UNIVERSITY OF FLORENCE

FACULTY OF MATHEMATICAL, PHYSICAL AND NATURAL SCIENCES

PHD IN GENETICS, XXIII CYCLE

STUDY OF *Bacillus subtilis* GENES INVOLVED IN CALCIUM CARBONATE BIOMINERALIZATION

A Thesis submitted for the degree of Doctor of Philosophy in Genetics; Doctorate School "Ubaldo Montelatici"; University of Florence, Dept. of Evolutionary Biology "Leo Pardi"

BIO19

PHD CO-ORDINATOR :
(Prof. Milvia Luisa Racchi)

TUTOR :
(Prof. Giorgio Mastromei)

CO-TUTOR :
(Dr. Brunella Perito)

CANDIDATE :
(Antonio Frandi)

Florence, December 2010

To my family.
Ohana means family.
Family means nobody gets left behind, or forgotten.

Acknowledgments

Far too many people to mention individually have assisted in so many ways during my work at DBE. They all have my sincere gratitude.

To start with, I would like to express my gratitude to my supervisors, Prof. Giorgio Mastromei and Dr. Brunella Perito, whose expertise, understanding, and patience, added considerably to my graduate experience. In this I would also like to include my gratitude to Mrs. Cristina Indorato for technical assistence and her extremely valuable experiences.

A very special thanks goes out to Dr. Emanuele Biondi, without whose motivation and encouragement I would never have been able to look forward. Thank you for your close guidance as well as for the debates, exchanges of knowledge, skills, and venting of frustration during my graduate program, which helped enrich the experience.

A penultimate thank you goes to my wonderful friends, Giorgio Mattiuz "Giorigo", Matteo Brilli, Francesco Pini "Pino", and Count Marco Fondi, ho believed in me as a scientist and always supporting me.
I would like to thank Dr. Alessio Mengoni and Prof. Renato Fani; you were always available and very helpful. Thanks for our insightful discussions.

Finally, my parents: Armando Frandi and Renza Innocenti Frandi. They gave me my name, they gave me my life, and everything else in between. I pride myself in having words for everything, but they truly shut me up when it comes down to describing how much I love them and appreciate the efforts they have put into giving me the life I have now. They are the

reason I did this; they are the reason I thrive to be better.

My final, and most heartfelt, acknowledgment must go to the love of my life Giulia. Her support, encouragement, and companionship has turned my journey through the PhD into a pleasure. For all that, and for being everything I am not, she has my everlasting love. Thank you and I love you.

Publications

Peer-reviewed publications:

1. Papaleo M.C., Russo E., Fondi M., Emiliani G., **Frandi A.**, Brilli M., Pastorelli R., Fani R. (2009) Structural, evolutionary and genetic analysis of the histidine biosynthetic "core" in the genus *Burkholderia*. Gene. 448(1):16-28.

2. **Frandi A.**, Mengoni A., Brilli M.(2010) Comparative Genomics of VirR regulons in *Clostridium perfringens* strains. BMC Microbiology 10:65, doi:10.1186/1471-2180-10-65.

3. **Frandi A.**, Zucca P., Marvasi M., Sanjust E., Mastromei G., Perito B. (2010) *Bacillus subtilis fadB(ysiB)* encodes for an Enoyl-CoA hydratase. Annals of Microbiology, doi: 10.1007/s13213-010-0121-5.

Contents

List of Figures

List of Tables

1 INTRODUCTION

The term *biomineralization* refers to the processes by which organisms form minerals. The control exerted by many organisms over mineral formation distinguishes these processes from abiotic mineralization. The latter was the primary focus of earth scientists over the last century, but the emergence of understanding the past and future evolution of the Earth are moving biological mineralization to the forefront of various fields of science. Of the wonderful topics that are receiving renewed attention, the study of biomineral formation is perhaps the most fascinating. It is, however, by no means a new field. Until the early 1980s the field was known as calcification, reflecting the predominance of biologically formed calcium containing minerals. As more and more biogenic minerals were discovered that contained other cations, the field became known as biomineralization. An extensive knowledge base has been established in literature with styles as varied as the biominerals themselves [70, 102, 72, 114, 66, 40, 47].

1.1 BIOMINERALIZATION

1.1.1 WHAT IS A BIOMINERAL ?

The term biomineral refers not only to a mineral produced by organisms, but also to the fact that almost all of these mineralized products are composite materials comprised of both mineral and organic components. Furthermore, having formed under controlled conditions, biomineral often have properties such as shape, size, crystallinity, isotopic and trace element compositions quite unlike its inorganically formed counterpart. The term biomineral reflects all this complexity. Figure 1.1 illustrates this by

comparing part of a single calcite crystal formed by an echinoderm to synthetic single crystals of calcite and that produced by *B. subtilis*.

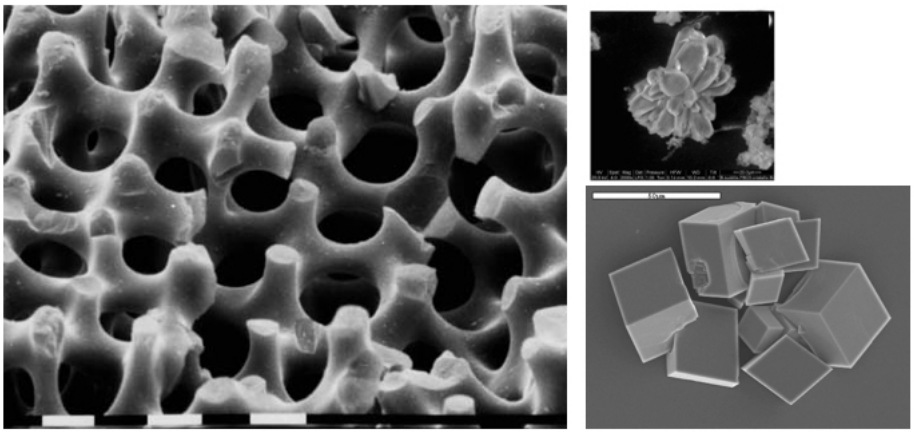

Figure 1.1: Comparison of calcite single crystals: (*left*) stereom of echinoderm, and (*right*) synthetically produced rhombohedral forms, (*upper right*) produced by *B. subtilis*.

1.1.2 GENERAL CHARACTERISTICS OF BIOMINERALS

As indicated in Table 1.1, calcium is the cation of choice for most organisms. The calcium-bearing minerals comprise about 50% of known biominerals [70]. This comes as no surprise because calcium fulfills many fundamental functions in cellular metabolism [69, 102, 16]. This dominance of calcium-bearing minerals is what led to the widespread usage of the term calcification.
The calcium carbonate minerals are the most abundant biogenic minerals, both in terms of the quantities produced and their widespread distribution among many different taxonomic groups [70]. Of the eight known polymorphs of $CaCO_3$, seven are crystalline and one is amorphous. Three of the polymorphs calcite, aragonite and vaterite are pure calcium carbonate, while two monohydrocalcite and the stable forms of amorphous calcium carbonate contain one water molecule per calcium carbonate [1]. One of the major challenges in the field of biomineralization is to understand the mechanism(s) by which biological systems determine which polymorph will precipitate. This is genetically controlled and is almost always achieved with high fidelity.
Phosphates comprise about 25% of the biogenic mineral types. Except for struvite and brushite, most phosphate minerals are produced by controlled

mineralization. The most abundantly produced phosphate mineral is carbonated hydroxyapatite, also called dahllite [70]. It is the mineral present in vertebrate bones and teeth, as well as in the shells of inarticulate brachiopods.
The iron biominerals are not readily evaluated by mineral class because they have significant occurrences as oxides, hydroxides, and sulfides [12, 64]. The iron biominerals are of particular significance because they comprise approximately 40% of all minerals formed by organisms [11], and magnetite formation is believed to be the most ancient matrix-mediated biomineralizing system. As such, Kirschvink and Hagadorn [62] have suggested that magnetic biominerals, and iron minerals in general, could contain clues to other aspects of controlled mineralization.
Table 1.1 also contains a group of so-called organic minerals. Despite the contradiction in terms, these are crystalline phases formed by organisms probably by the same underlying strategies used for normal mineral formation. This suggests that many of these minerals remain to be discovered, and exploring the functions they perform will be fascinating.

Table 1.1: Names and chemical compositions of minerals produced by biologically induced and controlled mineralization processes

Names	**Formula**
Carbonates	
Calcite	$CaCO_3$
Aragonite	$CaCO_3$
Vaterite	$CaCO_3$
Protodolomite	$CaMg(CO_3)_2$
Amorphous calcium carbonate (at least 5 forms)	$CaCO_3 \cdot H_2O$ or $CaCO_3$
Phosphates	
Octacalcium phosphate	$Ca_8H_2(PO_4)_6$
Brushite	$CaHPO_4 \cdot 2H_2O$
Francolite	$Ca_{10}(PO_4)_6F_2$
Carbonated-hydroxylapatite (dahllite)	$Ca_5(PO_4,CO_3)_3(OH)$
Whitlockite	$Ca_{18}H_2(Mg,Fe)_2{}^{+2}(PO_4)_{14}$
Struvite	$Mg(NH_4)(PO_4) \cdot 6H_2O$
Vivianite	$Fe_3{}^{+2}(PO_4)_2 \cdot 8H_2O$
Sulfates	

CONTINUED ON NEXT PAGE

Names	**Formula**
Gypsum	$CaSO_4 \cdot 2H_2O$
Barite	$BaSO_4$
Celestite	$SrSO_4$
Jarosite	$KFe_3{}^{+3}(SO_4)_2(OH)_6$
Sulfides	
Pyrite	FeS_2
Hydrotroilite	$FeS \cdot n$H2O
Sphalerite	ZnS
Wurtzite	ZnS
Galena	PbS
Greigite	Fe_3S_4
Mackinawite	$(Fe,Ni)_9S_8$
Acanthite	Ag_2S
Hydrated silica	
Amorphous silicates	$SiO_2 \cdot nH_2O$
Oxides	
Magnetite	Fe_3O_4
Amorphous Ilmenite	$Fe^{+2}TiO_3$
Amorphous Iron Oxide	Fe_2O_3
Amorphous Manganese Oxide	Mn_3O_4
Hydroxides & Hydrous Oxides	
Goethite	α-FeOOH
Lepidocrocite	γ-FeOOH
Ferrihydrite	$5Fe_2O_3 \cdot 9H_2O$
Todorokite	$(Mn^{+2}CaMg)Mn_3{}^{+4}O_7 \cdot H_2O$
Birnessite	$Na_4Mn_{14}O_{27} \cdot 9H_2O$
Organic Crystals	
Earlandite	$Ca_3(C_6H_5O_2)_2 \cdot 4H_2O$
Whewellite	$CaC_2O_4 \cdot H_2O$
Glushinskite	$MgC_2O_4 \cdot 4H_2O$
Manganese Oxalate (unnamed)	$Mn_2C_2O_4 \cdot 2H_2O$
Sodium urate	$C_5H_3N_4NaO_3$
Uric Acid	$C_5H_4N_4O_3$

CONTINUED ON NEXT PAGE

Biominerals – Continued

Names	Formula
Ca tartrate	$C_4H_4CaO_6$
Ca malate	$C_4H_4CaO_5$
Paraffin Hydrocarbon	
Guanine	$C_5H_3(NH_2)N_4O$

Biominerals meet the criteria for being true minerals, but they can also hold other characteristics that distinguish them from their inorganically produced counterparts, figure 1.2. The most obvious feature is that biogenic minerals have unusual morphologies. Implicit in this complexity is the intriguing ability of organisms to impose a handedness upon the morphology of biominerals. Good examples are found from the microscopic (*e.g: bacteria*) to macroscopic (*e.g: shells*) organisms.
A second characteristic of biominerals is that many are actually composites or agglomerations of crystals separated by organic material. As biominerals becomes increasingly characterized a better understanding of the assembly of biominerals could change our interpretations of mineralization mechanisms.

1.1.3 How are produced biominerals ?

For nucleation and growth to occur, biomineral formation requires a localized zone that achieves and maintains a sufficient supersaturation. In most biological systems, the site of mineral deposition is isolated from the environment by a physical delimiting geometry. The actual size of that site or volume is sometimes ambiguous, but it is generally agreed that this region must limit diffusion into/out of the system or utilize a type of compartment. The extent of this isolation can be passive and minimal such as observed when bacteria cluster to form an intercellular zone that is diffusion-limited. At the other extreme, intracellular vesicles create compartmentalized environments where compositions can be precisely regulated. This compartment must be capable of modifying the activity of at least one biomineral constituent (usually the cation) as well as protons and

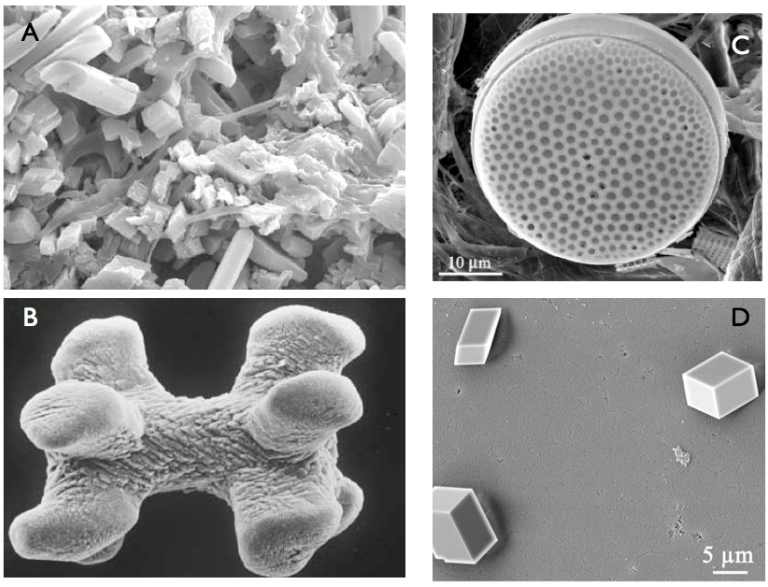

Figure 1.2: Diversity and complexity of bio-originated structures; (A) Calcite crystals produced by *Myxococcus xhantus* ; (B) $CaCO_3$ crystal produced by ; (C) Calcite crystals produced by Diatoms ; (D) Synthetically produced calcite crystals.

possibly other ions [1]. Ion supply (or removal) occurs by two means: active pumping near the sites of mineralization or passive diffusion gradients.

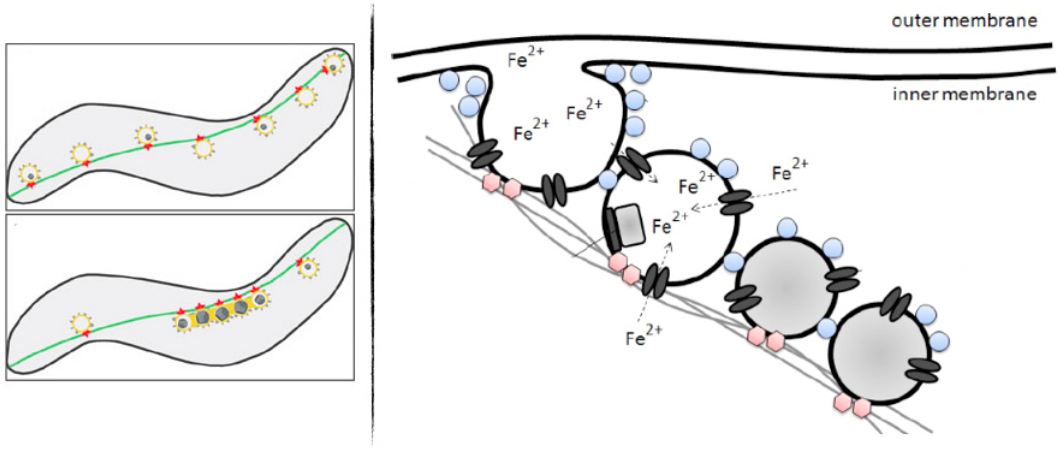

Figure 1.3: Schematics of magnetosome formation in magnetotactic bacteria [10, 100, 11]

Biomineralization processes are divided into two fundamentally different groups based upon their degree of biological control. Lowenstam (1981) introduced these as biologically induced and organic matrix-mediated [68], with the latter generalized by Mann (1983) to biologically controlled mineralization [71]. It is the specific nature and degree of control that is central to understanding the extent of biological control of the elemental compositions of biominerals.

Biologically controlled biomineralization

In biologically controlled mineralization, the organism uses cellular activities to direct the nucleation, growth, morphology and final location of the mineral that is deposited. While the degree of control varies across species, the results can be remarkably sophisticated, species-specific products that give the organism specialized biological functions. Biologically controlled mineralization processes can be described as occurring extra-, inter- or intracellularly. These distinctions refer to the locations of the mineralization site with reference to the cells responsible for mineralization.

[1] Any fluxes in ion chemistry must meet one constraint: the fluid must maintain electroneutrality

However, not all mineralization processes can be classified in this simple manner. In some cases, mineral formation begins within the cell and then proceeds outside the cell. Identifying what are in essence, end members at least helps us understand the overall complexity.

- Biologically controlled *extra*cellular biomineralization:
  In extracellular mineralization, figure 1.4. the cell produces a macromolecular matrix outside the cell in an area that will become the site of mineralization. The term matrix refers to a group of macromolecules comprised of proteins, polysaccharides or glycoproteins that assemble to form a three-dimensional framework. The matrix composition is unique in that many of its proteins contain a high proportion of acidic amino acids (especially aspartate) and phosphorylated groups. The structures and compositions of these organic frameworks are genetically programmed to perform essential regulating and/or organizing functions that will result in the formation of composite biominerals. In this approach, the cell works actively to supply cations to an external organic matrix for on-site nucleation and growth. This is distinguished from the epicellular nucleation and growth that occur during biologically induced mineralization.

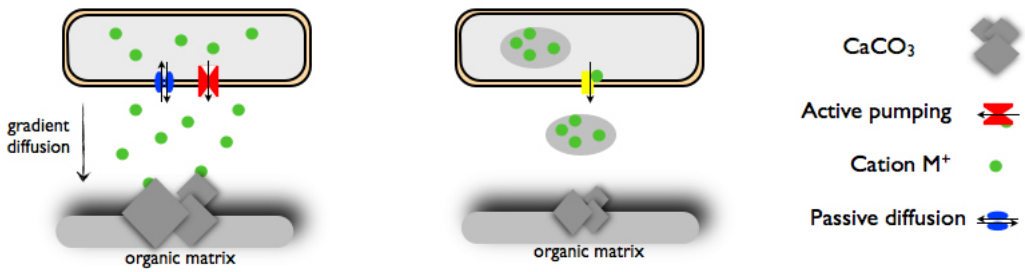

Figure 1.4: Biologically controlled *extra*cellular mineralization. *left* Cations are pumped across the cell membrane and move by passive diffusion through extracellular fluids to the site of mineralization. *right* Cations are concentrated intracellularly as aqueous ions into a vesicle that is subsequently secreted. Compartment breakdown at site of mineralization releases cations for biomineral formation.

- Biologically controlled *inter*cellular biomineralization, figure 1.5:
  It typically occurs in single-celled organisms that exist as a community. At first glance, intercellular mineral formation appears to be

a variant of extracellular mineralization. This might appear to be a type of biologically induced mineralization, but studies have shown that organisms directs the polymorph and shape of the biomineral that forms. An example is found in calcareous algae [21, 20].

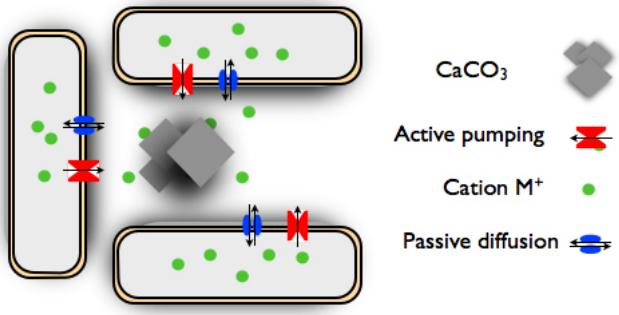

Figure 1.5: Biologically controlled *inter*cellular mineralization, the surfaces of cells are used as the organic substrate for nucleation and growth with a preferred orientation. Cations are pumped out of the cell and fluid compositions are regulated to maintain control over the biomineral type and extent of growth.

- Biologically controlled *intra*cellular biomineralization:
  controlled mineralization, figure 1.6, can also occur also within specialized vesicles or vacuoles that direct the nucleation of biominerals within the cell. This is a widespread strategy. These compartmentalized crystallization environments govern the resulting biomineral composition and morphology. In this situation, the cell has a high degree of control upon the concentrations of cation and anion biomineral constituents in an environment where an organic matrix may also be active as a nucleating template. The compartment membrane also regulates the pH, pCO_2. One of the best example is exhibited by the magnetosome-producing bacteria. These structures are membranebound, euhedral crystals of magnetite or greigite that assemble *via* their magnetic fields to produce biomineral chains [10, 100, 11], figure 1.3.

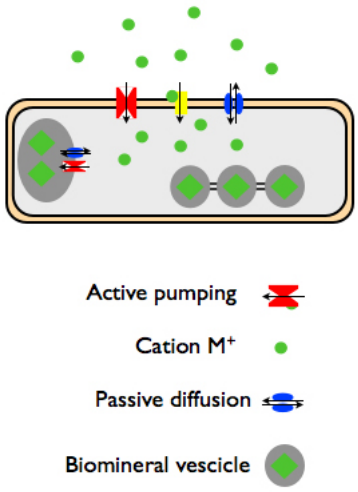

Figure 1.6: Biologically controlled *intra*cellular mineralization shows that nucleation occurs within the cell in a specialized vesicle. Biomineral is nucleated within a compartment in the intracellular environment. These growth units may be assembled intracellularly for subsequent secretion or the biomineral unit remains within the cell as a single growth unit organized into a higher order structure.

Biologically induced biomineralization

The secondary precipitation of minerals that occurs as a result of interactions between biological activity and the environment is termed biologically induced mineralization, figure 1.7. In this situation, cell surfaces often act as causative agents for nucleation and subsequent mineral growth. The biological system has little control over the type and habit of minerals deposited, although the metabolic processes employed by the organism within its particular redox environment mediate pH, pCO_2 and the compositions of secretion products [76, 52, 107, 11]. In some cases, biological surfaces are important in the induction stage because nucleation often occurs directly on the cell wall, and the resulting biominerals can remain firmly attached. In open waters, this epicellular mineralization can lead to encrustation so complete that gravitation overcomes buoyancy, and they settle through the water column. The sediment record attests to the extensive occurrence of this phenomenon [63, 109].
Heterogeneity is the hallmark of biologically induced minerals. Frankel and Bazylinski (2003) show that the compositions of minerals resulting from induced processes vary as greatly as the environments in which they

form [11]. This heterogeneity includes variable external morphology (typically poorly defined), water content, trace/minor element compositions, structure and particle size. Since these characteristics also typify inorganically precipitated minerals, unambiguous interpretations of the sediment and rock record continue to confound interpretations of earth environments.

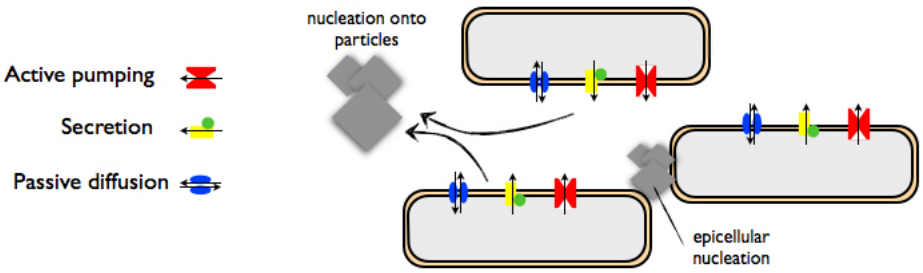

Figure 1.7: Biologically induced mineralization. Mineral precipitates form as a result of metabolic activities that affect pH, pCO_2, and secretion products. The cell is a causative agent only, without control over mineral type or habit.

1.2 Biomineralization induced by Bacteria

Organisms which have the capability of precipitating minerals are present in all the major groups, from Bacteria to Chordata. However, concerning the way in which these biominerals are formed, a marked difference exists among the different taxonomic groups [102]. The foremost group able to induce the precipitation of minerals is the animals. The second group is bacteria, then vascular plants and finally fungus and protozoa [70]. Bacteria are considered as agents that disperse, fractionate or concentrate material. As concentration agents, they can:

1. accumulate inorganic material via processes such as intracellular deposition, adsorption, and fixation at cellular level, and

2. precipitate of insoluble extra or intracellular compounds.

