)	165	0	0.000000000
Cthe_RF0097	-	(miscRNA)	105	0	0.000000000
Cthe_RF0098	-	(miscRNA)	90	0	0.000000000
Cthe_RF0099	-	(miscRNA)	85	0	0.000000000

3. Possible Transcription Start Sites

Position	TSS Score	TSS Count	Genes	Feature Type	Description	Feature Coverage
3667200	7.880	19	Cthe_R0067|Cthe_R0068	operon		2.411
1414679	5.436	57	Cthe_1180|Cthe_1181|Cthe_1182|Cthe_1183|Cthe_1184|Cthe_1185|Cthe_1186	operon	uncharacterized predicted metal-binding protein | transglutaminase-like protein | protein of unknown function DUF58 | ATPase associated with various cellular activities, AAA_3 | hypothetical protein | conserved hypothetical protein | Alpha/beta hydrolase fold-3	10.487

www.manaraa.com

3211504	4.122	543	Cthe_2724|Cthe_2725|Cthe_2726|Cthe_2727|Cthe_2728|Cthe_2729|Cthe_2730|Cthe_2731|Cthe_2732|Cthe_2733|Cthe_2734	operon	DNA-directed RNA polymerase, beta subunit | DNA-directed RNA polymerase, beta' subunit | LSU ribosomal protein L7AE | SSU ribosomal protein S12P | SSU ribosomal protein S7P | translation elongation factor 2 (EF-2/EF-G) | translation elongation factor 1A (EF-1A/EF-Tu) | RNA polymerase, sigma-24 subunit, ECF subfamily | hypothetical protein | hypothetical protein | VanW	131.718
1414681	3.814	40	Cthe_1180|Cthe_1181|Cthe_1182|Cthe_1183|Cthe_1184|Cthe_1185|Cthe_1186	operon	uncharacterized predicted metal-binding protein | transglutaminase-like protein | protein of unknown function DUF58 | ATPase associated with various cellular activities, AAA_3 | hypothetical protein | conserved hypothetical protein | Alpha/beta hydrolase fold-3	10.487
2042072	3.596	6	Cthe_1713|Cthe_1714|Cthe_1715|Cthe_1716|Cthe_1717|Cthe_1718|Cthe_1719|Cthe_1720|Cthe_1721|Cthe_1722	operon	phage major tail protein, phi13 family | conserved hypothetical protein | phage protein, HK97 gp10 family | phage head-tail adaptor, putative | conserved hypothetical protein | hypothetical protein | phage major capsid protein, HK97 family | peptidase S14, ClpP | phage portal protein, HK97 family | phage Terminase	1.668
705679	3.582	97	Cthe_0576|Cthe_0577	operon	ribulose-5-phosphate 3-epimerase | thiamine pyrophosphokinase	27.083
3667201	3.318	8	Cthe_R0067|Cthe_R0068	operon		2.411
2242923	3.140	4	Cthe_1888	CDS	hypothetical protein	1.274
994630	2.948	10	Cthe_0818|Cthe_0819|Cthe_0820	operon	hypothetical protein | ABC transporter related protein | hypothetical protein	3.392
2272053	2.835	12	Cthe_1908|Cthe_1909|Cthe_1910	operon	copper amine oxidase-like protein | copper amine oxidase-like protein | hypothetical protein	4.233
1579639	2.832	10	Cthe_1302	CDS	conserved hypothetical protein	3.531
451211	2.820	3	Cthe_0358	CDS	hypothetical protein	1.064
367774	2.755	3	Cthe_0292|Cthe_0293|Cthe_0294	operon	transposase, mutator type | hypothetical protein	1.089
1477765	2.551	10	Cthe_1232|Cthe_1233|Cthe_1234	operon	AMP-dependent synthetase and ligase | hypothetical protein | methyl-accepting chemotaxis sensory transducer	3.920
1993478	2.427	5	Cthe_1666	CDS	transposase, IS4	2.060
1706863	2.384	7	Cthe_1394|Cthe_1395|Cthe_1396|Cthe_1397	operon	two component transcriptional regulator, winged helix family | phospholipase	2.936

www.manaraa.com

					D/Transphosphatidylase | glycosyl transferase, family 8	
3289577	2.335	5	Cthe_2789|Cthe_2790|Cthe_2791	operon	NLPA lipoprotein | ABC transporter related protein | binding-protein-dependent transport systems inner membrane component	2.142
3127863	2.311	5	Cthe_2643|Cthe_2644|Cthe_2645|Cthe_2646|Cthe_2647|Cthe_2648|Cthe_2649	operon	Nucleotidyl transferase | DegT/DnrJ/EryC1/StrS aminotransferase | NAD-dependent epimerase/dehydratase | hypothetical protein | protein of unknown function DUF201 | sugar transferase | HpcH/HpaI aldolase	2.164
3722380	2.218	20	Cthe_3149|Cthe_3150|Cthe_3151|Cthe_3152|Cthe_3153|Cthe_3154	operon	aminoacyl-histidine dipeptidase | adenosylcobinamide-phosphate synthase | adenosylcobinamide kinase | cobalamin-5'-phosphate synthase | alpha-Ribazole phosphatase | conserved hypothetical protein	9.016
245828	2.191	24	Cthe_0197|Cthe_0198|Cthe_0199|Cthe_0200|Cthe_0201|Cthe_0202|Cthe_0203	operon	glutamate synthase (NADPH) GltB1 subunit | glutamate synthase (NADPH) GltB2 subunit | 4Fe-4S ferredoxin, iron-sulfur binding | FAD-dependent pyridine nucleotide-disulphide oxidoreductase | glutamate synthase (NADPH) GltB3 subunit | L-glutamine synthetase | response regulator receiver and ANTAR domain protein	10.955
2816481	2.184	7	Cthe_R0038|Cthe_R0039|Cthe_R0040	operon		3.205
2816512	2.184	7	Cthe_R0038|Cthe_R0039|Cthe_R0040	operon		3.205

المنارة للاستشارات
www.manaraa.com

VITA

Christopher Mark Gowen was born on December 12, 1983 in Memphis, Tennessee and is an American citizen. He graduated from Bolton High School in Arlington, TN with the class of 2002. He received a B.S. in Biosystems engineering and a minor in Bioengineering from Clemson University in May 2006, and a M.S. in Chemical and Life Science Engineering in December 2008.

المنارة للاستشارات
www.manaraa.com