Bacteria are able to induce the precipitation of minerals, either by highly controlled biomineralization [10], or by inducing the precipitation of minerals through processes that involve little control over the biomineralization (the so-called biologically induced biomineralization), for instance, by changing the chemistry of the environment as a result of the bacterial metabolic activity. Microorganisms have a geochemical activity which is responsible to a great extent for the deposition of minerals throughout the history of the Earth. Their central role over a wide range of mineralization processes has already been established. On one hand, bacteria can change the chemistry of the environment where they grow as a result of their metabolic activity. Therefore, they can create supersaturated conditions with respect to a particular mineral phase thus inducing the precipitation of such a phase. The processes that result in the precipitation of different biominerals are very different, due to the great variety of bacterial metabolic pathways. On the other hand, as it has been mentioned by numerous authors, bacteria may contribute to mineral precipitation not only actively, but also passively, by serving, both the cell, cell structures and the cell debries, as nucleation sites for mineral deposition [35, 111, 85, 15]. In this regard, bacteria can induce heterogeneous nucleation of minerals by providing not only a surface that lower the energy barriers for mineral precipitation, but also a stereochemical arrangement of the mineral components. Many studies place special emphasis on the role of microbial extracellular substances (EPS). These exopolymers and the biofilms are commonly dominated by negatively charged polysaccharides [42] and can contribute to mineral precipitation in numerous ways: 1) by trapping positively charged cations in the negatively charged sites of the EPS and

serving as a template for the nucleation of crystals [87] and/or 2) by trapping detrital seed crystals which act as nuclei for heterogeneous mineral precipitation [111].

1.2.1 Production of calcium carbonates

Bacterially induced or mediated carbonate mineralization is thought to be important in a range of processes including atmospheric CO_2 budgeting, carbonate sediment and rock formation, biogeochemical cycling of elements, metal-contaminated groundwater bioremediation, and conservation of ornamental stone [49]. It is known that a number of processes are involved in bacterially induced calcium carbonate precipitation, including non-methylotrophic metanogenesis, photosynthesis, ammonification, denitrification and sulphate reduction [33]. In fact, it has been observed that the number of bacteria capable of producing calcium carbonate is considerably high, among others: sulphate-reducing bacteria and cyanobacteria [10], *Bacillus* [33], *Myxobacteria* [14], *Halobacillus* [92] and *Pseudomonas* [9]. Calcite and aragonite, the two most stable polimorphs of calcium carbonate, are the most common microbial carbonates [49]. Elucidating how bacterial calcite and vaterite mineralization occurs may have further implications: it may lead to a better understanding of microbial carbonate mineralization, and help identify biosignatures both on Earth and elsewhere. Furthermore, it may shed light on the calcium carbonate polymorphism problem, which is one of the biggest challenges in the field of biomineralization.

1.2.2 Some example of induced bacterial calcium carbonate precipitation

Myxococcus xanthus induced calcium carbonates precipitation

It has been demonstrated that *M. xanthus*, a Gram-negative soil-sliding bacteria, has a remarkable ability to induce the precipitation of carbonates, mainly calcite, vaterite and Mg-calcite [14]. *M. xanthus* induces carbonate precipitation by changing the supersaturation conditions with respect to the particular carbonate phase. *M. xanthus* increases the alkalinity of the culture media as a consequence of the release of CO_2 and NH_3 resulting from its metabolic activity. Extracellular ammonia release raises pH values and therefore ${CO_{3(aq)}}^{2-}$ concentration, according to these equilibria:

$$NH_{3(g)} + H_2O \Leftrightarrow NH_4{}^+{}_{(aq)} + OH^-{}_{(aq)}$$

$$CO_{2(g)} + H_2O_{(l)} \Leftrightarrow H_2CO_{3(aq)} \Leftrightarrow HCO_3{}^-{}_{(aq)} + H^+{}_{(aq)} \Leftrightarrow CO_3{}^{2-}{}_{(aq)} + 2H^+{}_{(aq)}$$

$$Ca^{2+}{}_{(aq)} + CO_3{}^{2-}{}_{(aq)} \Leftrightarrow CaCO_{3(s)}$$

Calcite frequently precipitates from culture media that contain calcium and which have a starting supersaturation value relatively low.
Very interesting is the ability that *M. xanthus* has to induce the precipitation of vaterite, the most unstable polimorph of calcium carbonate. Vaterite has been found in calcareous sediments, metamorphic rocks and drilling muds, portland cement, and in mortars of the Florence Cathedral. Despite such occurrences, some controversy exists regarding whether vaterite can precipitate inorganically from natural waters. As mentioned above, most natural occurrences of vaterite are associated with biogenic activity. *M. xanthus* is able to precipitate vaterite in culture media containing calcium and which have relatively higher supersaturation conditions. Bacterial production of CO_2 and NH_3 and the transformation of the NH_3 to $NH_4{}^+$ and OH^-, thus increasing solution pH and carbonate alkalinity, set the physicochemical conditions (high supersaturation) leading to vaterite precipitation in the microenvironment around cells, and directly onto the surface of bacterial cells. In the latter case, fossilization of bacterium also occurred.
Moreover, when *M. xanthus* grows with high calcium and magnesium concentration, is able to induce the precipitation of magnesium calcite [14]. This was explained as a result of a series of local fluctuations in supersaturation as a consequence of, on one hand, the diffusion of the released CO_2 and NH_3 by bacterial activity from the zones of higher density of microbial growth and, on the other hand, the diffusion of ions in the opposite direction forced by the consumption of nutrients by bacterium. Therefore, the counter diffusion of microbial metabolites and culture ions produce spatio-temporal gradients of concentration that at certain moments and at certain points raise the supersaturation conditions to allow the precipitation of the specific mineral phase. Finally, *M. xanthus* calcified itself when it grows in presence of high [Ca^{2+}], as demonstrated by the findings of bacterial cells surrounded by nanometer-sized carbonate crystals. Nevertheless, it appears that the presence of bacterial cells is not a prerequisite for the formation of calcite, although bacterial activity does seem to be a prerequisite for calcite formation.

Bacillus subtilis induced calcium carbonates precipitation

Bacillus subtilis produces calcium carbonate crystals when cultivated in appropriate media supplemented with a Ca^{2+} source (*e.g:* $Ca(CH_3COOH)_2$). In calcite formation by *B. subtilis*, figure 1.8, precipitation was seen to involve metabolic pathways, ion exchange and pH as the major contributor. A step in the precipitation of calcium carbonate by *B. subtilis* has been described by the pioneeristics studies of T. Beveridge, as involving stoichiometric interactions between positively charged Ca^{2+} ions and the negatively charged bacterial cell wall. These Ca^{2+} ion-cell wall interactions produces changes in the overall charge of the bacterial cell wall, inducing the formation of the biomineral. It has been proposed that this high density of electronegative sites within the cell wall gives *B. subtils* the ability to form precipitates. There is some selctivity for different cations, and the wall of *B. subtilis* and Gram-positive bacteria is an especially good sink for protons (H^+). Metabolizing cells are continuosly pumping H^+ into the wall of a bacterium where negatively charged groups are situated, resulting in a delicate competition between H^+ and Ca^{2+} ions. Indeed metabolizing cells should bind less cations (*e.g.:* Ca^{2+}) to their walls than non-living or quiescently bacteria. In *B. subtilis*, the proton motive force, takes part, either directly or indirectly, in mechanical movement, chemotaxis, genetic processes, active transports, movement of charged solutes, and biomineralization as well.
Recent studies have reported the isolation of *B. subtils* mutants impaired in calcite precipitation, and the forces behind the impairment. The identification of four genes involved in calcite precipitation in *B. subtilis* have suggested that the process is genetically controlled, and the function of the four genes (*ysiA, ysiB, etfB,* and *etfA*) was seen to be involved in fatty acids metabolism. Despite the lack of knowledge about the mechanism through which these genes are involved in the biomineralization process, it seems to be relied on at least the *etfA* gene which has resulted to be essential for calcite precipitation. Further investigation, on the mechanism by which calcite precipitation in *B. subtilis* relied on, have demonstrated that pH variations are the main determinant for calcite precipitation.
However, the precise role(s) of *B. subtilis* in the process of calcite precipitation remain largely elusive but seem to fall categorically into three different yet related tracks, shared by most bacterial species able to precipitate calcium carbonate. First, mineralization occurs as a by-product of microbial metabolism that involves either autotrophic pathways inducing local CO_2 depletion or heterotrophic pathways leading to an increase of CO_3^{2-} in the media [33]. An alternative hypothesis proposes that the $CaCO_3$ pre-

cipitation is controlled by intracellular calcium metabolism [2, 76], instead of by changing $CO_3{}^{2-}$ concentration. Second, calcium carbonate nucleation takes place on cell walls due to ion exchange through the cell membrane (i.e. an active process, [33]) following some still poorly known mechanisms [92, 33, 34], or due to promotion by the negatively charged specific cell wall functional groups that adsorb divalent cations such as Ca^{2+} [92]. The third possibility involves the extracellular macromolecules because these biopolymers are shown to be able to trap calcium ions or serve as growth modifiers to control crystallization.

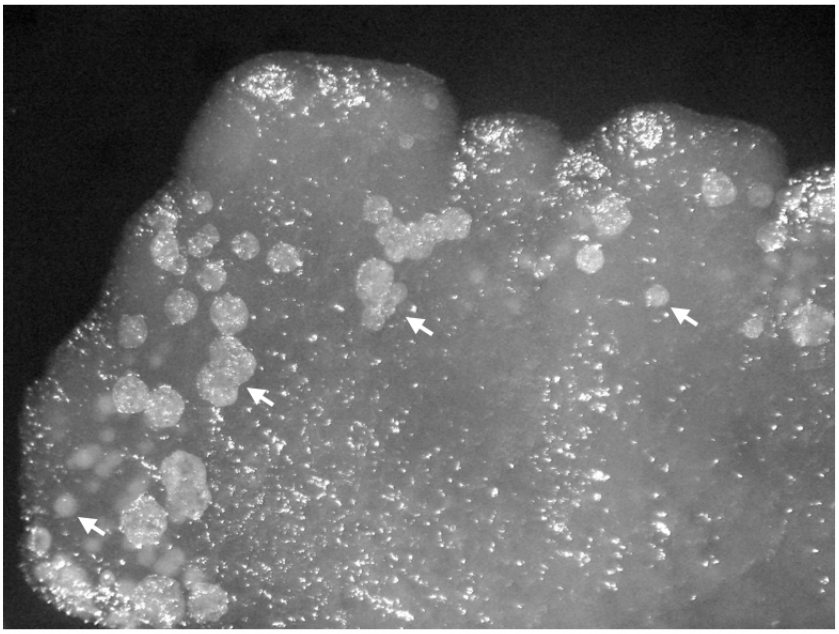

Figure 1.8: Calcite crystals produced by *B. subtilis* 168, white arrows marks $CaCO_3$ precipitates.

1.3 Proposed applications for bacterial induced biomineralization

Because biomineralization is such a wide phenomenon, there is an increasing interest to understand the mechanisms involved in such process, since they can provide unique information regarding, for instance, the origin and evolution of life in Earth and elsewhere in the solar system, bacterial interactions with the environment, environmental conditions of sediments of all ages and the recognition of the bacterial role of different human pathologies. Also, biomineralization has a high potential of being used in a wide range of technical applications, including for instance, nanotechnology, implants and restoration [33, 95].
The ecological implications of bacterially induced biomineralization are various and several papers deal with biomineralization [24, 117, 46]. Moreover, the ecological relevance of induced mineralization such as microbial calcium carbonate precipitation has been exploited in biotechnology, with applications ranging from bioremediation to control leaching [54], plugging-cementation of rocks [105], solid-phase capture of inorganic contaminants [112] and the consolidation of carbonate materials of monuments [8, 95].
In the field of stone conservation, the ability of different bacterial species, such as *B. pasteurii, M. xanthus, B. sphaericus, B. lentus,* and *B. subtilis* to act as limestone consolidants has also been investigated. Furthermore, the ability of *B. pasteurii* to precipitate carbonates has been investigated with respect to consolidation of sand columns [54, 105] and repair of concrete cracks [6]. Taken togehether, these applications open a new and promising future based on the development of methods to be applied in a wide range of strategy from restoration to biomeremediation.

2

PRESENTATION OF THE WORK

Calcium carbonate biomineralization is a widespread process among organisms from *bacteria* to *eukarya*. Calcium carbonate deposition by bacteria is a very diffuse phenomenon and represents a fundamental part of the calcium biogeochemical cycle [118]. It is generally accepted that this microbial activity can be influenced by environmental physical-chemical parameters, and it is correlated to both metabolic activities and cell surface structures [52, 44, 17]. Different mechanisms of bacterial involvement in calcification have been proposed, however the role played by bacteria in this process is still unclear [48, 49]. $CaCO_3$ precipitation by *Bacteria* is a complex process which was described by Castanier et al. [34] as involving metabolic pathways associate with photosynthesis, nitrogen and sulfur cycles, and ion exchange (Ca^{2+}/H^{+}). A step in the precipitation of calcium carbonate by *Bacteria* has been described as involving stoichiometric interactions between positively charged Ca^{2+} ions and the negatively charged bacterial cell wall. These Ca^{2+} ion-cell wall interactions produce changes in the overall charge of the cell wall, allowing interaction between differently charged groups. As a consequence of these changes in the overall ionic charge, bacteria aggregate to increase the size of the biomineral, and in turn, bacteria become the nucleus of the biomineral (25, 49,51). Extracellular polymeric substances (EPS) produced by bacteria have also been shown to be involved in the process of biocalcification by entrapping as serving as a nucleation site for crystal deposition. Carbonate precipitation by bacteria is known to be affected by several factors , such as temperature, ionic strength, composition of the growth media, and pH. Besides these factors the type of organism and the incubation time need to be taken into consideration since they can also affect the extent of the precipitation and the type of carbonate precipitated as seen in the studies of Ferrer et

al. 1988. The role of bacteria in calcification is still unknown, as there are many potential applications for this process, it is useful to understand the molecular and genetic basis for this process as would the process to be exploited in a controllable fashion and manipulated to be more effective. Applications include the fields of bioremediation,solid-phase capture of inorganic contaminants, and plugging-cementation in rock and other porous media fissures. Bacterial $CaCO_3$ mineralization has also been proposed as a new tool in the conservation of monumental calcareous stones.
In the Lab where I have been completing my PhD research program we are studying the genes involved in calcium carbonate biomineralization in *Bacillus subtilis* [7, 74]. *B. subtilis* is a model laboratory bacterium which can produce calcite precipitates on suitable media supplemented with a calcium source (B4 medium,[19]). The genes involved in this process are organized in a single transcriptional unit (*lcfA* operon, figures 3.1 and 4.1) which was firstly described by Wipat et al. 1996 [116] as a group of genes related to fatty acids β-oxidation. Despite the lack of knowledge about the molecular links between these genes and the biomineralization process, the mechanism seems specific and genetically controlled. This premise introduces the aims of my PhD research project: to understand how these genes participate or contribute to calcite precipitation (*Aim 1-2*) and to study their expression along with mineralization (*Aim 3*) in *B. subtilis*. The main goal of the current investigation is to understand the mechanism(s) through which the genes of the *lcfA* operon are involved in biomineralization. Results obtained could be helpful to implement proposed applications of $CaCO_3$ mineralization by bacteria.

AIM 1: *Defining the minimal set of lcfA operon genes indispensable for biomineralization in B. subtilis*

The *B. subtilis lcfA* operon involved in biomineralization contains five genes (*lcf*A, *ysi*A, *ysi*B, *etf*B and *etf*A). Five mutants carrying each of the five inactivated genes were constructed in our Lab [7] and all of these mutants, with the exception of that mutated in *lcf*A, were unable to form calcite crystals on B4 solid precipitation medium. We asked which are the *lcfA* operon genes indispensable for $CaCO_3$ biomineralization. A model has been proposed by *Barabesi et al., 2007* in which the *etf*A gene is essential for $CaCO_3$ precipitation in *B. subtilis*.
We planned to perform overexpression, complementation and functional analysis of *etfA* to unveil its meaning in calcium carbonate precipitation.

We have established that overexpression of *etfA* results in a rapid alcalinization of the media, while the *etfA* mutant acidifies the medium, hampering $CaCO_3$ deposition. This rapid alcaline shift leads to a premature development of calcite precipitates (3-4 days), compared to those of the wild type (WT) *B. subtilis* 168 (6-8 days). These observations suggest that the expression of *etfA* is essential to create the physiological conditions required for biomineralization. At the moment we are studying the ability of *etfA* to restore the calcification phenotype in strains where the operon transcription has been knocked-out, and the effects on calcite production along with pH.

AIM 2: *How is the function of these genes linked to biomineralization ?*

Many putative functions can be assigned to genes of the *lcfA* operon in bacteria, often linked to fatty acids metabolism and with different functional significance [53]. Furthermore the function of these genes are still based on *in silico* prediction with the sole exception of *ysiA* [75]. We have addressed this question guiding functional studies of these genes by purifying their products and analyzing their biochemical properties. Previous observations (*see* AIM 1) suggest that *etfA* expression is essential to create the proper conditions for biomineralization, but how this is achieved is poorly understood. The *etfA* gene encodes a putative α-subunit of prokaryotic heterodimeric flavoproteins involved in electron transport during fatty acid metabolism, and it functions as an hetero-dimer with the β-subunit encoded by the *etfB* gene from the *lcfA* operon . Thus we have planned to validate the function of EtfA through expression and purification in *E. coli*. In the framework of this functional studies we conducted protein-sequence analysis to search for conserved domains. Interestingly, we found that *etfA* has a putative calcium binding domain (CBD). Since calcium has been reported to regulate many processes in bacteria, we hypothesized that the domain found in *etfA* can regulate either the function or contribute to biomineralization. Therefore we have planned to characterize the function along with the binding ability of the newly identified CBD. At the moment, in a similar manner, we have characterized the *ysiB* gene, now *fadB*, demonstrating that its product acts as an enoyl-CoA hydratase.

AIM 3: *Exploring the expression of B. subtilis gene required for calcium carbonate precipitation*

Although the expression of genes belonging to the *lcf* A operon seems to be required for calcium carbonate precipitation [74, 7], the timing of their activation and expression during biomineralization are poorly understood. In *B. subtilis* precipitation of calcium carbonate occurs within 6-8 days of growth on solid B4 media. As far as we know biomineralization seems to be influenced by several environmental parameters (as pH, temperature, nutrients and calcium availability) and metabolic activity [52, 44, 17]. We investigated their expression by constructing *promoter::lacZ* transcriptional fusions. The resulted strains were used to measure and monitor gene expression during biomineralization and how this is influenced by perturbation of the environmental parameters known to affect the process. Through these experiments we have established that the expression of these genes occurs after 2-4 days of growth on solid B4, and could be affected by temperature variations and calcium availability.

COMPARATIVE GENOMICS OF VirR REGULON IN *Clostridium perfringens* STRAINS

During the PhD I had the opportunity to extend my interest to bioinformatics, learning and taking advantage of its instruments through a collaboration with Dr. Matteo Brilli (LBBE, Universit Lyon 1, FR). This collaboration resulted in a comparative analysis of the VirR regulon in pathogenic strains of *Clostridium perfringens*.
Clostridium perfringens, a Gram-positive anaerobic bacterium, causes severe diseases in both animals and humans, that are generally due to the production of powerful extracellular toxins. The main molecular trigger of toxin production is mediated by VirR, which participates in a two-component signal transduction system, and positively regulates the expression of toxin genes. Despite their high phenotypic variability in toxin production, only few targets are known for this regulator (mainly in one strain, *C. perf.* 13). Therefore, defining the genes belonging to the direct VirR regulon in closely related genomes, is important to understand the evolutionary patterns at the basis of the divergence between these strains and to develop effective strategies to counteract *C. perfringens* infections. We have implemented a two step computational strategy to infer from genome se-

quences the genes putatively regulated by VirR and then associated to pathogenesis. Our results showed the presence of a core (*evolutionary conserved*) and an accessory (*strain-specific*) VirR regulon in *C. perfringens*. Results obtained showed that the VirR regulon is variable in different *C. perfringens* strains with 4 genes controlled in all but one strains, and most genes controlled in one or two strains only. Interestingly we found that the type E strain JGS1987 has the largest predicted regulon with as many as 10 VirR targets not found in the other genomes. Within such regulon, several genes code for proteins linked to pathogenesis. This suggest that the relatively large predicted regulon for this strain could be responsible for its peculiar pathogenesis. Our prediction could help to better define strategies for pathogenicity studies and control in *C. perfringens*, and can be useful for wet-lab researchers working on less characterized *C. perfringens* strains that can thus design focused experiments reducing the search space of their experiments and increasing the probability of characterizing positive targets with less efforts.

3

ROLE OF *etfA* IN *Bacillus subtilis* BIOMINERALIZATION

3.1 BACKGROUND

Perito et al. 2000 [89] identified a gene cluster involved in calcium carbonate biomineralization in *Bacillus subtilis* strain 168. The *lcfA* cluster contains five genes and was originally described by Wipat et al. 1996 [116] as a group of genes involved in fatty-acids metabolism, figure 3.1. Five mutants carrying each of the five inactivated genes were constructed and all of these mutants, with the exception of that mutated in *lcfA*, were unable to form calcite crystals on B4 solid medium [7].

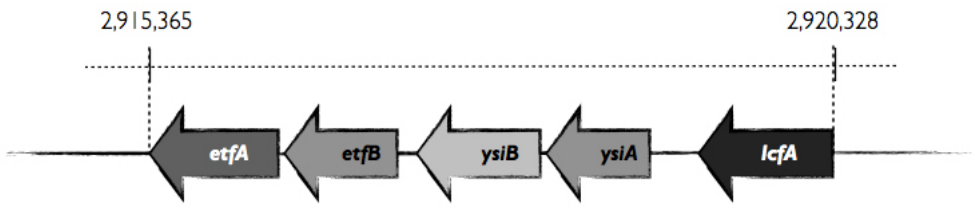

Figure 3.1: Schematics of the genetic organization of *lcfA operon*

The insertional mutagenesis experiments have suggested that all genes of the *lcfA* operon, with the exception of *lcfA*, are possibly involved in biomineralization but a polar effect on the downstream genes of the operon should not be excluded. Thus *B. subtilis* strains were constructed to evaluate this possibility, with an IPTG inducible promoter upstream of each one of the four genes. The growth and precipitation abilities of the four strains

cultivated on B4 supplemented with 1mM IPTG were compared to their counterparts grown on B4. In the absence of IPTG all strains are unable to produce calcite precipitates, while induction of the promoter upstream all the genes resulted in the formation of crystals. Thus the results reported in the work of Barabesi et al.,2007 [7] strongly suggested that the last gene, *etfA*, is essential for biomineralization.
However the physiology and the genetics behind this process is still unknown. Recently Marvasi et al. 2010 [74] through a comparative study involving the *B. subtilis etfA* mutant FBC5 have investigated the physiology of calcite formation and whether the biomineralization is exclusive of solid coltures. Growth comparison between the WT 168 strain and FBC5 in liquid coltures revealed no major differences in the growth kinetics, and change in pH, while a delay of at least 7-8 days in calcite precipitation were observed for FBC5 with respect to the WT strain 168. Indeed, significant pH variations, in terms of acidification of the medium, were noticed during FBC5 growth on solid B4 medium, figure 3.2. The impairment of calcite precipitation was only related to solid medium probably due to the particular spatial microenvironmental condition around the streak/growth and the decrease in pH during FBC5 growth was the main responsible for the lack in $CaCO_3$ precipitation. Calcite precipitation can be restored in FBC5 when the strain is incubated under alkaline conditions or using media bufferd at different pH values. These data suggested that precipitation of calcium carbonate is regulated by local pH, a similar behavior was reported for $CaCO_3$ precipitation by heterotrophic bacteria [93, 35]. It has been proposed that the pH decrease during FBC5 growth resulted from H^+ production exceeding that of OH^-, with the excess of H^+ arising from a misregulation of the EtfA pathway [74]. The *etfA* gene encodes the α-subunit of a prokaryotic heterodimeric flavoprotein, which could function in concert with the β-subunit EtfB. In *Clostridium acetobutylicum* the two proteins EtfA and EtfB are involved in the NADH oxidation pathway during butyrril-CoA biosinthesys, and share 64% and 57% homology with those of *B. subtilis*. It has been hypothesized that these proteins can act in a similar manner in *B. subtilis* as in *C. acetobutylicum*, supported by the fact that misregulation of the EtfA pathway resulted in increase NADH accumulation in FBC5. It has also been proposed that a possible link between NADH accumulation and proton extrusion in *B. subtilis* exists. NADH dehydrogenase in *B. subtilis* is involved in transmembrane proton pumping and catalyze the translocation of electron from NADH to metaquinone, thereby generating a proton-motive force, which can be utilized by the cell.
Based on the model proposed in *Barabesi et al. 2007* [7] & *Marvasi et al.*

2010 [74], in which the *etfA* gene is essential for $CaCO_3$ precipitation in *B. subtilis*, we performed overexpression, complementation and functional analysis of *etfA* to unveil its role in $CaCO_3$ precipitation.

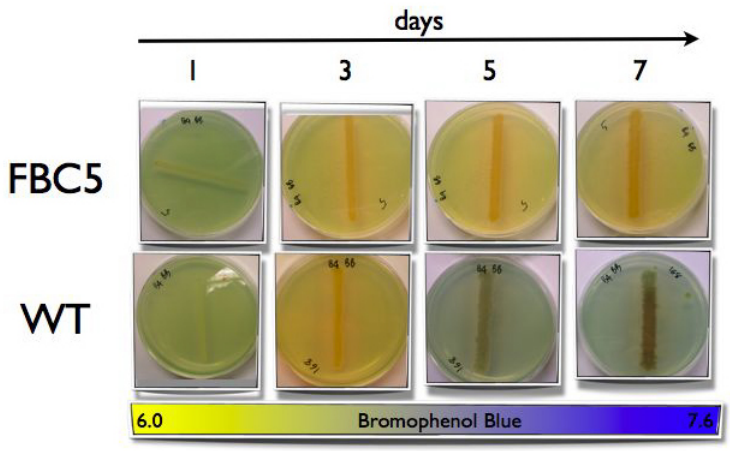

Figure 3.2: pH variations during growth of FBC5 mutant compared to WT strain 168, B4 medium supplemented with Bromothymol Blue

3.2 Experimental Procedures

Bacterial strains, plasmids, and growth conditions

The strains used in this studies are listed in Table 3.1. *B. subtilis* coltures were routinely grown in Nutrient Agar and B4 media (0.4% yeast extract , 0.5% glucose, 0.25% Ca-Acetate, 1.5% Agar in solid preparations, *Bouquet et al., 1973* [19], adjusted to pH 8.0 with 5M sodium hydroxide (NaOH)). *Escherichia coli* coltures were routinely grown either in LB solid and liquid medium for the construction of plasmids (*see Appendix Material and Methods for details*).

Cloning etfA under the control of P_{spac}-promoter

B. subtilis 168 genomic DNA was extracted as described by Marmur [73]. PCR technique was used for *etfA* isolation, PCR reaction was carried out using *pfu* Taq polymerase (Invitrogen). The PCR was performed in a volume of 50 μl, the reaction conditions were 95°C for 5 min, 30 times (95°C

Table 3.1: Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype	Source or reference
E. coli strains		
JM109	*endA1, glnV44, thi-1, relA1, gyrA96, recA1, mcrB*$^+$, Δ*(lac-proAB), e14*$^-$ [F′ *traD36 proAB*$^+$ *lacI*$_q$ *lacZ*Δ*M15*] *hsdR17(rK*$^-$ *mK*$^+$*)*	Promega Co.
JM109-*pDG148*	pDG148, amp^R	This study
JM109-*clone 10*	pDG148-*etfA*, amp^R	This study
B. subtilis strains		
168	*trpC2*	Lab. stock
FBC2	*trpC2, pJM103::ysiA, chf*^{R}	[7]
FBC5	*trpC2, pJM103::etfA, chf*^{R}	[7]
FBC15	*trpC2, pDG148, kan*^{R}	This study
FBC16	*trpC2, pDG148-etfA, kan*^{R}	This study
FBC17	*trpC2, pJM103::etfA, pDG148, chf*^{R}*, kan*^{R}	This study
FBC18	*trpC2, pJM103::etfA, pDG148-etfA, chf*^{R}*, kan*^{R}	This study
FBC19	*trpC2, pJM103::ysiA, pDG148-etfA, chf*^{R}*, kan*^{R}	This study
Plasmids		
pDG148	Vector for IPTG-inducible gene expression using the P$_{spac}$ promoter, kan^R, amp^R	[88]
pDG148-*etf*A	pDG148 derivative for inducible expression of *etf*A, kan^R, amp^R	This study

for 40 sec; 50°C for 40 sec; 68°C for 2 min) and 68°C for 10 min. Oligonucleotides used for the isolation of the *etfA* gene are listed in Table 3.2.

Shuttle vector containing the *etfA* gene was constructed by cloning the DNA fragment into the MCS of pDG148. Both fragments and pDG148 were digested with *Sal*I restriction endonuclease. Linearized pDG148 have been dephosphorylated by Shrimp alkaline phosphatase in order to prevent religation of the linearized vector. One unit of shrimp alkaline phosphatase was used for dephosphorylation of 50 ng linearized vector in dephosphorylation buffer (50 mM Tris-HCl, 0.1 mM EDTA pH 8.5) at 37°C for 60 min. The shrimp alkaline phosphatase was inactivated by heating at 65°C for 15 min. Then dephosphorilated vector and digested PCR fragment were ligated to yield pDG148-*etfA*. The vector was constructed in *E. coli* JM109 grown in Luria-Bertani (LB) medium at 37°C. Transformants were selected by including 100 μg mL^{-1} ampicillin in LB agar plates. Constructs were checked by PCR and DNA sequence analyses.

Wild-type and mutant *B. subtilis* strains were transformed with pDG148-*etfA*, and pDG148 as control, to yield the kanamycin series of strains listed in Table 3.1. Successful transformant were selected for by resistance against kanamycin, and confirmed by plasmid extraction and PCR using the forward and reverse primer matching the region upstream and downstream the MCS of pDG148, respectively Table 3.2.

Table 3.2: Oligonucleotides primer used for the isolation of *etfA* gene and to check the correctness of the insertion; Primers E11(f)/E9(r) were used for the amplification of the *etfA* gene, D7(f)/E9(r) for the amplification of the *etfBA* genes. (C) *Sal*I

Oligonucleotide primer name	5′-3′ sequence	Description
E11(f)	ACGCGTCGACCTTGGACATTCAAAG	(C)
E9(r)	ACGCGTCGACCGTACATGTCAACGAC	(C)
D7(f)-*Sal*I	ACGCGTCGACAGGGGATATGATCATG	(C)
pDG148-Fw	ACAAGGTGTGGCATAATGTG	
pDG148-Rev	TCCCTACAGTGTTATGGC	

pDG148 *shuttle vector for inducible gene expression*

Vector pDG148, figure 3.3, possess the replicon of pBR322 and the β-lactamase gene amp^R for the replication and ampicillin selection in *E. coli* (Stragier, et al., 1988). Moreover as a shuttle vector, pDG148 possess the replicon from pUB110 for multiplication in *B. subtilis* (McKenzie, et al., 1987) and also the *phl*^{R} and *kan*^{R} genes of pUB110 which permit a selection by neomycin and/or kanamycin. The presence of the P_{spac}-promoter with associated Lac-operator and the *lacI* encoding Lac-repressors from *E. coli* under control of the penicillinase promoter P_{pen} of *B. licheniformis* allowed the IPTG-induction expression of a promoterless genes. The multiple cloning site (MCS) *Hind*III-*Sal*I-*Sph*I allowed to clone interested genes into the vector under the control of the P_{spac}-promoter, figure 3.3.

Determination of pH development and crystal formation

pH changes of *B. subtilis* strains grown on B4 plates were monitored using two pH indicators (Sigma Co.): Bromothymol Blue (BB, pH range 6.0-7.6) and Phenol Red (PR, pH range 6.4-8.2). BB and PR stock solutions (20mg mL^{-1} in 0.1 N NaOH) were added to the B4 medium prior autoclaving at a final concentration of 0.05% (v/v). Plates were inoculated by streaking a 4-cm line on the center of the plate and incubated at 37°C for 21 days or until crystal formation. The experiments were carried out in triplicate, controls consisted in uninoculated B4 medium either with PR or BB pH indicator. Growth, production of carbonates crystals and pH changes were monitored collecting photos till the end of the experiments. To monitor

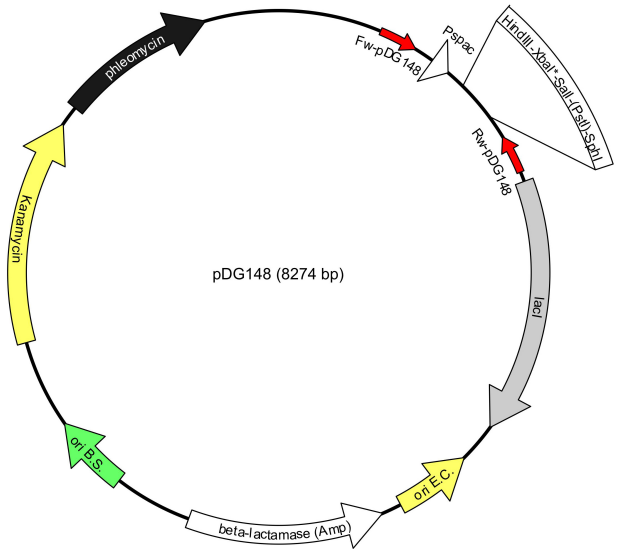

Figure 3.3: Physical map of shuttle-vector pDG148, red arrows on the map mark the annealing site of the forward and reverse primer of pDG148, Table 3.2.

pH change and calcite formation in liquid coltures, 500 mL bottles were inoculated with *B. subtilis* strains, OD_{600nm}= 0.2-0.3, and then incubated at 37°C without shaking for up to a month. pH changes of cultures were monitored by a pH meter (MicropH 2000 CRISON). A 2 mL aliquots of each coltures were used for each measurement.

3.3 RESULTS

Since the *etfA* gene seems to be essential to biomineralization and responsible for the impairment in calcite precipitation, we have planned to study the effect of EtfA overexpression on calcite $CaCO_3$ precipitation through the construction of strain FBC16, in which *etfA* is under the control of the IPTG inducible promoter P_{spac} on the multicopy pDG148 vector. As control in the following experiments we have used *B. subtilis* strains transformed with an empty pDG148, strains FBC15 ad FBC17.

3.3.1 *Overexpression of etfA affects the kinetics of calcite formation in B. subtilis*

It has been reported in Marvasi et al., 2010 [74] that the lack of *etfA* is probably responsible for the pH decrease observed in FBC5, due to the exceeding H^+ extrusion compared to that of WT strain 168, figure 3.2. To investigate whether the pH changes during the growth are associated with calcite production, strains FBC16, WT strain 168, and FBC15 were grown on B4 solid media supplemented with pH indicators as Phenol Red (PR) and Bromothimol Blue (BB) either supplemented or not with 1mM IPTG (*see Determination of pH development and crystal formation for details*).
After three days of growth FBC16 exhibited basification of the medium when cultivated in presence of 1mM IPTG, which leads to a premature (usually 2-3 days before WT and controls) production of calcite crystals compared to those of the WT strain 168 and FBC15, figure 3.6. In contrast no major differences were observed for the WT strain 168 and FBC15 in pH variations during growth in B4 medium supplemented with 1mM IPTG, figure 3.6. During the growth of FBC16 strain in B4 medium supplemented with 1mM IPTG we observed that the basification of the medium starts all around the boundary of the streak, then this halo increase in size and extend its bound to the entire plate, figure 3.6. Interestingly calcite crystals begin to appear on the border of the streak where the growth rate is higher, and the "basification" halo begin to extend, figure 3.4. During the whole incubation period we assist to a production of calcite crystals which cover the entire streak and are smaller in size, in contrast to those observed in controls where calcite crystals are randomly distributed and bigger in size.

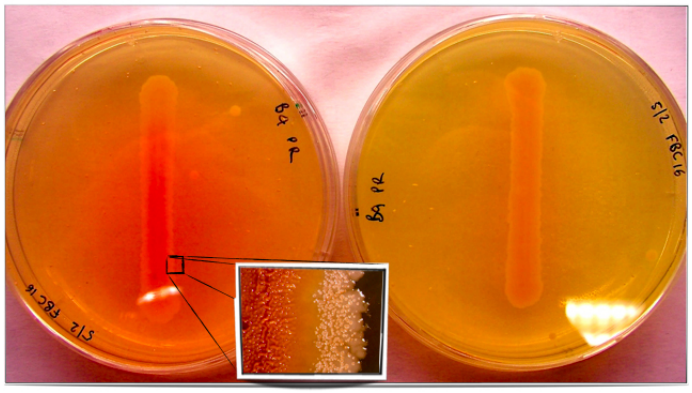

Figure 3.4: Comparison of FBC16 ± 1mM IPTG at 5 days of growth on B4 supplemented with 0.05% PR, to notice the "alkalynization halo", and calcite precipitates on the bound on FBC16 (1mM IPTG) *left plate*.

Although any significant difference during the first week of growth in liquid B4 media, in terms of pH changes among FBC5 and WT was reported [74], we asked if the overexpression of *etfA* in FBC16, in contrast to that observed for FBC5, could significantly modify pH and calcite precipitation as seen for solid coltures. Therefore we have compared the pH changes and crystal formation for FBC16, 168 grown in liquid B4 medium, figure 3.5. Changes in coltures pH were routinely monitored as described in *Experimental procedures*. No major differences were observed, in terms of changes in pH, when FBC16 coltures either supplemented or not with 1 mM IPTG, were compared, figure 3.5. On the converse significant differences can be found by comparing pH changes of WT to those of FBC16 ± 1mM IPTG. The pH increment toward alkaline values reported for strain 168 occurs 7-10 days in advance of FBC16, figure 3.5.

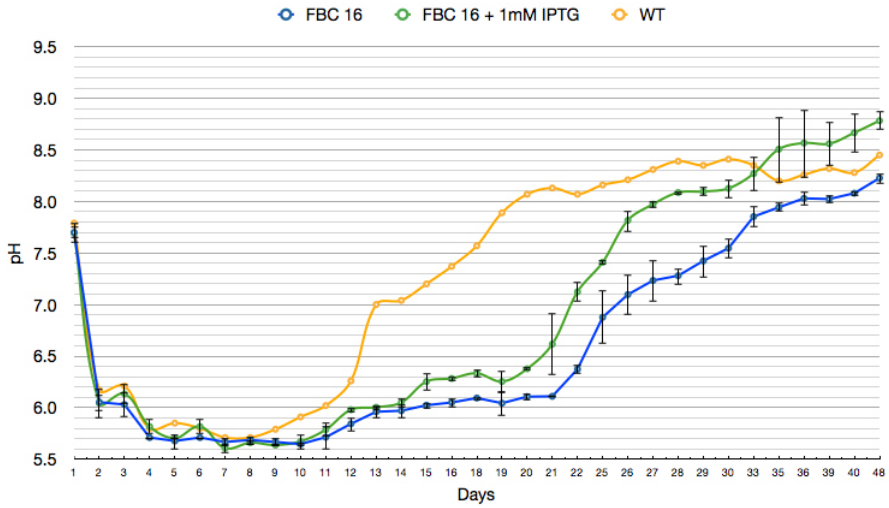

Figure 3.5: pH changes of strain FBC16 ± 1 mM IPTG, and WT strain 168 grown in B4 liquid medium

3.3.2 *Defining the minimal set of genes required for $CaCO_3$ precipitation in B. subtilis*

As noted above (*see previous section*) *etfA* expression is essential to create the proper condition for biomineralization, but how this is achieved is poorly understood. Moreover we still do not know if EtfA works alone to promote calcite precipitation.
The *etfA* gene encodes a putative α-subunit of prokaryotic heterodimeric flavoproteins involved in electron transport during fatty acid metabolism, and it would function as an hetero-dimer with the β-subunit encoded by the *etfB* gene of the *lcfA* operon. Thus, we asked if the sole *etfA* was sufficient or on the contrary which is the minimal set of genes required for calcite precipitation in *B. subtilis*. We have addressed this question guiding functional studies in strains FBC18 and FBC19, Table 5.1.
Strain FBC18, resulted from the transformation of strain FBC5 with plasmid pDG148-*etfA*, is able to produce calcite crystals only when 1mM IPTG was added to B4 medium. Changes in pH during calcite production were monitored by adding Phenol Red to the medium. FBC18 has the same behaviour of the WT strain 168 when cultivated under 1mM IPTG induction, sharing the same timing of alkalynization as well as biocalcification, while without IPTG induction, acidification of the medium was observed as in FBC5, as well as absence of precipitates, Figure 3.7. The current finding in-

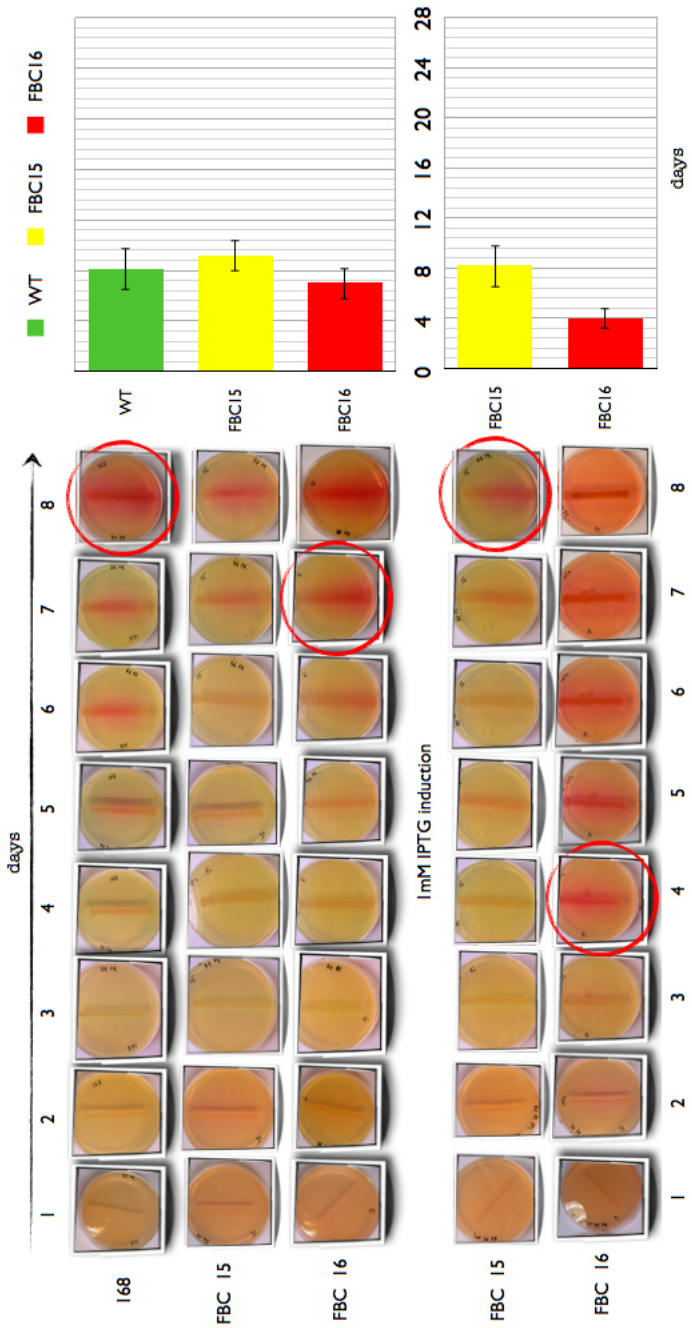

Figure 3.6: pH variations during growth of strains FBC15 and FBC16 compared to WT strain 168, B4 medium supplemented with 0.05% PR, red circles mark the day in which calcite precipitation started. Bar graphs indicate the timing of calcite production between strains used in the experiment, bars show standard deviation among 4 independent replicas.

dicates that *etfA* is able to restore the bio-calcification phenotype in strain FBC5, confirming that the function performed by *etfA* is an essential requirement for biomineralization.
Since EtfA would function as an hetero-dimer with EtfB, we asked if EtfA is able to restore the calcification phenotype even in the absence of EtfB and/or other proteins encoded by the *lcfA* operon. Thus we constructed strain FBC19 which resulted from the transformation of strain FBC2[1] with the pDG148-*etfA*, Table 5.1. First of all we have tested its capabilities to restore the bio-calcification phenotype compared to those of FBC18, WT strain 168 and of FBC2. FBC19 showed poor production of calcite precipitates after two or more weeks of growth under 1mM IPTG induction, figure 3.7. We have tested this strain also in B4 medium supplemented with Phenol Red to investigate the pH changes during growth. Results obtained showed that strain FBC19 under IPTG induction are responsible for a small increment of the pH towards alkaline values compared to that of FBC18, Figure 3.7.

[1]FBC2 mutant carries the interruption of the *ysiA* gene which in turn hamper transcription from P_{ysiA}.

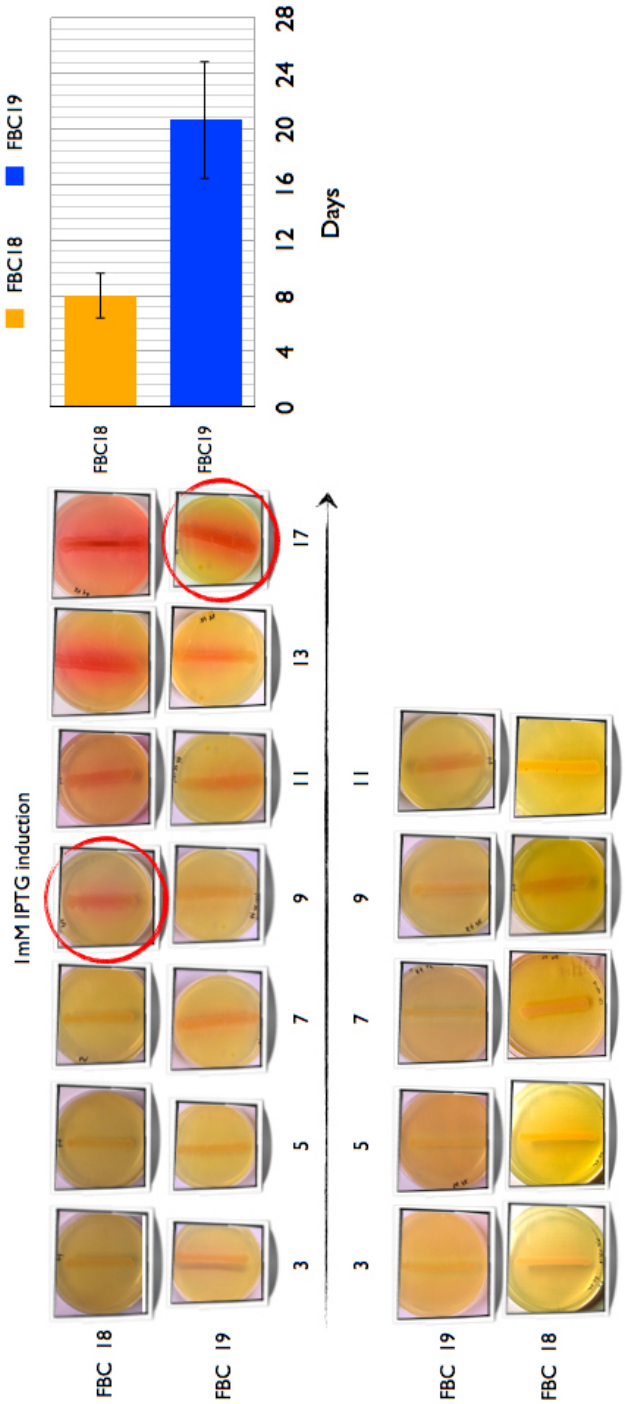

Figure 3.7: pH variations during growth of FBC18 and FBC19 cultivated on B4 medium supplemented with 0.05% PR, red circles mark the day in which calcite precipitation has started. Bar graphs indicate the timing of calcite production between strains used in the experiment, bars show standard deviation among 4 independent replicas.

3.4 CONCLUSIONS

The EtfA protein encoded by *etfA* gene is essential for calcite precipitation on solid media [7]. Recent studies, focused on the physiology behind the lack of calcite precipitation affecting the EtfA-disruptant FBC5, have revealed that the pH decreasing mechanism is the main responsible for the impairment of calcite precipitation, supporting its direct involvement in *B. subtilis* biomineralization [74].
In the current study we have investigated how calcite biomineralization is affected by the *etfA* overexpression. In our experiments the overexpression of *etfA* in strain FBC16 resulted in a precocious calcite production, usually 2-3 days before the parental strain 168. Consequently, we have monitored the pH of solid cultures to study if overexpression of *etfA* resulted in pH changes of the culture media as seen in its counterpart FBC5, figure 3.2, [74]. We found that strain FBC16 leads to an early alkalinyzation of the medium figure 3.6, which is, in almost all the conducted experiments, coincident with the production of $CaCO_3$ precipitates. This findings account for the local pH regulation as the main mechanism responsible for calcite precipitation in *B. subtilis* on solid media. In liquid cultures, we did not report any significant difference for pH development between FBC16 grown in liquid B4 either supplemented or not with 1mM IPTG, supporting the possibility that precipitation was only related on solid media due to its particular spatial micro-environmental condition.These data suggested that overexpression of EtfA is the main responsible of the effects observed in pH and calcite precipitation on solid media.
Further investigation have explored if EtfA is able to restore calcite precipitation even in the absence of EtfB and/or other proteins encoded by the *lcfA* operon. In the mutant background of strain FBC5 expression of *etfA* in strain FBC18, under IPTG induction, was able to restore the calcite production, while without induction precipitation were completely inhibited, and thus proving that the function performed by *etfA* is an essential prerequisite for calcite precipitation. In addition we have investigated if/how the EtfA expression in the mutant background affects pH. Alkalinyzation of the medium was observed in induced FBC18, with a lower intensity than in FBC16 and WT, figure 3.7 and 3.6. However in FBC18 as in FBC16 and controls, calcite precipitation is in almost the experiments coincident with a pH shift towards alklynization. Once established that the expression of the sole *etfA* restores calcite precipitation in FBC5 mutant background, we have investigated if the phenotype observed for FBC18 could be achieved with the expression of EtfA in the FBC2 mutant background, by strain FBC19. The expression of EtfA under IPTG induction in FBC19 resulted in

calcite production after two or more weeks of incubation, in addition we observed a significant delay in pH changes towards alkaline values with respect to FBC18 and FBC16 as well, figure 3.7. From this findings arises the possibility that other functions, together with EtfA, are required for the complete restoration of the calcite precipitation phenotype in *B. subtilis*.

4

EXPLORING THE EXPRESSION OF *Bacillus subtilis* GENES REQUIRED FOR CALCIUM CARBONATE PRECIPITATION

4.1 BACKGROUND

The *lcfA, ysiA, ysiB, etfB* and *etfA* genes are clustered in this order, figure 4.1. The product of these genes are known to be involved in β-oxidation of fatty acids in *E. coli* and homologous could be well assigned to different enzymatic steps in this pathway. RT-PCR analyses, in *B. subtilis* cells grown on B4 medium, has showed that the region from *lcfA* gene to the *etfA* gene is expressed as a single transcriptional unit [7], but it did not exclude that additional transcripts could exist, starting from internal *lcfA* promoters.

The *ysiA* product is a member of the TetR family of bacterial helix-turn-helix transcription regulatory proteins. YsiA was supposed to be the functional homolog of *E. coli* FadR, a global transcriptional regulator of fatty acid metabolism [53], which negatively regulates more than 10 genes, the majority of which most likely partecipate in fatty acids β-oxidation. As for *E. coli*, *B. subtilis* FadR (YsiA) is supposed to be induced by long chain acyl-CoAs (14-20 C) which interacts with YsiA preventing its binding with the YsiA boxes [75], located immediately upstream the genes regulated by YsiA.
DNA microarray analyses were performed to find candidate YsiA-regulated genes through the comparison of the transcriptomes from cells of the *ysiA* disruptant and parental wild type strain (*B. subtilis* 168) [75]. The analysis revealed candidate YsiA repressed genes within five operons. The YsiA regulon includes 15 genes organized into five operons (*lcfA-ysiA-*

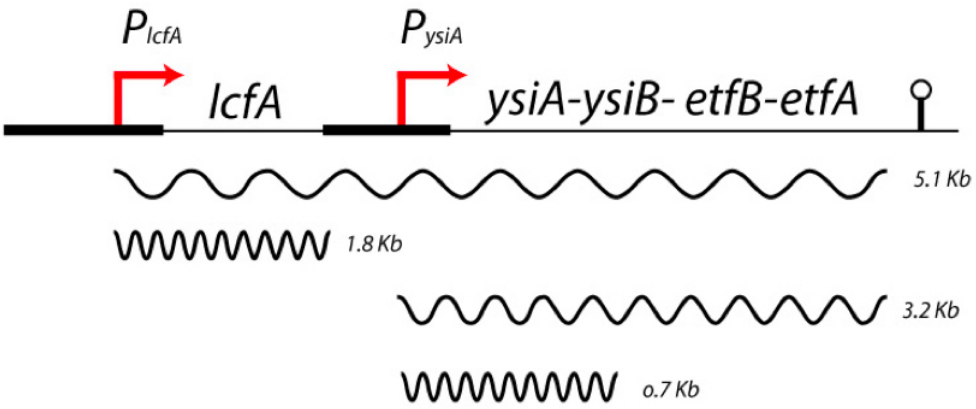

Figure 4.1: Map of the *B. subtilis lcfA* operon. Wavy lines indicates transcript, and their length, produced by P_{lcfA} and P_{ysiA}, [75].

ysiB-etfB-etfA; yusM-L-K-J and ywjF-acdA-rpoE). Their transcripts were analyzed by Northern Blotting [75], figure 4.1. Four transcripts were detected with the *lcfA, ysiA and ysiB* probes, the 1.8Kb and the 0.7Kb seem to contain only the *lcfA* and the *ysiA* genes respectively, whereas the 5.1Kb and the 3.2Kb contain the whole cluster and the genes from *ysiA* to *etfA* respectively, since a ρ-independent terminator was found immediatly downstream *etfA*. Primer extension analysis revealed transcription initiation, and promoter regions immediately upstream the YsiA regulated genes showing the presence of highly similar palindromic sequences very close to the transcription initiation sites, the YsiA boxes.
Barabesi et al. 2007 [7] showed that the LcfA-disruptant, in which the transcription of the 5.1Kb and 1.8Kb transcripts starting from *lcfA*, was shut down, is still able to produce calcite crystals indicating that *lcfA* is not required for precipitation, supporting that the 3.2Kb transcript starting from *ysiA* and containing the genes from *ysiA* to *etfA* is that one essential for biomineralization. The expression of these genes and the timing of their activation during biomineralization have never been correlated. We investigated the expression of the *lcfA* operon by constructing promoter::*lacZ* transcriptional fusions. The resulting strains were used to measure and monitor gene expression during biomineralization and to test if perturbations of environmental parameters known to affect biomineralization act at the transcritptional level.

4.2 EXPERIMENTAL PROCEDURES

Bacterial strains plasmids and growth conditions

Strains used in this study are listed in Table 4.1. Coltures were routinely grown on Nutrient Agar (NA) and B4 medium *Bouquet et al., 1973* [19] adjusted to pH 8.0 with 5M sodium hydroxide (NaOH). When assayed for the expression through the *lacZ* fusion system, cultures were grown on B4 medium supplemented with the appropriate antibiotics and X-Gal (Sigma Co.), in solid preparation, at the final concentration of 80 μg mL^{-1}. Liquid cultures were cultivated in B4 medium supplemented with the appropriate antibiotics in conical flasks at 150 rev min^{-1} into a gyrotory shaker.

Table 4.1: Bacterial strains and plasmids

Strain or plasimd	Genotype		Source or reference
E. coli strains			
JM109	*endA1, glnV44, thi-1, relA1, gyrA96, recA1, mcrB*$^+$, Δ*(lac-proAB), e14*$^-$ [F′ *traD36 proAB*$^+$ *lacI*$_q$ *lacZ*Δ*M15*] *hsdR17(rK*$^-$ *mK*$^+$*)*		Promega Co.
JM109-4A *clone*	pJM116-P$_{lcfA12}$, amp^R		This study
JM109-8B *clone*	pJM116-P$_{lcfA13}$, amp^R		This study
JM109-14C *clone*	pJM116-P$_{ysiA11}$, amp^R		This study
JM109-13D *clone*	pJM116-P$_{etfBA10}$, amp^R		This study
JM109-pJM116 *clone*	pJM116, amp^R		This study
B. subtilis strains			
168	*trpC2*		Lab. stock
FBC10	*trpC2, pJM116-P*$_{etfBA10}$*::amyE, chf*^{R}		This study
FBC11	*trpC2, pJM116-P*$_{ysiA11}$*:: amyE, chf*^{R}		This study
FBC12	*trpC2, pJM116-P*$_{lcfA12}$*:: amyE, chf*^{R}		This study
FBC13	*trpC2, pJM116-P*$_{lcfA13}$*:: amyE, chf*^{R}		This study
FBC14	*trpC2, pJM116:: amyE, chf*^{R}		This study
Plasmids			
pJM116	Integrational vector for *lacZ* fusion construction		[88]
pJM116-P$_{lcfA13}$	(2.920.384-*Kpn*I - 2.920.234-*Bam*HI)-*lacZ*, amp^R, chf^R	A8(f)/A9(r)	This study
pJM116-P$_{lcfA12}$	(2.920.493-*Kpn*I - 2.920.234-*Bam*HI)-*lacZ*, amp^R, chf^R	A7(f)/A9(r)	This study
pJM116-P$_{ysiA11}$	(2.918.780-*Kpn*I - 2.918.470-*Bam*HI)-*lacZ*, amp^R, chf^R	A10(f)/B9(r)	This study
pJM116-P$_{etfBA10}$	(2.917.294-*Kpn*I - 2.917.059-*Bam*HI)-*lacZ*, amp^R, chf^R	C11(f)/D9(r)	This study

Construction of promoter::lacZ transcriptional fusions

B. subtilis 168 genomic DNA was extracted as described by Marmur [73]. PCR technique was used for gene fusion construction, PCR reaction was carried out using *pfu* Taq polymerase (Invitrogen). The PCR was performed in a volume of 50 μl, the reaction conditions were 95°C for 5 min, 30 times (95°C for 15 sec; 50°C for 30 sec; 68°C for 1 min) and 68°C for 10 min. Oligonucleotides primers used for the isolation of the regulative regions within the *lcfA* operon are listed in Tables 4.2 and 4.1.

Table 4.2: Oligonucleotides primers used for the amplification of the putative promoter region in *lcfA* operon, restriction site are underlined, (A)*Kpn*I, (B)*Bam*HI

Primer name	5′-3′ sequence	
A7(f)	CGGGGTACCGCTTGCGAAAATACTGCTCGC	(A)
A8(f)	CGGGGTACCGCCGTCCATCTGCTATAATG	(A)
A9(r)	GCGGGATCCTCTGTCAGGATGGATTGCAGG	(B)
A10(r)	CGGGGTACCGCGAGATCACGTCTAGCTC	(A)
B9(r)	GCGGGATCCGTAGCCGTTTTCTGCAATGAC	(B)
C11(f)	CGGGGTACCGTGTATTCCTATGAAGGCAGC	(A)
D9(r)	GCGGGATCCCCATTCCGCTCCGTCATC	(B)

Integrational vectors containing the regulative regions of the *lcfA* operon, Tables 4.1 and 4.3, (P_{lcfA12}, P_{lcfA13},P_{ysiA11}, and $P_{etfBA10}$) were constructed by cloning DNA fragments into the MCS of pJM116, figure 4.2. Both fragments and pJM116 were digested with *Kpn*I and *Bam*HI restriction endonucleases and ligated to yield the pJM116-P_{lcfA13}, pJM116-P_{lcfA12}, pJM116-P_{ysiA11} and pJM116-$P_{etfBA10}$, Table 4.1. Recombinant vectors were constructed in *E. coli* JM109 grown in Luria-Bertani (LB) medium at 37°C. Transformants were selected by including 100 μg mL^{-1} ampicillin in LB agar plates. Constructs were checked by PCR and DNA sequence analyses.
Wild-type *B. subtilis* strains were transformed with recombinant plasmids to yield the chloramphenicol series of strains listed in Table 4.1. Successful integrants into the *amyE* locus by homologous recombination were selected for by resistance against chloramphenicol. Integration into the *amyE* locus was confirmed by an amylase negative phenotype of cells plated on LB agar containing 1% soluble starch.

β-Galactosidase assay

β-Galactosidase activity was assayed *qualitatively* on LB agar plates containing the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (80 μg mL^{-1}).
Quantitative assay of β-Galactosidase was performed as follows.
One milliliter of cells were harvested from each culture at each time point (60′), and resuspended in an equal volume of Z buffer (40 mM NaH_2PO_4, 60 mM Na_2HPO_4, 1 mM $MgSO_4$, 10 mM KCl, 38 mM β-mercaptoethanol). To each sample, lysozyme was added to a final concentration of 0.2 mg mL^{-1} and incubated at 30°C for 15 min. Each sample was diluted appropriately in 500 μL of Z buffer and the reaction was started with 100 mL of 4 mg mL^{-1} 2-nitrophenyl-β-D-galactopyranoside (in Z buffer) and stopped with 250 mL of 1 M Na_2CO_3. The OD_{420} of the reaction mixtures was recorded, and the β-galactosidase specific activity was calculated according to the equation: $[OD_{420}/(t \times OD_{600})] \times d \times 1000$ [1].

pJM116 *integrational vector for lacZ fusion construction*

To analyze the expression of *etfA*, as well as of the other genes of the *lcfA* operon, we used pJM116 integrational vector, figure 4.2, for constructing fusions to specific cloned genes. pJM116 *lacZ* integrational fusion vector, carries antibiotic resistance markers selectable either in *B. subtilis* (*cat*) and *E. coli* (*β-lactamase*), and a promoterless copy of the *lacZ* gene preceded by multiple cloning site (MCS), figure 4.2. To study the trascriptional potential of specific DNA fragments of the *lcfA* operon, such fragments were placed in front of *lacZ*. Trascriptional fusions to *lacZ* are placed in the chromosome in the amylase gene (*amyE*), opposite to its direction of transcription, by a double cross-over event[2]. This avoid the problems observed for vectors that integrate by a single cross-over event. We have choosed this approach because it offers several advantages. First, the sequence at the normal chromosomal loci of the gene(s) used to create the fusion(s) remains intact. Second, such construct are highly stable in the chromosome, since excision cannot occur, thus it can be easily transferred from strain to strain by transformation or transduction in order to analyze the behavior of the fusions in different backgrounds.

[1] OD_{420} is read from the reaction mixture, OD_{600} reflects the cell density just before assay, t time of the reaction in minutes, d diluition factor

[2] In order to achieve a double cross over event, recombinant plasmids used to transform *B. subtilis* strains were linearized with *Pst*I endonuclease.

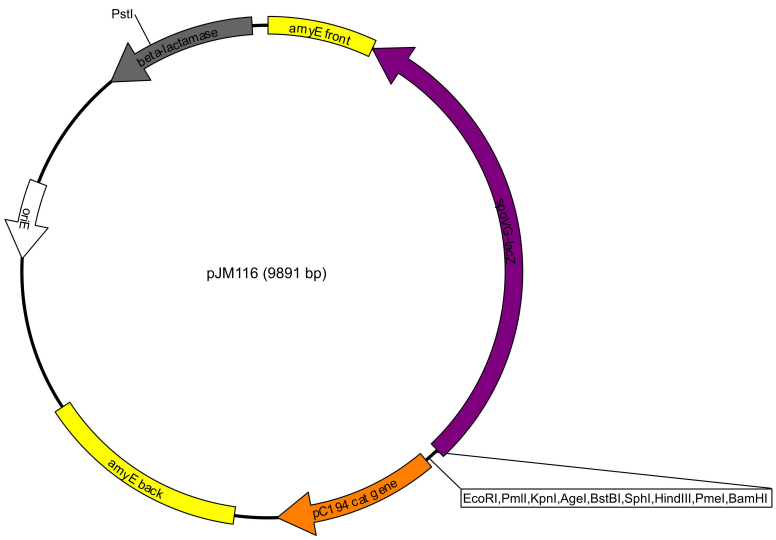

Figure 4.2: Physical map of integrational vector pJM116

4.3 RESULTS

In order to study the expression of these genes and the timing of their activation along with calcite precipitation, we have generated a set of strains harboring promoter::*lacZ* transcriptional fusions, Table 4.1 and 4.3. The *lcfA* operon transcription has been widely studied, and has been demonstrated that is under the control of two promoters, P_{lcfA} the promoter leading to the transcription of the whole cluster, located upstream *lcfA*, and the P_{ysiA} internal promoter. Interestingly the preliminary *in silico*[3] analyses we

[3]Neural Network promoter Prediction, NNPP, is a method that finds eukaryotic and prokaryotic promoters in a DNA sequence. The function of the promoter as a initiator for transcription is one of the most complex processes in molecular biology. It has been shown that multiple functional sites in the primary DNA are involved in the polymerase binding process. These elements, such as the TATA-box and the transcription start site ("Initiator") for eukaryotes, are known to function as binding sites for Polymerase II, transcription factors, and other proteins that are involved in the transcription initiation process. These promoter elements are present in various combinations separated by various distances in the sequence. The basis of the NNPP program is a time-delay neural network [90, 91]. The time-delay network consists mainly of two feature layers, one for recognizing the TATA-box and one for recognizing the "Initiator", which is the region spanning the transcription start site. Both feature layers are combined into one output unit, which gives output scores between 0 and 1.

performed before starting this study, have revealed the presence of a putative promoter region ($P_{etfBA10}$) in the inter-genic region between the *ysiB* and *etfBA* genes, moreover we found an inverted repeated sequence immediatly upstream the P_{lcfA} (P_{lcfA13}) which could account for an additional regulation. Therefore we have included in the following study also these putative newly regulative regions, Tables 4.1 and 4.3.

Table 4.3: Genomic location and strains harboring the regulative regions of the *lcfA operon* used for the construction of the transcriptional fusions. *Score* of the regulative regions cloned obtained from the analysis with NNPP program.

STRAINS	REGULATIVE REGION NAME	GENOMIC LOCATION	***Score***	SOURCE
FBC12	P_{lcfA12}	2.920.493-2.920.234	0.99	[75, 90, 91]
FBC13	P_{lcfA13}	2.920.384-2.920.234	0.34	[75, 90, 91]
FBC11	P_{ysiA11}	2.918.780-2.918.470	0.98	[75, 90, 91]
FBC10	$P_{etfBA10}$	2.917.294-2.917.059	0.91	[75, 90, 91]

To study the transcriptional potential of specific DNA fragments of the *lcfA* operon, Table 4.3, such fragments were cloned in front of the promoter-less copy of *lacZ* carried by pJM116 integrative vector. Transcriptional fusions were placed into a neutral loci (*amyE*) on the chromosome, allowing the sequences at the normal chromosomal loci of the genes used to create the fusions to remain intact. Thus, this transcriptional fusion system would allow us to monitor the expression along with calcite precipitation, and possibly when and where, the transcription of these genes is switched on. However, once activated, we can only qualitatively measure the activation, while changes in the expression during the incubation period on solid media were not easily detectable. Therefore we have decided, to collect photos at 48 h (2 days) of incubation when expression in almost all strains is maximized.

4.3.1 *Influence of temperature, pH and Ca^{2+} availability on the expression of genes involved in biomineralization*

Temperature

Generally, carbonate precipitation by bacteria increases with time and temperature of incubation. Temperature affects not only the amount of precipitation but also crystal quality and morphology [30, 29]. We investigated the effect of temperature on the expression of the promoters of the *lcfA* operon and on biomineralization. Strains FBC10, FBC11, FBC12, FBC13 and FBC14 as control, were cultivated on B4 solid media supplemented with 80 μg mL^{-1} X-Gal and incubated at 30°C, 37°C, 42°C. The coltures were routinely monitored both for *lacZ* expression and calcite production. The images reported in figure 4.3 show that expression of *lacZ* and thus the magnitude of the activation of the promoter increases following the increment of incubation temperature, in all strains except for FBC10 and FBC14 which do not show *lacZ* expression in any condition, even after longer incubation, figure 4.3. As regards the timing of the activation of the *lcfA* operon promoters, we report that expression usually occurs after 24-48 hours depending on the temperature of incubation (*i.e:* 24 h-42°C; 48 h-30-37°C). At 42°C, in particular, expression of *lacZ* was detectable in strains FBC11-13 within the first 24 h of incubation, while at 30-37°C it was detectable at 48 h. Although we can only qualitatively measure the precise moment of activation, we have to notice that expression, in strains FBC11 and FBC13 at 30-37°C, observed at 48 h usually tend toward an apparent slight increment within the following 120 h (5 days), while in FBC12 changes are not of easy detection due to the strenght expression observed within the first 48 h.
Because of the spatial distributed growth of calcite crystals on the surface of the *B. subtilis* biofilm, we would have expected that also the expression of these genes would have followed the same pattern. Our experiments showed that the expression in FBC11-13 strains is distributed over the entire surface of the streak starting from the center to the bound. Calcite was produced by all the strains tested in 6-8 days at 37°C, and 2-4 days at 42°C. We did not observe any calcite precipitates for strains cultivated at 30°C even after 21 or more days of incubation.

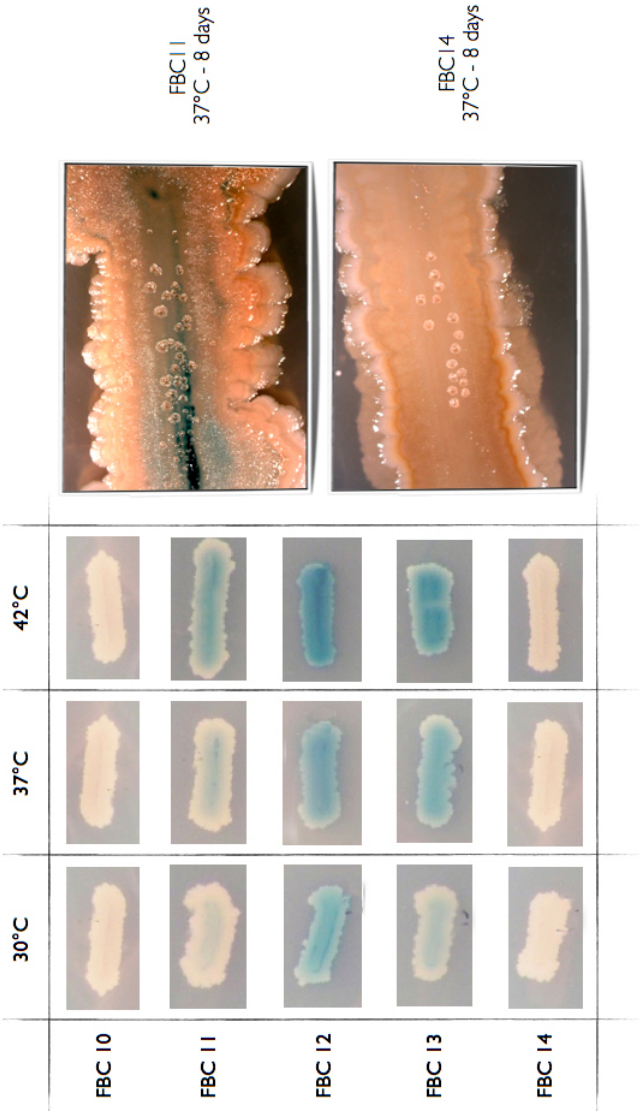

Figure 4.3: Qualitatively assays of β-gal activities of strains FBC10-14 on B4 agar plates supplemented with X-Gal (80 μg mL^{-1}), cultivated at three different temperature at 48 h of incubation.

pH

Changes in pH during growth of *B. subtilis* are thought to be the main mechanism responsible for calcite production. As a matter of fact, using B4 medium buffered at different pH values it is possible to control the biomineralization abilities of the WT strain 168 and FBC5 [7, 74]. Therefore we have tested if the expression of the *lcfA* operon genes as well as the $CaCO_3$ precipitation undergo to significant changes, in strains FBC10-13 and FBC14 as control, cultivated in B4 buffered at different pH values.
Strains harbouring the *lacZ* transcriptional fusions were streaked on B4 buffered media supplemented with X-Gal (80 μg mL^{-1}) and routinely monitored for expression and biomineralization. We have selected two pH values pH 7.7 and pH 8.2, which would inhibit and promote biomineralization respectively.
FBC11-13 when cultivated on buffered pH 7.7 B4 medium (37°C) did not displays any significant difference in the expression when compared to their counterparts grown on pH 8.2 buffered B4, figure 4.4. Activation of the promoter, in all strain except for FBC10,14, usually occurs 48 h after incubation has started, with different strength for each strain, figure 4.4. As regarding the magnitude we have noticed that FBC12 displays the highest level of activation in terms of color development, on the contrary FBC13 displays the lowest, without any significant difference comparing pH 7.7 and pH 8.2 plates. FBC11 behavior, always concerning the magnitude of activation, seems to be placed in the middle between those observed for FBC12 and FBC13, and, again, any difference between pH 7.7 and pH 8.2 was detected, figure 4.4.
Although any significant difference in the expression can be reported for growth on buffered B4 media, we showed that calcite production was affected by the pH of the culture media. Strains FBC10-14 cultivated in pH 7.7 buffered B4 did not produce calcite precipitate, on the converse their counterparts cultivated on pH 8.2 buffered B4 media were able to produce calcite after 2-4 days of incubation[4].
We have tested the possibility that temperature could still influence the expression of these genes when cultivated in buffered B4 media. We tested the expression at two different temperatures (30°C-42°C) and two different pH values (7.7-8.2). Qualitatively β-Gal assays have showed that any significant difference can be observed in the expression in FBC11 and FBC12 strains, apart for those due to temperature, when buffered plates were compared to their unbuffered counterparts at the same temperature, fig-

[4] 2-4 days before the WT strain 168, and FBC10-14 cultivated in unbuffered B4 at the same temperature

ures 4.5 and 4.6. The differences we observed in FBC11 and FBC12 are the same found for their unbuffered replicas except for that of FBC13 between 30°C and 42°C in both pH 7.7 and pH 8.2 plates.

Ca^{2+} availability

Intracellular and extracellular Ca^{2+} concentration must be tightly regulated because of the key role of Ca^{2+} in the regulation of many fundamental processes and its potential harm to cell structures [103]. $CaCO_3$ precipitation has been proposed as a detoxification mechanism to immobilize the excess of dangerous extracellular calcium in bacteria [2]. Considering the importance of the biomineralization process for cell life, it seems unlikely that it would be relied only on aspecific mechanisms in bacteria. It seems more likely to hypothesize more than one bio-precipitation mechanism with a different involvement of bacterial cells [48, 49]. Thus we have investigated if the calcium concentration could influence the expression of the genes involved in $CaCO_3$ precipitation. Strains FBC10-13 and FBC14 as control were cultivated at 37°C on B4 media supplemented with different Ca^{2+} concentration. Our experiments showed that the expression starting from P_{lcfA}, in FBC12,13 is not affected by the Ca^{2+} availability, while in FBC11, used to figure out the expression starting from P_{ysiA}, we observed that expression is more intense at low Ca^{2+} (0.01-0.25%) concentration rather than at higher concentration (0.5-1.0%), figure 4.7 AB. We have observed that in FBC11 rather than in FBC12,13 a magnification of the expression at higher calcium concentration (0.5-1.0%) starting after 4-6 days of incubation, even though calcite have been already precipitated, figure 4.7 C. During these experiments we have also tracked the effect of Ca^{2+} availability concerning calcite precipitation. We noticed that $CaCO_3$ precipitation is inhibited in plates supplemented with Ca-acetate concentration from 0.01% to 0.1%, while at higher Ca-acetate concentration ranging from 0.5% to 1.0% crystals were produced after 2-3 days of incubation at 37°C, figure 4.7 C.

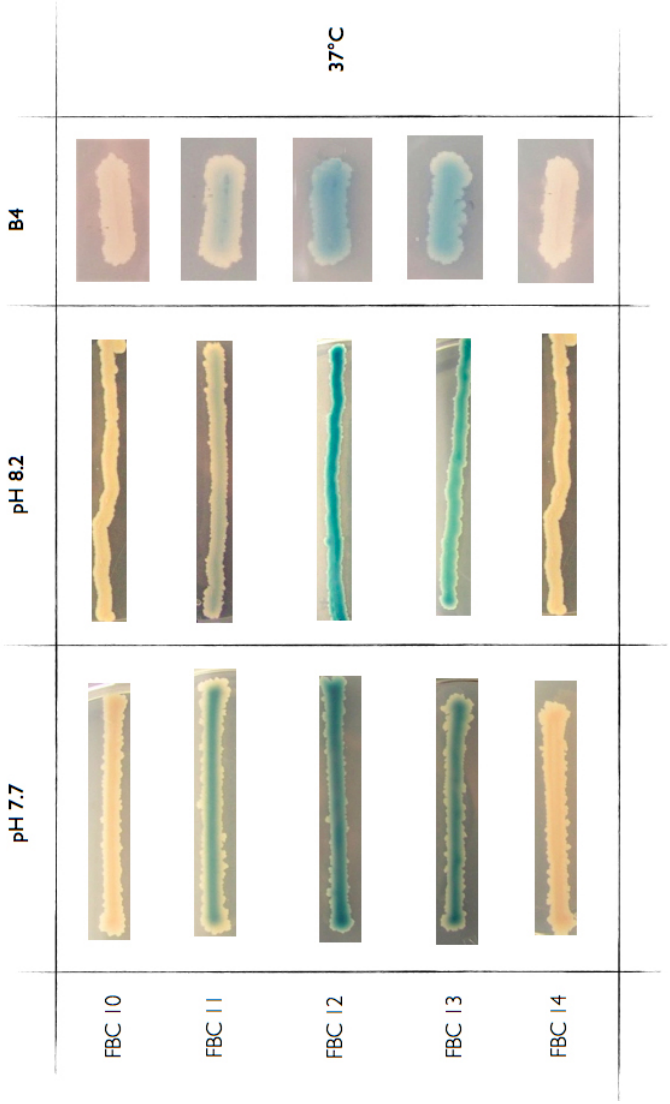

Figure 4.4: Qualitatively assays of β-gal activities of strains FBC10-14 on buffered, pH 7.7 and pH 8.2, compared to unbuffered B4 agar plates cultivated at 37°C. (48 h incubation)

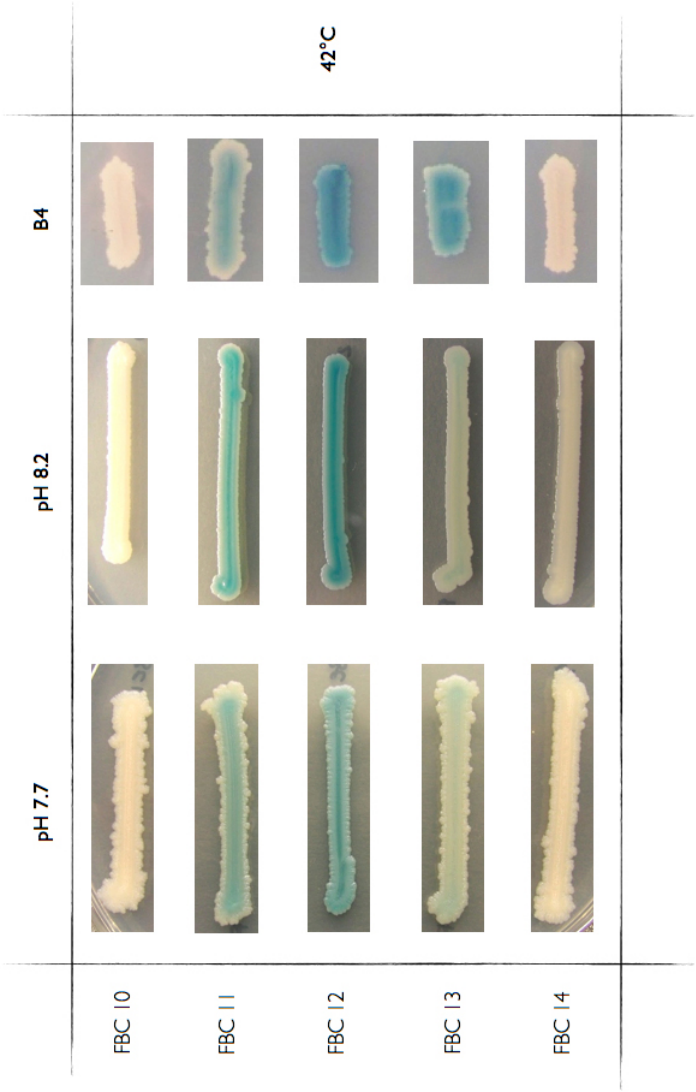

Figure 4.5: Qualitatively assays of β-gal activities of strains FBC10-14 on buffered, pH 7.7 and pH 8.2, compared to unbuffered B4 agar plates cultivated at 42°C. (48 h incubation)

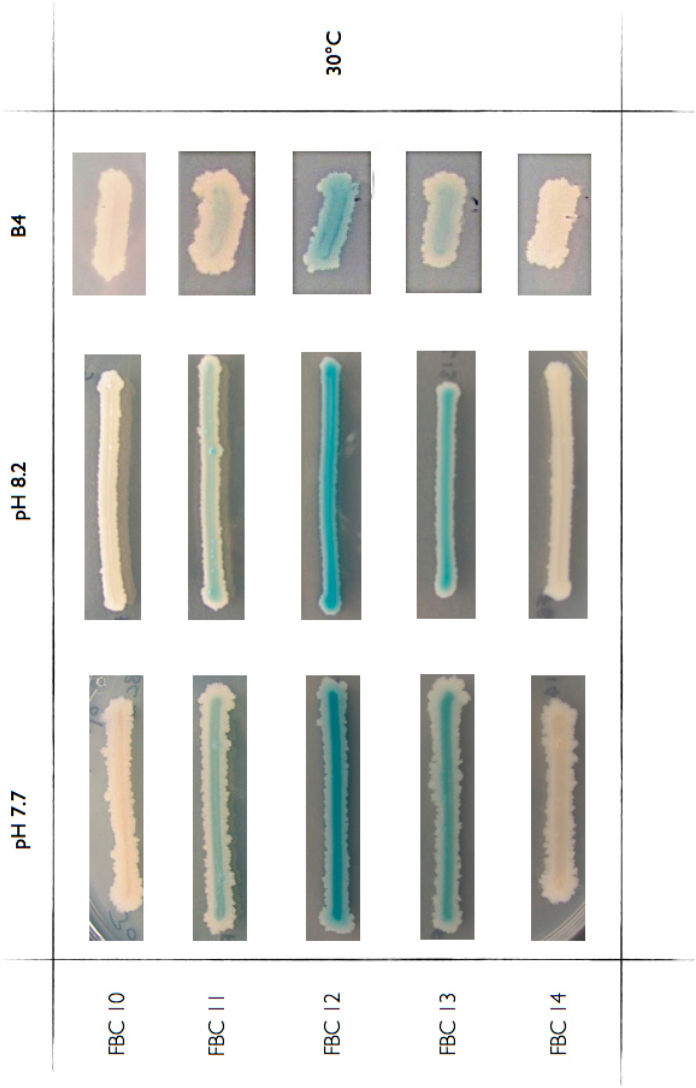

Figure 4.6: Qualitatively assays of β-gal activities of strains FBC10-14 on buffered, pH 7.7 and pH 8.2, compared to unbuffered B4 agar plates cultivated at 30°C. (48 h incubation)

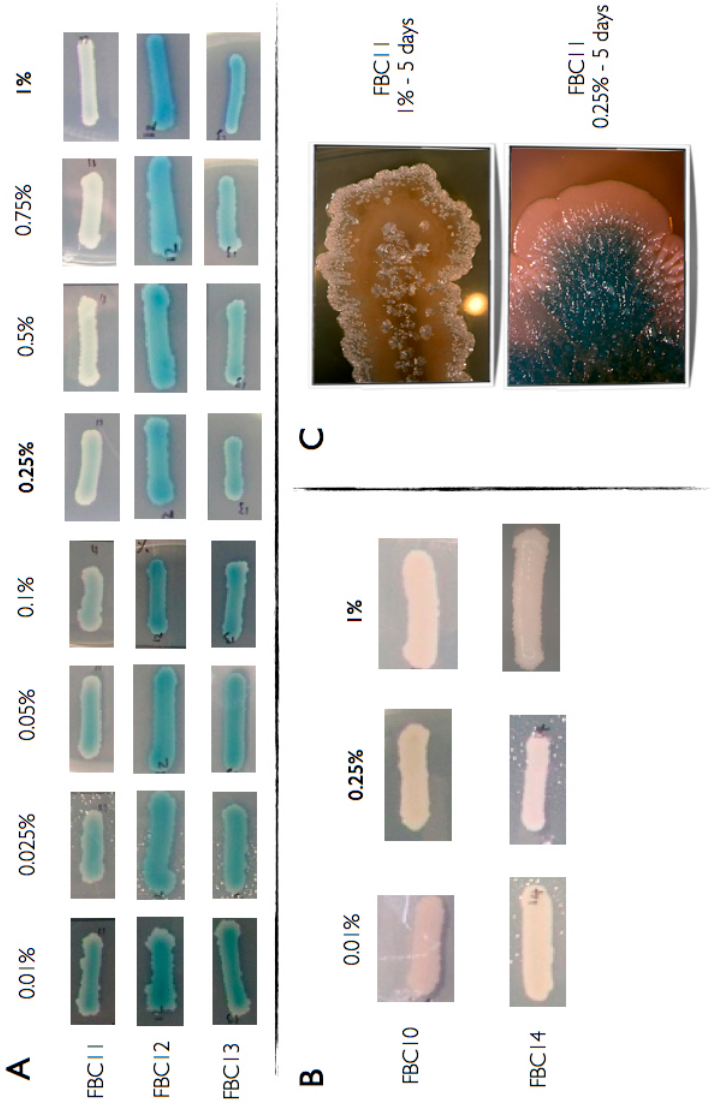

Figure 4.7: (A) Qualitatively assays of β-gal activities of strains FBC11,12 and 13 grown on different calcium concentration (37°C, 48 h incubation); (B) Qualitatively assays of β-gal activities of strains FBC10 and FBC14 grown on different calcium concentration (37°C, 48 h incubation); (C) Calcite crystals on FBC11 after 5 days incubation with 1.0% calcium acetate (up), beta-gal activity pattern on FBC11 after 5 days on 0.25% calcium acetate (down)

INFLUENCE OF TEMPERATURE AND CA^{2+} AVAILABILITY IN LIQUID CULTURES

In this section we have investigated, how different temperatures and Ca^{2+} availability affect the expression of the *lcfA* operon in B4 liquid cultures by comparing the β-Galactosidase (LacZ) activity in strains FBC11, FBC12 and FBC14 as control. Cultures were grown in liquid B4 media (*see* EXPERIMENTAL PROCEDURES), samples were collected at each time point (60′) for nine hours or until five hours after entry the stationary phase. Samples were then assayed for β-Gal synthesis as described in EXPERIMENTAL PROCEDURES.

Strains FBC11,12 and 14 were grown in B4 liquid media at 37°C and 42°C. Any difference were detected among the growth rate of the cultures at 37°C and 42°C, figure 4.8. As regards the β-Gal activity (Miller Units, MU), that of FBC12 showed the higher values at t_0 and during the early stage of growth with high values at 42°C compared to that at 37°C, then at both temperatures β-Gal activity tends to decline toward a plateau which was reached at the entry of stationary phase, figure 4.9 B right panel. On the converse β-Gal activity in FBC11 strain is maximized during the first 60 min at 42°C, while at 37°C the maximum was reached after 120 min of incubation, figure 4.9 B left panel. In a similar way to that observed for FBC12, FBC11 activity once has reached the maximum tends toward oscillations around a lower β-Gal activity value, and these oscillations are coincident with the entry in the stationary phase, figure 4.9 B left panel. Although the trend of activity of strains FBC11 and FBC12 shares similarities, magnitude of activation is higher in FBC12 than in FBC11, figure 4.9 B left panel.

We tested β-Gal activity of strains FBC11, FBC12 and FBC14 as control in B4 liquid cultures supplemented with different concentration of Ca-acetate. We have not used Ca-acetate concentration over the 0.25% (corresponding to a Ca^{2+} concentration of 15μM) because of higher concentration resulted in the formation of precipitates which inhibited the measurement of the β-Gal activity, thus we have assayed FBC11,12 and FBC14 in presence of 0.01%-0.1% and 0.25% of Ca-acetate (corresponding to a Ca^{2+} concentration of 0.6μM, 6.0μM, and 15μM respectively). As for temperature, growth rate of cultures was not affected by the different Ca^{2+} concentrations, figure 4.8.

The β-Gal activity measurements of strain FBC11, at all [Ca^{2+}], have showed that the highest expression is achieved during the first 60 min of incubation, figure 4.9 A left panel, with a β-Gal activity spike after 60 min at 0.25% Ca-acetate (15μM Ca^{2+}); then tends to decline during the early/mid-

log phase till the entry in stationary phase (300 min) after that a slight increase of the expression until the end of the measurements is shown, figure 4.9 A left panel, this trend is similar at the three [Ca^{2+}], but with great difference in the activity values (MU). In FBC12 β-Gal activity level are maximized during the early stage of growth (0-60 min), figure 4.9 A right panel, with highest activity reported for low [Ca^{2+}]: 0.1% Ca-acetate (6.0μM) at t_0, and 0.01% (0.6μM respectively) at t_{60}. Proceeding with incubation, β-Gal activity tends to decrease toward a low activity level, which was reached at the entry of the stationary phase for all [Ca^{2+}], and remains stable until growth end, figure 4.9 A right panel. Also in this case, as seen for temperature the activity values of FBC12 are higher than in FBC11 (*see* Miller Units in figure 4.9 A right panel).

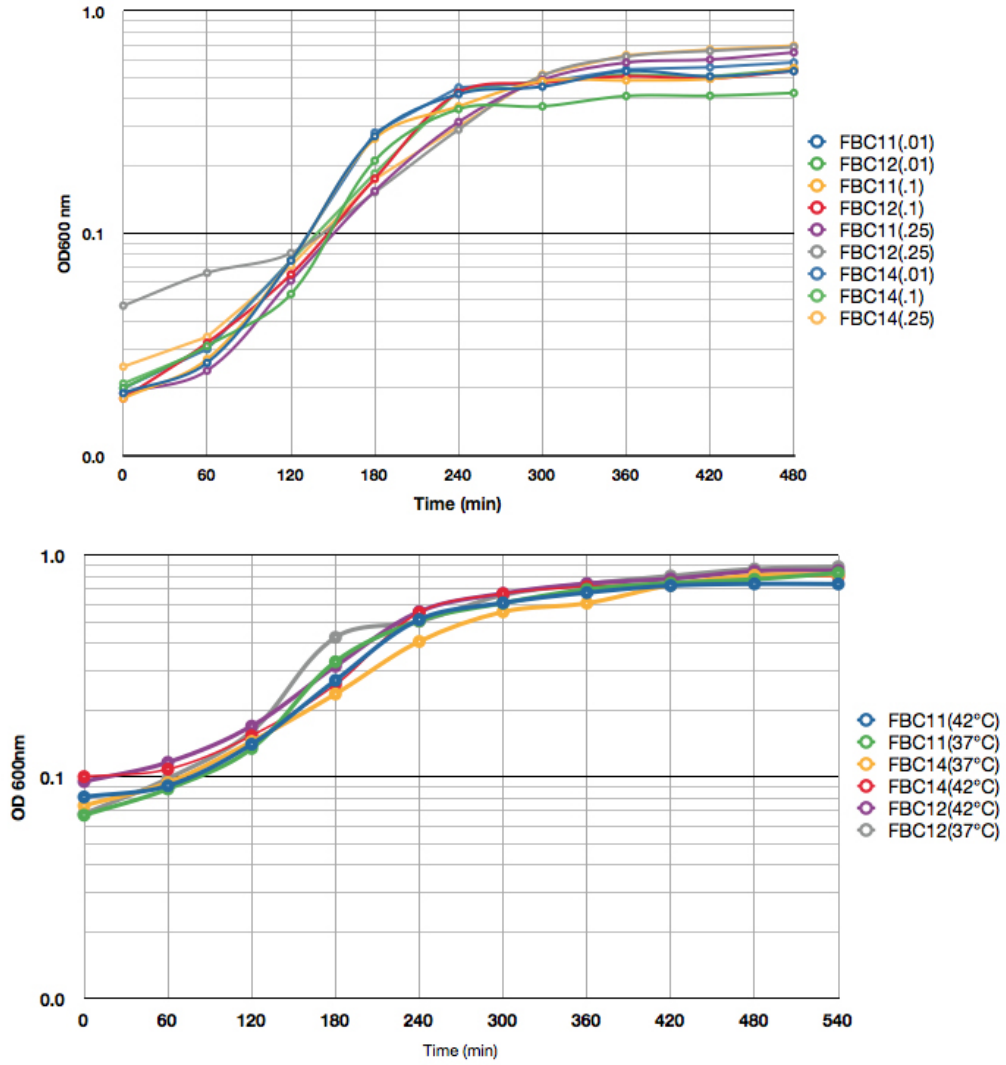

Figure 4.8: *Upper panel* growth kinetics of strains FBC11,12 and 14 in B4 liquid cultures supplemented with different calcium concentrations at 37°C; *Lower panel* growth kinetics of strains FBC11,12 and 14 in B4 liquid cultures at different temperatures

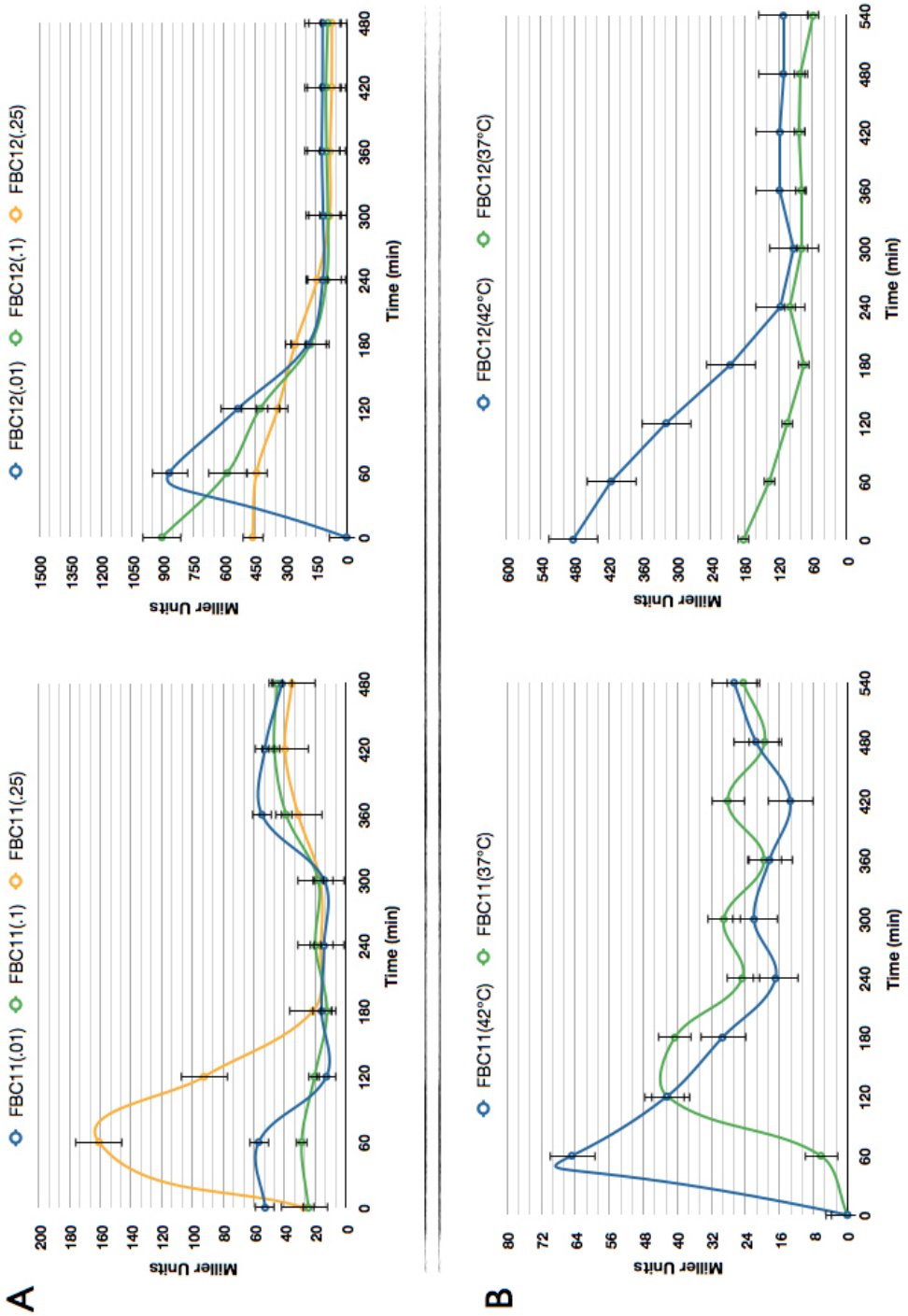

Figure 4.9: (A) *left panel* β-Gal activity of strain FBC11 grown with different calcium concentration, *right panel* β-Gal activity of strain FBC12 grown with different calcium concentration; (B) *left panel* β-Gal activity of strain FBC11 grown at different temperatures; *right panel* β-Gal activity of strain FBC12 grown at different temperatures

4.4 CONCLUSIONS

In this section we have explored, by constructing promoter::*lacZ* trascrptional fusions, the expression of the *lcfA* operon genes and the timing of their activation along with biomineralization. We have investigated how perturbation of temperature, pH ,and Ca^{2+} availability affect the expression and biomineralization. We have focused our attention, mainly, on strain FBC11 which harbours the transcriptional fusion with the P_{ysiA} thought to be the main responsible for the expression of *etfA*.
It has been reported from several authors that calcite precipitation by bacteria increased along with temperature of incubation [30, 29], thus we have tested how temperature affects the expression and calcite precipitation in our model *B. subtilis*. We have qualitatively monitored the expression of the *lcfA* operon, in solid B4. Our investigation have showed that temperature could affect the magnitude of β-Gal expression in strains FBC11 and FBC13, figure 4.3, while in FBC12 we did not observe significant variations between different temperatures, figure 4.3. Timing of the activation[5], depends on temperature oscillating around 24-48 h, however proceeding with incubation, additional variations of the expression can be observable only for strains FBC11 and FBC13 in which the starting expression level is lower than that observed in FBC12, figure 4.3.
In particular, strains FBC11 and FBC13 display a similar trend of expression, in which β-Gal expression tends to increase following the temperature augmentation, figure 4.3. In addition, when we compared strains grown at the same temperature, a great difference, in terms of expression strength, can be observed among FBC12 and FBC11,13, while a slight one can also be detectable among FBC11 and FBC13. Despite our *in silico* analysis showing that a putative promoter region could be located into the inter-genic region between *ysiB* and *etfBA*, we did not obtain β-Gal expression in strain FBC10 in all the assayed temperatures, figure 4.3.
Concerning calcite precipitation, we noticed that at the lowest temperature (30°C) we have cultivated strains FBC10-14, precipitation was completely inhibited, although β-Gal expression in FBC11,12 and 13, was observed. On the contrary when strains are grown at 37°C they showed the same behavior observed for the WT strain 168, with calcite precipitates usually produced after 6-8 days, while when they have been grown at 42°C crystals are produced within 2-4 days. By comparing the β-Gal expression and calcite precipitation in strains FBC11, FBC12 and FBC13 we observed that

[5]We consider as timing of activation the point (in time) in which the degradation of X-Gal leading to the development of the characteristics blu, becames visible

they did not occur simultaneously, indeed expression starts usually 2-4 days before calcite production.
Several works have reported that pH changes in the medium were the main determinant of calcite production on solid media, and it has been demonstrated that calcite precipitation could be controlled, even in a mutant impaired in precipitation, by cultivating it in alkaline buffered B4. Therefore we have tested if pH could affect the expression of the *lcfA* operon genes, and how biomineralization is affected.
Qualitatively assays of β-Gal expression in FBC10-14 strains cultivated in B4 buffered medium pH 7.7 and pH 8.2 at 37°C did not reveal any significant difference when compared to those cultivated on unbuffered B4, figure 4.4. Similarly, when cultivated on buffered B4 at different temperature 42°C and 30°C we did not observed any significant difference in β-Gal expression compared to unbuffered B4, figures 4.5 and 4.6, even-though a slight difference can be observable in FBC13 cultivated at 30°C on both pH 7.7 and pH 8.2 buffered B4, figure 4.6. Considering calcite precipitation it was completely inhibited on pH 7.7 plates at all temperatures tested, in contrast to pH 8.2 plates incubated at 42°C and 37°C in which it was not inhibited. Calcite production were observed after 6-8 days and 2-4 days of growth on pH 8.2 B4 at 37°C and 42°C, respectively.
Few systematic studies have been devoted to investigating the role of Ca^{2+} as an intracellular messenger in *B. subtilis*, providing evidence for a regulatory role for Ca^{2+}. Considering the key role of calcium in the regulation of many fundamental cellular processes, we have explored how different calcium concentrations affect expression of the *lcfA* operon genes along with biomineralization.
Qualitatively assays of β-Gal expression in strains FBC10-14, cultivated in diverse $[Ca^{2+}]_i$ ranging from 0.01 to 1.0%, were performed. Interestingly β-Gal expression in FBC11 seems to be influenced by the $[Ca^{2+}]_i$, with higher β-Gal expression level at low concentration (0.01-0.25%), and low expression level at higher ones (0.5-1.0%), figure 4.7. Although only qualitative observation can be made through this transcriptional fusion system on solid media, increment of the expression can be detected proceeding incubation in 0.5-1.0% Ca-acetate, despite that calcite was already precipitated. Furthermore we showed that the β-Gal expression in FBC12, 13 is not affected by the calcium concentration in the media, figure 4.7. Calcite precipitation was completely inhibited at low concentration, while at high level (0.5-1.0%) of Ca^{2+} crystals begin to appear after 2-4 days after incubation.
Our data suggested that there was no obvious connection between the precipitation of calcite and the expression of *lcfA* operon genes starting from

P_{lcfA}, and P_{ysiA} as well. The only factor among those studied, which seems to affect the expression of P_{ysiA} and calcite precipitation in a concording manner is temperature, figure 4.3. Increase of temperature of incubation reduces the timing of calcite production and at the same time results in increased intensity of the β-Gal expression level in FBC11 and FBC13. On the converse pH and $[Ca^{2+}]$ seem to influence calcite precipitation even in absence of β-Gal expression. Thus, results presented here indicated that several factors could affect biomineralization, rather than the expression, suggesting that the expression of *lcfA* operon genes is essential but not indispensable. This supports the hypothesis of several authors which assume that calcium carbonate precipitation is a complex process affected by temperature, composition of the growth media together with the expression of the *lcfA* operon genes, raising that during biomineralization different mechanisms has an essential influence on the formation of crystals.

5

FUNCTIONAL ANALYSIS OF FADB AND ETFA

PREFACE

In previous works the genes belonging to the *lcfA* operon have been described and set out their involvement in calcite precipitation in *B. subtilis* [7]. In Perito et al., 2000 [89], through a large scale screening of insertional mutants, were identified two strains, with an insertion into the *ysiB* and *ysiA* genes, impaired in crystal formation; then Barabesi et al., 2007 suggested that the latest gene of the operon *etfA* is essential for biomineralization (*see* CHAPTER 3). Recently Marvasi et al., 2010 [74] have focused on the physiology behind the impairment of calcite precipitation in the *etfA*-disruptant, concluding that the main process responsible for the lack of calcite precipitation in this strain is the decrease of the pH in the medium probably due to a misregulation of the EtfA pathway. These genes were originally described in the nineties as a group of genes related to fatty acids β-oxidation [116], and thought to be the functional homologs of the Fad regulon genes of *E. coli* in *B. subtilis* [53]. Within these genes only the *ysiA* gene has been recently characterized, as a transcription factor regulating the expression of fatty acids β-oxidation enzymes [75]. The function of the other genes and their involvement in this pathway remain unknown. Thus we have embraced the possibility to start a functional study on these genes, determining their function and possibly their influence on biomineralization. In our analysis we have focused on two genes of the operon, *ysiB* and *etfA* which enconde a putative enoyl-CoA hydratase and the α-subunit of prokaryotic heterodimeric flavoprotein, respectively. The translational product of *ysiB* has been found to be homologous to the FadB protein in *E. coli* and to the *rpfF* product of *Xanthomonas campestris* involved in the pathway for the production of a factor (*cis*-11- methyl-2-

dodecenoic acid) regulating phytopathogenicity. On the other hand EtfA shared more than 60% similarity with prokaryotic electron transfer flavorprotein. Particularly interesting is the similarity shared with that of *Clostridium acetobutylicum* in which its *co*-expression with the β-subunit, encoded by *etfB*, is essential for the butanol-butyrate synthesis pathway. These observation suggested a possible link between the function of the *lcfA* operon genes and biomineralization. This could be ascribed to possible pleiotropic effects that mutantions in fatty acids pathways may have on cellular metabolism or to a direct involvement in biomineralization through regulating the synthesis of such fatty acids intermediate (*e.g.*: *rpfF* of *X. campestris*).

5.1 PURIFICATION AND FUNCTIONAL ANALYSIS OF FADB$_{Bs}$

5.1.1 BACKGROUND

Fatty acids are essential components of membranes and are an important source of metabolic energy. Fatty acid degradation and biosynthesis pathways have been mainly studied in the model prokaryote *E. coli*. Fatty acids that are intracellularly formed or extracellularly supplied are degraded through β-oxidation when cells are starved for carbon source. In *E. coli* the degradation pathway is catalysed by the enzymes encoded by the *fad* regulon which is responsible for the transport and activation of long-chain fatty acids and their β-oxidative cleavage into acetyl-CoAs [32, 31]. *B. subtilis* possesses a considerable number of genes, organized in five operons, that are possibly involved in the β-oxidation of fatty acids due to their similarity with corresponding *E. coli* genes; these are : *lcfA, lcfB(yhfL), fadB$_{Bs}$(ysiB), acdA, fadNAE(yusLKJ)* [75] and the *mmgABC* genes transcribed by the σ^E-RNA polymerase [28].
The conservation of protein sequences among different species, suggested that β-oxidation plays indispensable functions under certain physiological conditions in *B. subtilis*, such as sporulation [55] as well as calcium carbonate biomineralization [7]. Among the genes products involved in fatty acids β-oxidation in *B. subtilis* only YsiA has been recently characterized as a transcriptional regulatory protein, belonging to the *TetR* family, which negatively regulates the expression of majority of β-oxidation genes (*ysiA, fadB$_{Bs}$, etfB, etfA, fadNAE, ykuFG, lcfB, and ywjF-acdA-rpoE*) including those belonging to the *lcfA* operon [53, 75].
However the involvement of the YsiA regulon members in β-oxidation is only inferred on data based on BLASTP similarity of their gene products, therefore some compelling evidence is much needed to demonstrate what these genes actually do. The *B. subtilis* FadB protein has previously been found to be homologous to eukaryotic and prokaryotic proteins belonging to the crotonase superfamily. The crotonase superfamily is comprised of mechanistically different proteins that share a conserved quaternary structure. Some enzymes in the superfamily have been shown to display hydratase and isomerase activity as well as the hydrolysis of thioesthers. Bacterial members of this superfamily have been found to be involved in different metabolic pathways such as fatty acids beta oxidation FadB in *E. coli* [43], polyhydroxyalkanoate PHA [106, 98] and butanol biosynthesis CRT of *C. acetobutilycum* [59]. In this section we report the cloning and the

expression of *B. subtilis* FadB(YsiB) and its functional characterization as an enoyl-CoA hydratase.

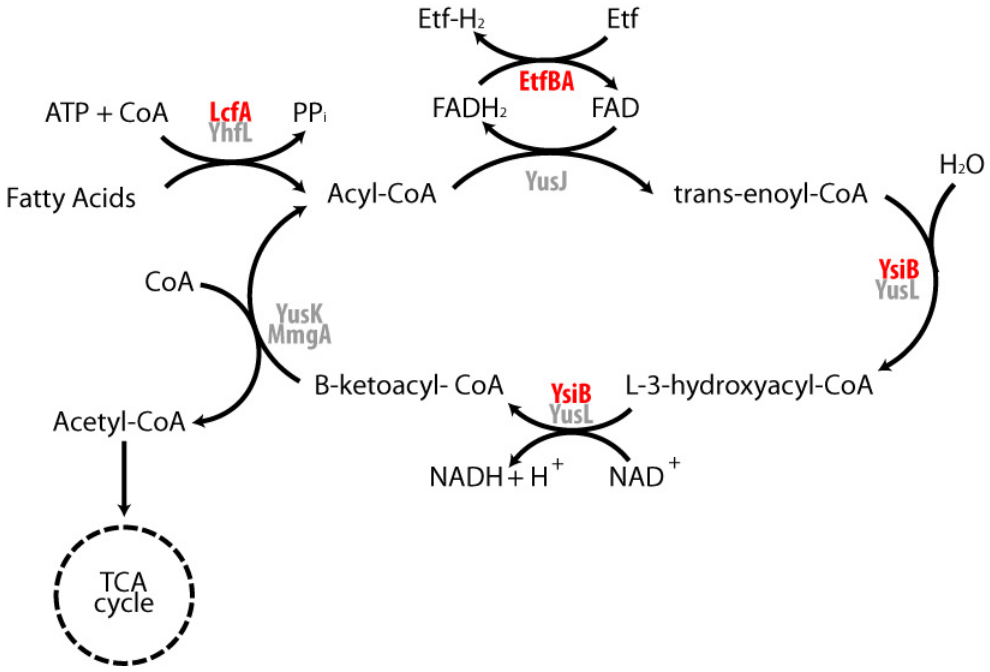

Figure 5.1: fatty acids β-oxidation pathway and the probable involvement of *lcf* A operon gene products, [53, 75]

5.1.2 Results & Discussion

In order to assign a function to *fadB*$_{Bs}$ gene product, database search was carried out using default BLASTP parameters (http://blast.ncbi.nlm.nih.gov), results showed that FadB$_{Bs}$ has the highest identity with enoyl-coA hydratase of *Bacillus amyloliquefaciens* FZB42 (77%, ABS74918) and with putative enoyl-coA hydratase from *Bacillus licheniformis* DSM13 (73%, AAU41868) and ATCC14580 (73%, AAU24508), both only putatively attributed to their function. We found an enoyl-CoA hydratase/isomerase (ECH) domain in FadBBs through SWISS-PROT and TrEMBL database search with the Expasy ScanProsite tool. The ECH domain is also shared with the crotonase of *C. acetobutylicum*, encoded by the *crt* gene belonging to a gene cluster showing a similar organization to that of the *lcfA* operon [7]. Moreover, sequence alignement of the ECH domain found in FadB$_{Bs}$ with those of known enoyl-CoA hydratases from different sources displays significant

identities with the residues known to be important for catalysis [3]. These

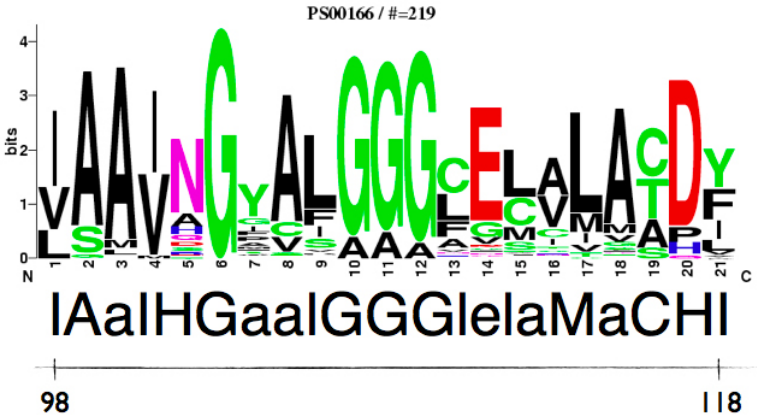

Figure 5.2: Comparison of the domain found in YsiB to the sequence logo obtained from the alignment of ECH domains found in ProSiTe. A sequence logo is a graphical display of a multiple sequence alignment consisting of colour-coded stacks of letters representing amino acids at successive positions. Sequence logos provide a richer and more precise description of sequence similarity than consensus sequences and can rapidly reveal significant features of the alignment that could otherwise be difficult to perceive. The total height of a logo position depends on the degree of conservation in the corresponding multiple sequence alignment column. Very conserved alignment columns produce high logo positions. The height of each letter in a logo position is proportional to the observed frequency of the corresponding amino acid in the alignment column. The letter of each stack is ordered from most to least frequent, so that it is possible to read the consensus sequence from the top of the stacks.

findings suggested that FadB$_{Bs}$ could be active towards crotonyl-CoA. In order to test this hypothesis, FadB$_{Bs}$ was produced in *E. coli* and purified by a Ni-column (Macherey-Nagel). The *E. coli* crude extract and the eluate after purification showed an intense band of approximately 30.5 KDa (Theoretical MW= 28.3 KDa) after SDS-PAGE separation Both crude extract and eluate were tested for proteins content, and enoyl-CoA hydratase activity using crotonyl-CoA as substrate by following absorbance decrease at 263 nm, according to described protocols [57](Hartmanis and Gatenbeck, 1984). The activity of the crude extracts was 231 EU/mg of proteins (One EU of enoyl-CoA hydratase was defined as the amount of enzyme able to hydratate 1 μmol of crotonyl-CoA per min (using ϵ263 = 6.7 x 103 M^{-1} cm^{-1}) and the first FadB$_{Bs}$ eluate showed enoyl-CoA hy-

dratase activity of 1,603 EU/mg of proteins (yield 15% and purification fold 7). We performed a kinetic analysis on the purified enzyme measuring both kinetic and specific constant parameters in the direction of the hydratation towards crotonyl-CoA. K_M value was 46.9 μM, very similar to other reported values for bacterial and mammals crotonases, *e.g.: C. acetobutylicum* 30 μM [6], *E. coli* 50 μM [18], *Homo sapiens* 30 μM [60], *Sus scrofa* 13 μM [51], confirming an analogue behavior for this ubiquitous enzyme. This evidence further suggests the identification of a real crotonyl-CoA activity. The values of $FadB_{Bs}$ specific constants were: k_{cat} 32 $\pm$ 7 min^{-1}, k_{cat}/K_M 690 $\pm$ 190 min^{-1} μM^{-1}. *B. subtilis* $FadB_{Bs}$ has been proposed to be involved in the fatty acid β-oxidation pathway as a subunit of the putative $FadB_{Bs}$/FadN(YusL) complex (enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase [53]. Our data revealed that *B. subtilis* FadB is active in the direction of the hydratation of crotonyl-CoA, supporting the possibility of its direct involvement in the β-oxidation pathway. The findings here reported contribute to add a new and relevant piece of information to the poorly characterized β-oxidation pathway in *B. subtilis*.

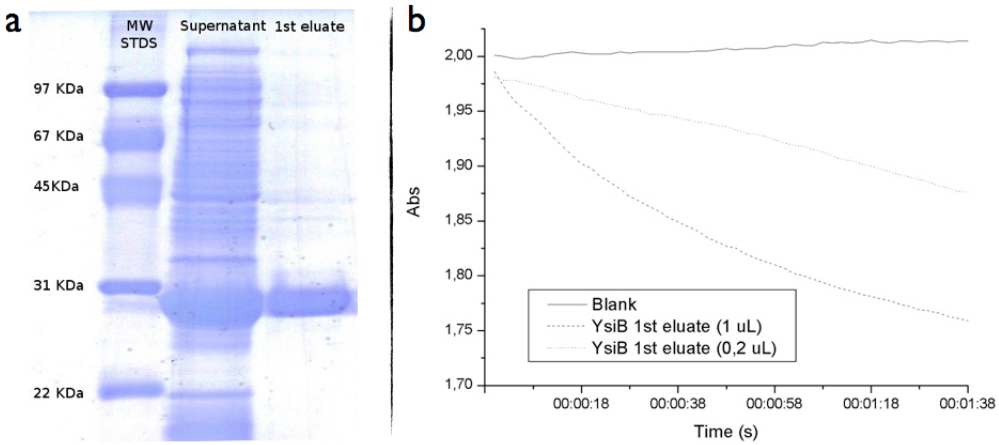

Figure 5.3: (A) SDS-PAGE 12% polyacrilamide of purified $FadB_{Bs}$, *supernatant* stand for crude extract, 1st eluate from Protino columns; (B) Kinetic analysis of $FadB_{Bs}$.

5.2 PURIFICATION OF ETFA AND CALCIUM BINDING DOMAIN CHARACTERIZATION

5.2.1 BACKGROUND

Electron transferring flavoproteins (ETFs) catalyze electron transfer between other flavoproteins [83, 108]. The pig liver enzyme is the best characterized ETF. It is found in the mitochondrion, where it functions to transfers electrons from flavoprotein dehydrogenases involved in the metabolism of fatty acids to a further flavoprotein named ETF-CoQ oxidoreductase. The latter enzyme passes the electrons to the terminal electron transfer chain [108]. Mammalian ETF is a heterodimer that contains one FAD and one AMP [97, 45]. The 3-D structure of the recombinant enzyme has showed that the FAD interacts mainly with the α-subunit, while the AMP is buried in the β-subunit [94]. Protein similar to mammalian ETF have been isolate from bacteria. ETFs from *Paracoccus denitrificans* and the methylotroph W3A1 are also heterodimers with one FAD and one AMP per molecule [45, 58, 104]. The mammalian and *P. denitrificans* enzymes share very similar catalytic properties [113], and in both organisms, the genes are arranged in tandem with the gene for the β-subunit preceeding the α-subunit gene [13, 36]. Sequence analysis identified similar genes in a variety of other organisms *i.e.*: those of nitrogen fixing and butanol producing bacteria.There is evidence that the proteins encoded by the *fixAB* genes are involved in nitrogen fixation, even-though their precise functions has not yet been established [61, 56]. Proteins encoded by *fixA* and *fixB* genes have similar sequences to the α- and β-subunits, respectively, of ETF and their genetic organization is similar to that of bacterial Etf [50, 13, 36, 4]. Furthermore a similar organization of the *etfBA* genes has been noted also in the BCS operon of *C. acetobutylicum* involved in the conversion of acetoacetyl-CoA to butyryl-CoA in the butanol-butyrate synthesis pathway in *clostridia* [22]. Recent analysis of the genomes of the sequenced *clostridia* have revealed multiple homologues of *etfA* in each organism [84]. Interestingly in *Clostridium perfringens* EtfA catalyse the formation of DPA (Dipicolinic Acid) from the product of DHDPA synthase [84], which is essential for the formation of stable endospores. Many sporulation-specific genes are highly conserved across all spore formers and several different genera, but it is unclear what the functions of these homologues may be, even-though *etfA* in various genera of bacteria have been implicated with different pathway such as butyrate production [59], DPA biosynthesis [84], and $CaCO_3$ mineralization [7]. In the majority of reactions utilizing these

proteins, EtfA is thought to form an heterodimer with the smaller EtfB molecule, but it should not ruled out the possibility that EtfA is complexed with another protein *in vivo* functioning without an EtfB counterpart. In *B. subtilis* the products of *etfB* and *etfA* genes belong to the *lcfA* operon are thought to be involved in fatty acids β-oxidation [116, 53] and in $CaCO_3$ precipitation [7]. As noted in Barabesi et al., 2007 [7] the *lcfA* operon shares a similar organization to that of the BCS operon in *C. acetobutylicum*. EtfA is thought to catalyse the reduction of acyl-CoAs during β-oxidation of fatty acids [53], and is essential for the production of calcite precipitates in *B. subtilis* [7, 74]. Thus we have planned to study if a link between its function and biomineralization exist. To do so we have cloned and purified the product of the *etfA* gene in order to characterize its function along with its involvement in $CaCO_3$ precipitation.

5.2.2 Results & Discussion

The function of the two EtfB-EtfA proteins in *B. subtilis* was only putative, they share 64%(EtfB) and 57%(EtfA) similarity with those of *C. acetobutylicum*. A database search through the Expasy Scan Prosite tool with the EtfB and EtfA protein sequences has revealed the presence of a FAD binding domain in both the α- and β-subunit, figures 5.4 and 5.5. In EtfA the search for conserved motif has revealed a new domain similar to bacterial EF-hand of known bacterial calcium binding domain (CBDs). This type of domain consists of a twelve residue loop flanked on both side by a twelve residue α-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu (E) or Asp (D) at position 12 provides two oxygens for liganding Ca (bidentate ligand). The basic structural/functional unit of EF-hand proteins is usually a pair of EF-hand motifs that together form a stable four-helix bundle domain. The pairing of EF-hand enables cooperativity in the binding of Ca^{2+} ions. Since calcium has been reported to regulate many processes in bacteria, we hypothesized that the CBD found in EtfA could be either regulated or to contribute directly to $CaCO_3$ biomineralization. In bacteria, Ca^{2+} is implicated in a wide variety of cellular processes, and could control various protein functions by affecting stability, and their enzymatic activity. Thus we have planned to investigate if the CBD of EtfA is able to bind Ca^{2+} and if/how affect the enzymatic activity of EtfA. To test our hypothesis we have cloned *etfA* in pET21b vector, produced recombinant His_6-Tagged

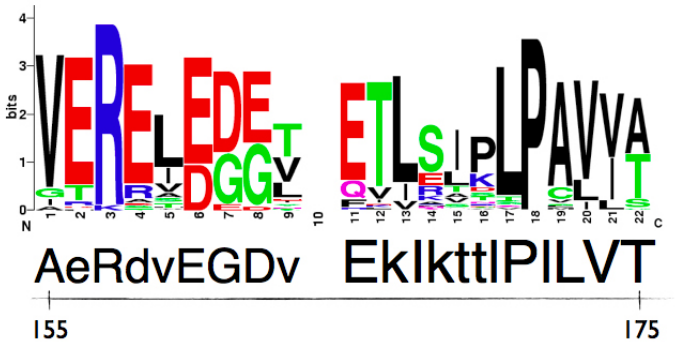

Figure 5.4: Comparison of FAD binding domain sequence of EtfB with the sequence logo obtained from the multiple sequence alignment of FAD domains found in ProSiTe.

EtfA in *E. coli* and purified through a Ni-column as described in EXPERIMENTAL PROCEDURES. The *E. coli* crude extract showed an intense band of approximately 30 KDa (Theoretical MW = 34.55 KDa[1]), figure 5.6. In the majority of reactions utilizing these proteins, EtfA is thought to form an heterodimer with the smaller EtfB molecule (Theoretical MW = 28.52 KDa), so that we have also cloned the *etfA* and *etfB* genes, together on the same pET21b vector, in order to express bot genes and evaluate the formation of the heterodimer. By this construct we were not able to purify the heterodimer, probably because of no covalent interaction are required to achieve the formation of the EtfBA heterodimer, figure figure 5.6. Further investigation are needed to purify the heterodimer and analyze how it functions along with the enzymatic activity and biomineralization.

[1]Theorethical molecular weight with the His_6 tail is 35.7KDa

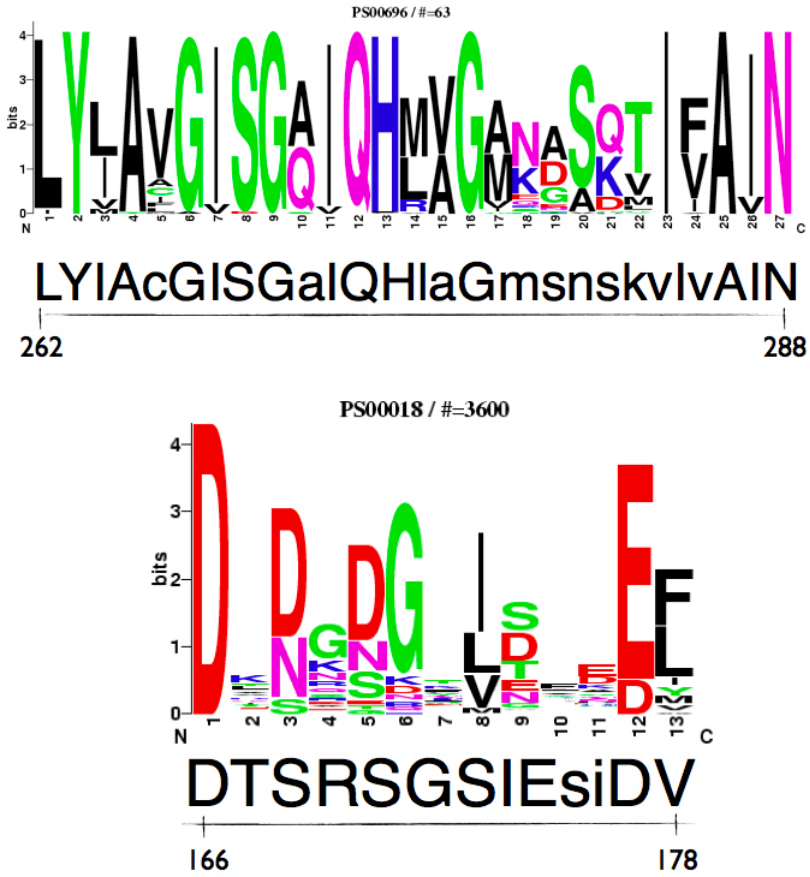

Figure 5.5: *upper panel:* Comparison of FAD binding domain sequence of EtfA with the sequence logo obtained from the multiple sequence alignment of FAD domains found in ProSiTe. *lower panel:* Comparison of the EF-HAND domain sequence found in EtfA to the sequence logo obtained from multiple sequence alignment of bacterial EF-HAND domains found in ProSiTe.

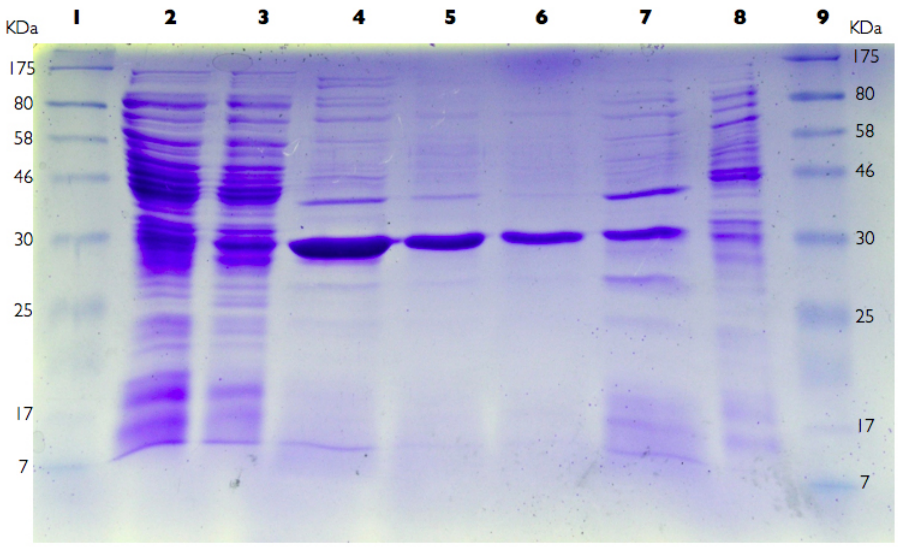

Figure 5.6: SDS-PAGE 12% polyacrilamide of purified His_6-EtfA and crude extract of BL21 DE3 (pET21b-*etfA*); Lanes: (1) Molecular weight protein standard; (2) crude extract of BL21DE3 (pET21b-*etfA*); (3) crude extract of BL21DE3 (pET21b-*etfA*) induced with 1 mM IPTG; (4) 1st eluate from Protino columns of His_6-EtfA; (5) 2nd eluate from Protino columns of His_6-EtfA; (6) 3rd eluate from Protino columns of His_6-EtfA; (7) 1st eluate of BL21DE3 (pET21b-*etfAB*) induced with 1 mM IPTG; (8) crude extract of induced BL21DE3 (pET21b-*etfBA*); (9) Molecular weight protein standard.

5.3 EXPERIMENTAL PROCEDURES

Bacterial strains & plasmids

Bacterial strains used in this study are listed in Table 5.1.

Table 5.1: Bacterial strains and plasmids

Strain or plasimd	Genotype	Source or reference
E. coli strains		
JM109	*endA1, glnV44, thi-1, relA1, gyrA96, recA1, mcrB*$^+$, Δ*(lac-proAB), e14*$^-$ [F′ *traD36 proAB*$^+$ *lacI*$_q$ *lacZ*Δ*M15*] *hsdR17(rK*$^-$ *mK*$^+$*)*	Promega Co.
BL21 DE3 Gold	F$^-$ *ompT, gal, dcm, lon, hsdSB(rB*$^-$ *mB*$^-$*)* λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]), , *tet*^{r}	Novagen Inc.
JM109-pET21b	*pET21b, amp*^{r}	This study
JM109-*ysiB*	*pET21b-ysiB, amp*^{r}	This study
JM109-*etfA*	*pET21b-etfA, amp*^{r}	This study
JM109-*etfB-etfA*	*pET21b-etfB-etfA, amp*^{r}	This study
BL21-pET21b	*pET21b, tet*^{r}*, amp*^{r}	This study
BL21-*ysiB*	*pET21b-ysiB, tet*^{r}*, amp*^{r}	This study
BL21-*etfA*	*pET21b-etfA, tet*^{r}*, amp*^{r}	This study
BL21-*etfB-etfA*	*pET21b-etfB-etfA, tet*^{r}*, amp*^{r}	This study
B. subtilis strains		
168	*trpC2*	Lab. stock
Plasmids		
pET21b	Vector, amp^r	Novagen Inc.
pET21b	*ysiB*, amp^r	Novagen Inc.
pET21b	*etfA*, amp^r	Novagen Inc.
pET21b	*etfBA*, amp^r	Novagen Inc.

DNA manipulation

Isolation of total genomic DNA and plasmids, digestion of DNA with restriction endonucleases and transformation of *E. coli* were carried out by standard procedures. The *fadB*$_{Bs}$*(ysiB)*, *etfA*, and *etfBA* genes were amplified by PCR using the oligonucleotides primer listed in Table 5.2, and DNA of *B. subtilis* strain 168 as template. The amplified DNA was purified with

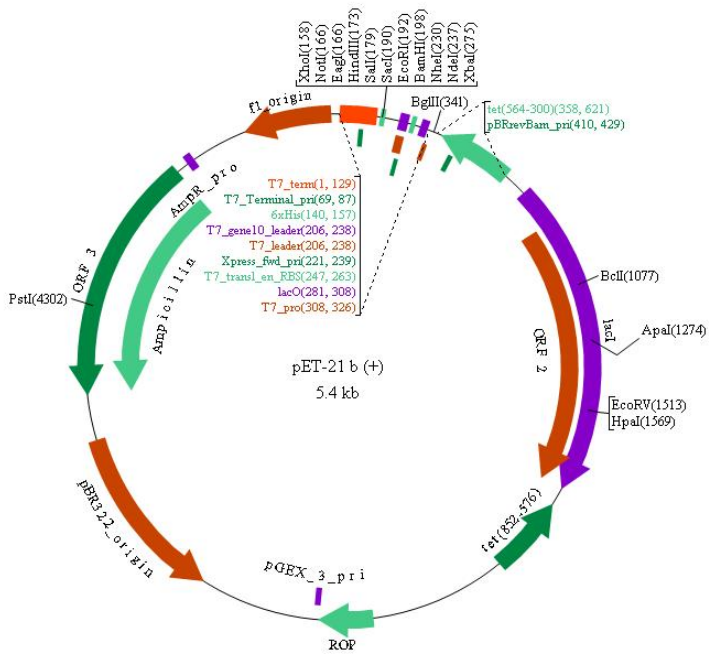

Figure 5.7: Physical Map of the cloning expression vector pET21b(+) (Novagen).

Wizard Promega. Purified *fadB*$_{Bs}$, *etfA*, *etfBA* and pET21b, figure 5.7, were digested with *Nde*I/*Ava*I, and *Nde*I/*Xho*I, respectively (New England BioLabs). The resulting fragments were ligated and then used for transformation of *E. coli* BL21 DE3 Gold. Correct cloning of *fadB*$_{Bs}$, *etfA*, and *etfBA* in plasmid pET21b was confirmed by nucleotide sequencing.

Production and analysis of His$_6$-Tagged proteins

Recombinant *E. coli* BL21 DE3 cells harboring pET21b vectors (Table 5.1) were inoculated in 10 ml of LB medium containing Ampicillin 100 mg l^{-1} and Tetracyclin 12.5 mg l^{-1}. After an over-night growth at 37°C in a reciprocal shaker, cells were diluted in 50 ml of fresh LB medium (Amp 100 mg l^{-1}, Tet 12.5 mg l^{-1}) in a 500 ml flask (OD_{600} about 0.2-0.3) and then cultivated at 37°C with vigorous aeration for three hours, followed by 1 mM IPTG addition and an additional incubation for two hours to achieve His$_6$-Tagged proteins production. After IPTG induction, cells were harvested (8.000× g, 20 min at 4°C) and resuspended in 50 mM MOPS pH 7.00 and 300 mM NaCl. The suspension was incubated on ice for 30 min with 1 mg/ml of lysozime (Sigma Co.) and then mechanically homogenized for

Table 5.2: Oligonucleotides primer used for the isolation of *fadB*$_{Bs}$, *etfA*, and *etfBA* genes and to check the correctness of the transformation are listed in Table 5.2, restriction site are underlined.

Oligonucleotide primer name	**5′-3′ sequence**
Fw-*etfA*-*Nde*I	CAGGGGGATGTTCATATGGG
Rw-*etfA*-*Xho*I	GGCCTCGAGGAGTGTATATTCAGCTG
Fw-D7(*etfB*)-pET-*Nde*I	GGGAATTCCATATGAATCTATTTGTACTG
Fw-*ysiB*-*Nde*I	GGAATTCCATATGAATGCAATTTCACTT
Rw-*ysiB*-*Ava*I	CCCTCGAGTTCGCCTTTGAACTGAGG

FadB$_{Bs}$, while for the purification of EtfA and EtfBA was directly sonicated sonicated[2], without lysozime digestion. After centrifugation (20.000 $\times$ g, 30 min at 4°C), the resultant soluble fraction (crude extract) was loaded directly onto a Protino© 150 Column (Macherey-Nagel) and eluted with 50 mM MOPS pH 7.00, 300 mM NaCl, 250 mM Imidazole. All purification procedures were carried out at 4C to prevent protein denaturation and loss of activity. Sodium Dodecyl Sulfate (SDS) polyacrylamide gel electrophoresis 12% was performed according to Laemni protocol [65] and stained with Blue Brilliant Coomassie R250. Total proteins were assayed using Coomassie Blue Dye staining protocol [23].

Assay of Enoyl-CoA hydratase activity of FadB$_{Bs}$

The enoyl-CoA hydratase activity was measured by assaying the hydration of crotonyl-CoA (Sigma Co.) at 263 nm (UltroSpec 2100 pro, Amersham Bioscience, Milan, Italy) through the decrease in absorbance of the conjugated double-bond band at 263 nm. In a final volume of 3 mL, 0.1 mL of enzyme solution were added to 44 mM Tris HCl buffer (pH 7.5), 0.0044% bovine serum albumin, 0.67 mM EDTA and 0.107 mM crotonyl-CoA. Decrease in absorbance at 263 nm was measured at 25°C. One unit (EU) of enoyl-CoA hydratase was defined as the amount of enzyme able to hydratate 1 μmol of crotonyl-CoA per min (using ϵ_{263nm} = 6.7 x 103 M^{-1} cm^{-1}). The specific activity of enoyl-CoA hydratase was defined as the activity of enoyl-CoA hydratase per milligram of protein. The kinetic constants, maximum reaction rate (V$_{max}$), and Michaelis constant (K$_M$), and their standard deviation were determined using R 2.5.1 software (R Foun-

[2]6 cicles of 30″ with 60″ intervals; 25W-50 amplitude

dation for Statistical Computing, Vienna).

CONCLUDING REMARKS

In our studies on *B. subtilis* biomineralization we tried to extend the understanding of the mechanism(s) relying on $CaCO_3$ precipitation.

In CHAPTER 3 we have investigated how $CaCO_3$ precipitation is affect by *etfA* over-expression. Investigation of changes in pH on solid media of strain over-expressing *etfA* have revealed that over-expression of the *etfA* gene may be responsible for the alkaline shift observed in that strain (FBC16 1mM IPTG, figures 3.6 and 3.4). When *etfA* is over-expressed, we have reported that calcite precipitates became visible 4-5 days before WT strain 168 and controls, figure 3.6. A further significant finding of the work is that expression of *etfA* is able to restore the bio-calcification phenotype in the EtfA-disruptant FBC5 [7], rather than in the $FadR_{Bs}$(YsiA) mutant FBC2, figure 3.7. Although past evidence lead to hypothesize that *B. subtilis* bio-calcification is relied to solid medium only [74], we investigated changes in pH of liquid cultures of induced and un-induced strain FBC16 and compare to those of WT strain 168. Surprisingly we found that shift toward alkaline values in the WT strain 168, which lead to $CaCO_3$ precipitation, occurs 7-10 days before induced FBC16 and its control, figure 3.5. This delay seems to be related also to calcite precipitates which became to be visible 7-8 days after the WT strain 168 in both induced FBC16 and control, figure 3.5.
Taken together the results of this research provide some clues toward the understanding of the complex mechanism(s) involved in biomineralization. This findings account for local pH regulation as the major forces driving calcite precipitation on solid media. In this mechanism seems to be essential to create the proper condition for calcification the *etfA* gene

encoded by the *lcfA* operon. The lack of *etfA* gene has been found to be responsible for the lack of calcite precipitation in strain FBC5 due to acidification of the medium, figure 3.2, while *etfA* over-expression resulted in early alkalinyzation of the media, figures 3.4 and 3.6. As proposed for the *etfA* disruptant, in which the interruption of *etfA* may cause a misregulation of the EtfA pathway, over-expression of *etfA* may cause a deregulation of the EtfA pathway with opposite effects to that reported for FBC5.
Over-expression of *etfA* in the mutant background of FBC2 did not result in marked pH changes and did not completely restore the bio-calcification when compared to the effects observed in FBC5 mutant, figure 3.7. This findings account for the involvement of other genes encoded by the *lcfA* operon, *i.e.* EtfA cannot exert its function without its counterpart EtfB. Few papers have described the involvement of the sole EtfA in diverse processes, such as the DPA synthesis in *C. perfringens*, even-though in the majority of the processes involving EtfA it is essential to accomplish its function its counterpart EtfB. Therefore we have cloned the *etfBA* genes in order to investigate if bio-calcification could be relied on the heterodimer EtfB-EtfA, in which the pivotal role may be played by EtfA. At the moment we have achieved the *E. coli* strain harboring the *etfBA* recombinant plasmid, and confirmed the nucleotide sequence of the two genes through nucleotide sequencing. Then we have planned to transform the mutant strain FBC2 in order to investigate the effect of *etfBA* over-expression on bio-calcification and changes in pH. The results of these experiments may extend the understanding of a part of the complex process which is bio-calcification, and could also be useful in defining the minimal set of genes required for *B. subtilis* biomineralization.

Investigations in CHAPTER 4 have explored the expression of the *lcfA* operon genes and the timing of their activation along with biomineralization. Moreover we have investigated how perturbation of environmental parameters affect expression and bio-calcification. It has been reported that the expression of the *lcfA* operon genes is an indispensable requirement for biomineralization, and it is mainly controlled by two regulative regions. P_{lcfA12}, the promoter upstream of the whole cluster, and P_{ysiA} an internal promoter leading to the transcription of the *ysiA, fadB, etfB* and *etfA* genes, figure 4.1. In Barabesi et al., 2007 [7] has been reported that in strain FBC1 where transcription from P_{lcfA} is interrupted, biomineralization was not affected. Therefore it is clear that transcription starting from P_{ysiA} is that essential for proper $CaCO_3$ mineralization. Preliminary *in silico* analyses have showed the presence of at least two additional putative

regulative regions within the *lcfA* operon, located up-stream P_{lcfA} and in the inter-genic region encompassing *fadB* and *etfB*, Table 4.3.
In order to study the transcriptional potential of these regulative regions, we cloned such fragments in front of a promoter less copy of the *lacZ* gene carried by the ectopic integrational vector pJM116, figure 4.2. One of the main goals of this investigations would have been to look for a spatial pattern of expression in the *B. subtilis* streak, possibly reflecting that of calcite precipitates. Our observations showed that expression of these genes did not reflect the uneven distribution of calcite crystals. Indeed expression of the genes seems to be homogeneously distributed all over the streak, figures 4.3, 4.4, 4.5, 4.6, and 4.7. Our research has established the timing of activation of the regulative regions within the *lcfA* operon, rather than changes in the expression during the whole incubation, due to the transcriptional fusion system chosen for this study. Expression by the two alternative P_{lcfA} occurs within the first 24 h, while that controlled by P_{ysiA} arises after 48 h. Qualitative observation showed a different intensity of the expression in each of the regulatory region tested (*intensity:* FBC12$\geq$FBC13$\geq$FBC11, figures 4.3, 4.4, 4.5, 4.6, and 4.7), except for that found in the inter-genic region between *fadB* and *etfB*. Moreover expression of *ysiA*, *fadB*, *etfB* and *etfA* genes trough P_{ysiA} seems to be influenced by temperature and calcium availability rather than by pH, figures 4.3 and 4.7, indeed that of the two alternative P_{lcfA} was found to be unaffected by perturbation of environmental parameters. Temperature affects the expression by P_{ysiA} increasing its activity along with temperature.
As regards calcite precipitation, it occurs within 2-4 days (42°C, [Ca-acetate] $\geq$ 0.75%, and in buffered pH 8.2 B4), while in strains cultivated at 37°C and 0.25% $\leq$ [Ca-acetate] $\leq$ 0.75% precipitation occur within 6-8 days of incubation. When timing of the activation was compared to that of biomineralization, no obvious connection between expression and calcite production can be found. The activity of the two alternative regulative regions up-stream *lcfA* start 4-5 days in advance than the development of calcite precipitates, while that of the region up-stream *ysiA* 3-5 days before the appearance of first $CaCO_3$ crystals. Next steps in this investigation will involve other methods to quantitative measure the expression and the contribution of the *lcfA* operon genes in biomineralization. In order to characterize also small expression changes of these genes along with biomineralization we have planned to perform qRT-PCR experiments on the transcripts of the *lcfA* operon. Indeed to study the contribution of these genes and perhaps to determine the half-life of the *lcfA* operon genes products in biomineralization, it would be interesting to perform western blotting experiments with antibody against each gene product within the operon.

Finally in CHAPTER 5 we started a functional study to determine the function of the *lcfA* operon gene products, in order to characterize their influence in biomineralization. The function of the putative translational products of the *lcfA* operon genes are related to fatty acids metabolism, figure 5.1. They are considered as the *B. subtilis* functional homologs of *E. coli fad* genes. However, except for YsiA recently characterized as the FadR ortholog of *E. coli* [75, 53], their function are only based on database similarity search.
Characterization of these gene products has involved the cloning and the purification of recombinant proteins as well as database search to find specific domain features. We reported that YsiB possesses an ECH domain common to all the enoyl-CoA hydratase, figure 5.2. Interestingly we found in EtfA, thought to play a pivotal role in bio-calcification, an EF-hand like domain which might interact with calcium ions, figure 5.5. Therefore we have planned to characterize the function of the newly domain found in EtfA and to study its involvement in *B. subtilis* biomineralization.
At the moment we have characterized YsiB, now FadB, as acting as an enoyl-CoA hydratase. Studies on the calcium binding domain (CBD) found in EtfA are still in progress, at present time we have purified the recombinant EtfA, figure 5.6, next steps will be to test the ability of the CBD to bind Ca^{2+} by using a fluoresce detection method, as that proposed by Tatsumi et al., 1997 [96]. Results we will obtain by this investigation could be helpful to better understand the role of EtfA in calcite precipitation.

6

COMPARATIVE GENOMICS OF VIRR REGULONS IN *Clostridium perfringens strains*

6.1 BACKGROUND

Clostridium perfringens is a Gram-positive anaerobic species able to form heat-resistant endospores and to live in many habitats, from marine sediments to animal gut, to soil. The genus *Clostridium* comprises species causing severe diseases such as botulism, tetanus, gas gangrene and pseudomembranosus colitis that are generally due to the secretion of powerful toxins. *C. perfringens* is the most prolific toxin producer within the genus; several of its extracellular toxins and enzymes have been identified as for instance α-toxin (plc, phospholipase C), β-toxin (hemolysin family toxin), ϵ-toxin, θ-toxin (pfoA), κ-toxin (colA, collagenase) and others. Toxins are thought to act synergistically in the development of pathogenesis, and *C. perfringens* strains show a high degree of phenotypic and pathogenic variability, so that understanding the control of the expression of toxin genes is critical to help in fighting diseases caused by this bacterium. The identification of similarities and differences in the set of pathogenic instruments (i.e. genes) of different strains will help to define effective strategies of infection control. Pathogens usually have precise control mechanisms for toxin production so that expression only takes place when required e.g. when the density of the bacterial population overcomes a certain threshold, or when the bacterium reaches a certain cell-type/organ. In bacteria, quorum sensing and environmental signal detection and transduction depend on the activity of dedicated two component systems consisting of a membrane bound sensor histidine kinase and a response regulator. The kinase activity of the sensor is activated by specific signals, trigger-

ing phosphorylation of the cognate response regulator. The phosphorylated regulator then actively changes gene expression of its target genes through binding of specific DNA motifs [115]. In *C. perfringens* a major role in integrating environmental signals with virulence competes to the two-component VirR/VirS system, where VirR is the response regulator and VirS the membrane anchored sensor protein [37] figure 6.1. The first VirR regulated promoters have been located upstream of toxin genes [5] and subsequent works showed that VirR target sequences are formed by a pair of imperfect direct repeats, separated by 7-8 nucleotides (depending on how the repeat is defined) [39]. These repeats are known as VirR box1 and VirR box2 (VB1 and VB2) and are located within a core region of about 50 base pairs located immediately upstream of the -35 element of the promoter of regulated genes. The two VirR boxes are both required for VirR mediated transcriptional activation, and mutation of either of them drastically reduces the expression level of target genes. The binding of VirR to its boxes is required for the efficient positioning of the RNA polymerase to the promoter. Furthermore in all the upstream regions of genes directly regulated by VirR, the two boxes are in the same relative position with respect to the promoter and are on the same face of the helix. DNA spacing and helical phasing play a crucial role in the transcriptional activation by VirR, as demonstrated by the insertion or deletion of 5 base pairs in the region between VB1 and VB2 that displaces them on opposite faces of the DNA double helix: in this situation a pronounced reduction of the expression level of genes controlled by VirR was observed [38].

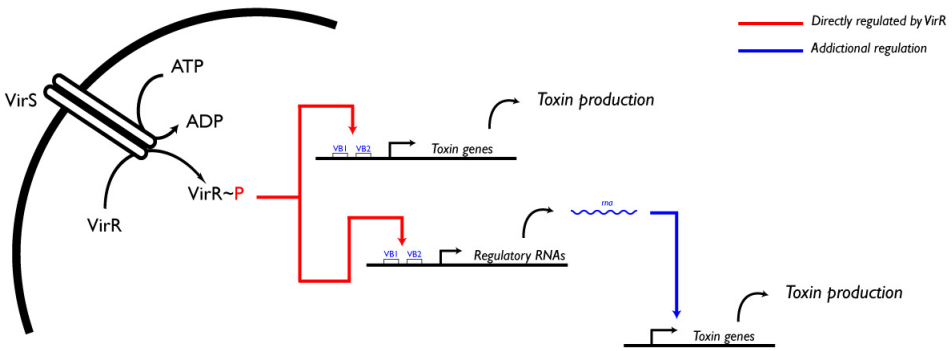

Figure 6.1: The two component system VirR/VirS and its cellular targets are here schematically represented.

The VirR/VirS regulatory network is not only involved in direct control of toxin encoding genes figure 6.1, but also of several other genes such as *hyp7* (*vrr*) a gene encoding a regulatory RNA (VR-RNA) which controls the rate of transcription of *colA*, *plc*, *ptp* (protein tyrosine phosphatase) and *cpd* (encoding 2′,3′-cyclic nucleotide phosphodiesterase) [101]. A recent paper dealing with the *in silico* identification of VirR regulated promoters in *C. perfringens* str. 13 followed by experimental validation, allowed to identify additional direct VirR targets, namely *virT*, *virU* and *ccp* (α-clostripain gene) [82]. The former two genes are particularly interesting because they are regulators of gene expression. Two genes only appeared to be controlled by *virT* (*pfoA* and *ccp*), while *virU* is active with respect to *pfoA*, *ccp*, *hyp7*, and *virT*. A mutational analysis revealed a clear parallel with what observed for *hyp7*, because the gene expression level of their targets is unchanged in *virT* or *virU* nonsense mutants, with respect to the wild-type, allowing to conclude that the functional forms are the *virT* and *virU* RNA [81]. Moreover, three additional genes regulated by VirR and coding for hypothetical proteins, were found in different *C. perfringens* strains: CPF_1074, CPF_0461 in *C. perfringens* ATCC13124 and CPR_0761 in *C. perfringens* SM101 [80]. It is now clear that the two component VirR/VirS system is at the top of a hierarchical regulatory cascade where it directly stimulates the transcription of several virulence-related genes including three different regulatory RNAs that are in turn able to control several other genes [101]. Because of the large heterogeneity in toxin production by *C. perfringens* strains [**?**], it is interesting to define the genes belonging to the direct VirR regulon in closely related genomes to assess the degree of evolutionary conservation of the VirR regulon. This could also clarify the evolutionary patterns that are at the basis of the divergence between these strains from a common ancestor. However the experimental strategy cannot be easily implemented for all strains, so that it is necessary to integrate information from different strains in a bioinformatics protocol. In this work we extend the bioinformatic approach of [81] to scan the genomes and plasmid sequences of all available genomes of *C. perfringens* strains (Table 1), and identify genes that are putatively controlled by the VirR/VirS system. We implemented a two step strategy allowing to consider information concerning VirR binding sites in all these genomes and defining the core (evolutionary conserved) and accessory (strain-specific) VirR regulons in different strains. Results obtained could help to better define strategies for pathogenicity studies and control strategies in *C. perfringens* and can moreover be used to design focused wet-lab experiments.

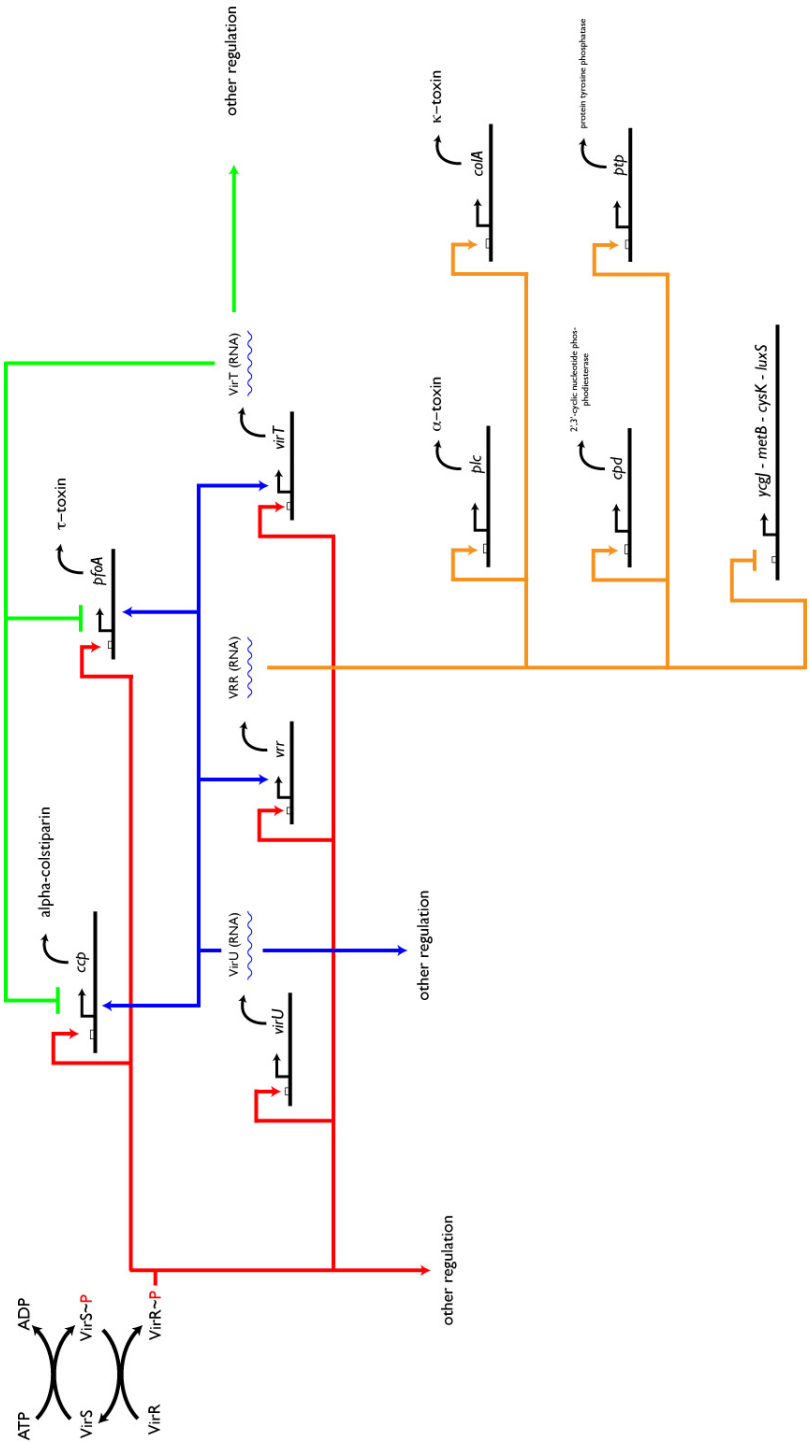

Figure 6.2: *V*irR regulon in *C. perfringens str. 13*

6.2 EXPERIMENTAL PROCEDURES

Binding sites identification

To identify motifs corresponding to the binding site of VirR we devised the following strategy (illustrated in Figure 6.4). Using experimentally validated VirR targets (CPE0163, CPE0846, CPE0845, CPE0920, CPE0957 from [81] and CPF_1074 and CPF_0461 from [**?**]), we derived a position weight matrix describing the region encompassing the VirR box 1 and 2, for a total of 34 nucleotide positions. This matrix was used for a first scanning of the genomes, allowing to retrieve all orthologous to known targets in other strains. All the motifs identified upstream of known targets were then used to build a second PWM that was used for a second round of genome scanning to identify candidate VirR targets. Genome scanning was performed with a sliding window approach from first nucleotide to genome length - L, where L is the motif's length. Each 34-mer was scored using the function proposed in [**?**]:

$$S_i = \frac{1}{L} \sum_j \left[2 + log_2(F_{ij})\right],$$

where F_{ij} is the frequency of the ith base at the jth position. S_i is an information-based measure of potential binding sites. We retained only motifs having a score larger than or equal to the lowest score for an experimentally validated target, corresponding to a threshold of 0.88. Each motif found along the genome was then associated with a gene when located within the region going from 100 nucleotides downstream to 600 nucleotides upstream of the corresponding first codon and on the same strand of the motif.

Clustering protein sequences

Protein sequences of candidate targets were clustered using the MCL algorithm coupled with Blast2Network [25], whose source code was changed accordingly. Blast2Network performs an all-against-all Blast comparison of protein sequences, followed by translation of the results in an adjacency matrix. From here on we changed the B2N code to allow the use of the MCL with a similarity measure corresponding to the normalized alignment bit score between homologous sequences:

$$S_{ij}^{norm} = \frac{S_{ij}}{S_{ii}},$$

where S_{ii} is the maximal score attainable using the ith query and it corresponds to the query aligned with itself. The adjacency matrix is normalized to make it stochastic, a prerequisite for the MCL algorithm used to define clusters of orthologous sequences. The MCL algorithm simulates flow alternating two algebraic operations on matrices: *expansion* of the input matrix ($\mathbf{M^{out}} = \mathbf{M_{in}} * \mathbf{M_{in}}$) models the spreading out of flow and *inflation* ($m_{ij} = m_{ij}^r$). Parameter r controls the granularity of the clustering and it is set to 2. After these two steps we apply diagonal scaling to keep the matrix stochastic and ready for the next iteration. Inflation models the contraction of flow, and it is thicker in regions of higher current and thinner in regions of lower current. The consequence is that the flow spreads out within clusters while evaporating in-between clusters leaving at convergence an idempotent matrix revealing the clusters hidden in the original adjacency matrix.

Plasmid analysis

Concerning the identification of VirR targets, we analysed plasmids with the same procedure used ofr genomes. Phylogenetic profiling and the hypergraph describing the similarity in gene contents of different plasmid molecules were calculated using the software Blast2network [25] and visualization with the software Visone [110]. The phylogenetic profiling technique is described in detail in several papers, e.g. [86, 41] so that we will not discuss it here in detail, it is enough to say that by comparing the distribution of different genes in different plasmids we can quantify the extent at which proteins tend to co-occur which is an indication of the degree of functional overlapping between different proteins. We want to spend some word concerning the hypergraph shown in Figure 6.7. Let's suppose to have an adjacency matrix describing homologies between proteins encoded by several different plasmids. In this matrix, element m_{ij} corresponds to the similarity between sequences i and j. However these matrices can be quite large (i.e. the total number of proteins in the study set), so that it is possible to apply some dimensionality reduction approach to extract the information we are interested in. In our case, given the mobility of genes encoded on plasmids, we wanted to assess the degree of similarities between them in term of gene content, and to identify the most plausible routes for gene exchange in the strains under analysis. One way to do that is to calculate the similarity in the phylogenetic profiles of each plasmid and then reduce the original matrix to a new one whose size corresponds to the number of plasmids in the dataset. In this new matrix, the values correspond to the similarity in gene content between every pair of plas-

mids. Given the binary nature of phylogenetic profiles calculated by B2N, it is possible to to quantify the level of similarity between them using the Jaccard similarity coefficient. Plasmids with highly similar gene content will then give very tight clusters, and plasmids in-between different clusters (sharing some of their genes with plasmids in one clusters and some other genes with an otherwise unrelated cluster of plasmids) could be important because they share genes with different molecules i.e. they could represent preferential routes for the passage of genes between plasmids that are not in contact.

Alignments and Phylogenetic analysis

The alignment of *rrnA* operons was performed using the software muscle with default parameters. The alignment has a total of 4719 nucleotides, 32 of which are variable, and was used as input to the software mega to build a phylogenetic tree. The algorithm used was the Neighbor-Joining with different rates for transitions and transversions and 100 bootstrap replicates.

Comparison of intergenic sequences

The comparison of intergenic sequences was performed as follows: all intergenic sequences were extracted from the genome of Str. 13 using gene annotations and were then filtered for a minimum length of 100 nucleotides, obtaining 1633 sequences. These sequences were then blasted against the other genomes. We retained each first blast hit when the e-value of the alignment was less then 1E-06. The boxplots shown in have been obtained for the totality of matches for a genome.

6.3 Results and Discussion

6.3.1 *Comaprison of Clostridium perfringens strains*

As a preliminary analysis we studied the variability of the selected genomes using both standard phylogenetic techniques and a comparison of all intergenic sequences. The alignment of *rrnA* operons for a total of 4719 nt was used to build a Neighbor-Joining tree revealing that these strains are closely related. In agreement with a low differentiation on ribosomal operon sequences, bootstrap support for the branching pattern was quite low; in fact, 32 variable sites only were found in the alignment, which were evenly distributed between strains. However, the comparison of a large number of intergenic sequences extracted from the genomes revealed that some of them are quite variable between the different strains with respect to the very conserved *rrnA* operon (down to 82% with respect to *C. perfringens* Str. 13.

6.3.2 *Regulon prediction in sequenced* C. perfringens *strains*

We scanned available *C. perfringens* genomes using the VirR position weight matrix (PWM) derived from experimental observations, following the procedure reported in Figure 6.4 (see Methods for details).

At the time we performed this analysis (April, 2009), the NCBI microbial genome database stored three different complete genomes for *C. perfringens* corresponding to strains 13, ATCC 13124 and SM101, plus the draft genomes of five strains (ATCC 3626, JGS1721, JGS1987, F4969, JGS1495) in the form of whole genome shotguns (Table 1). Despite the higher probability of errors in gene assignments characterizing draft genomes, we decided to include them to expand the scope of our genomic comparison. A whole genome scanning was performed using a PWM derived from the region comprising several experimentally validated VirR binding sites [82, 80]. A new PWM was generated from the targets identified in the first scanning by using 30 motifs found in the promoters of genes that are orthologous to known targets and then used for a second genome scanning. In this way we avoid the biases that affect the first matrix, obtained from only a few sequences mainly coming from one strain. After our two-step strategy, we collected all genes with a motif scoring more than 0.88, which is the lowest value observed for an experimentally tested VirR target gene (corresponding to gene CPF_1074, [80]). At this threshold we retained at end 53 occurrences of the VirR motif. Most of them were distributed in one peak only, present at around 100 bp from the beginning of the gene

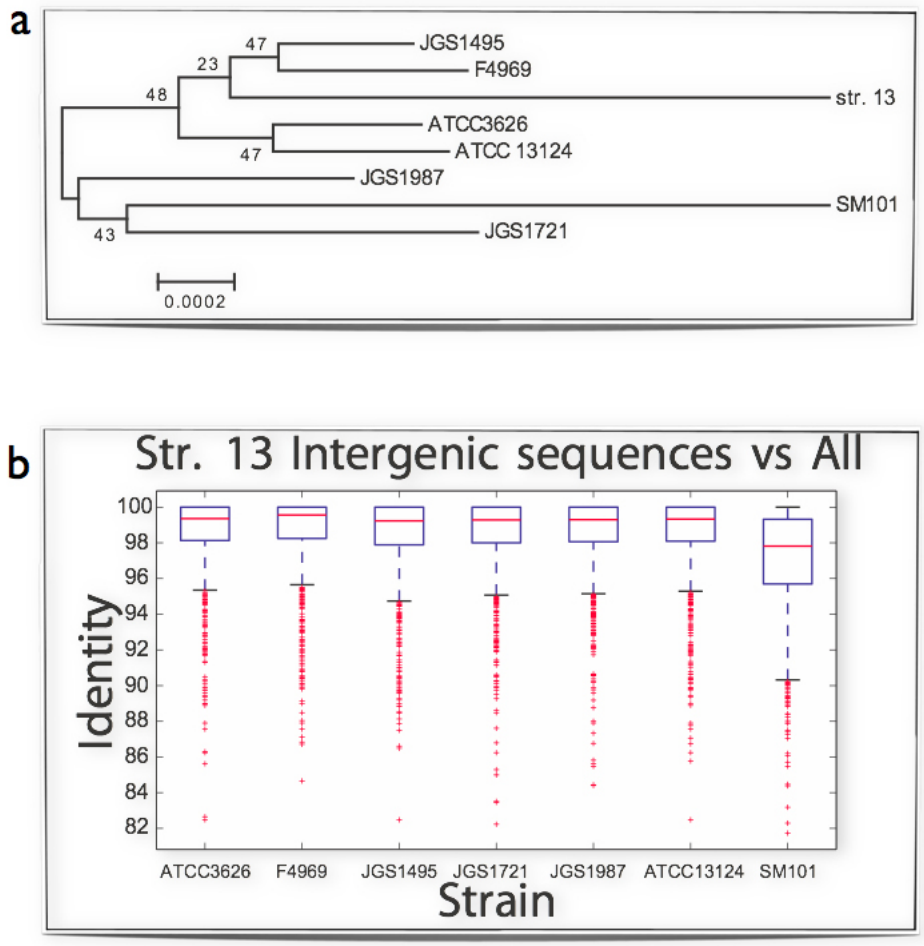

Figure 6.3: a) Phylogenetic tree of *rrnA* operons of the eight strains used. Numbers at the nodes indicate bootstrap support on 100 total replicates. The bar at the bottom is in substitutions per site indicating a very low variability of *rrnA* operons. b) Number of differences between strains confirming the previous observation. c) Boxplots summarizing the variability of the intergenic sequences of seven strains with respect to Str. 13. All intergenic sequences were extracted from the genome of Str. 13, filtered to retain only those longer than 100 nt and blasted against the other genomes using an E-value threshold of 1E-06.

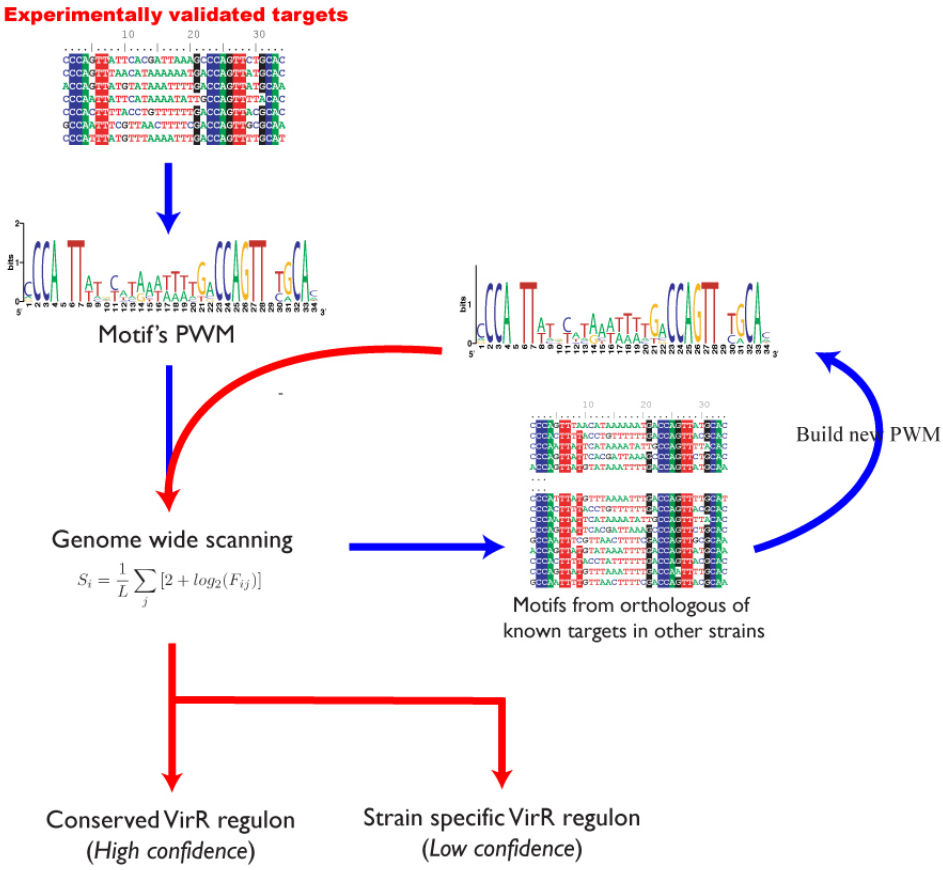

Figure 6.4: As discussed in Methods, we performed a two step strategy that allowed to use information coming mainly from a single strain to do the predictions on all other strains in a less biased way. In the first step a position weight matrix (PWM) calculated from a limited number of experimentally validated motifs. This PWM is used to scan the genome to make a list of possible targets. Within that list we looked for orthologous to known targets, retrieved their motifs and built a second PWM. This includes the variability of the motif in several strains and was used for the final scan of the genomes.

Figure 6.5. The finding that in some case a larger distance between ATG and VirR boxes was observed may be due to differences in transcription start sites (with larger 5′ untranslated regions) or may account for some different level of regulation for those genes.

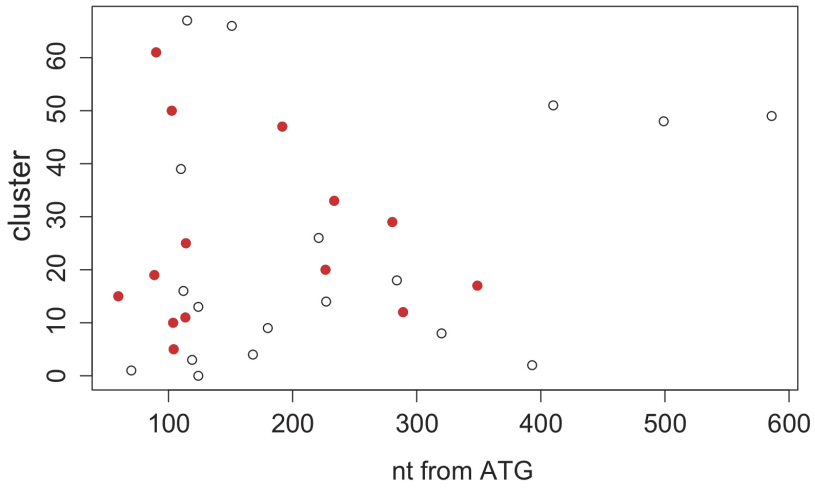

Figure 6.5: Here we show the distance of the motifs identified with respect to the translation start site (X-axis), grouped by homology of the downstream gene (cluster identifier on the Y-axis). Most of the targets are located in the first 200 nt from the start of the gene, but some of them (and notably several corresponding to characterized ones) are located at larger distances. Red circles correspond to clusters containing several genes.

The list of genes putatively regulated by VirR was splitted in three different groups after clustering similar sequences (see Methods), by defining the : i) conserved VirR regulon as formed by chromosomal genes retrieved in at least two different genomes; ii) the accessory regulon with chromosomal genes present in a single strain; iii) the mobile regulon, including genes found on plasmids.

The conserved VirR regulon

The conserved regulon (Table 2), appeared to contain all known target genes [82, 80] except for CPR_0761 and *virT*. The former can be identified in the genome of strain SM101 only, while the latter has been found in strain 13 and ATCC3626; in both cases we were able to identify a VirR binding motif in their promoter (Table 4.x).

Table 6.1: Genomes and plasmids analyzed List of genomes and plasmids used in this study. The *Type* column indicates if a sequence is a genome (G) or a plasmid (P) in that case we also indicate the name of the plasmid within round parentheses.

Organism	Type (name)	Sequencing Status	Genes	Length (nt)
Clostridium perfringens 13	G	Finished	2905	3085740
Clostridium perfringens ATCC 13124	G	Finished	3066	3256683
Clostridium perfringens B ATCC 3626	G	Draft	3427	3896305
Clostridium perfringens C JGS1495	G	Draft	3254	3661329
Clostridium perfringens CPE F4969	G	Draft	3118	3510272
Clostridium perfringens D JGS1721	G	Draft	3485	4045016
Clostridium perfringens E JGS1987	G	Draft	3729	4127102
Clostridium perfringens NCTC 8239	G	Draft	2925	3324319
Clostridium perfringens SM101	G	Finished	2748	2921996
Clostridium perfringens	P (pBCNF5603)	Finished	36	36695
Clostridium perfringens	P (pCP8533etx)	Finished	63	64753
Clostridium perfringens F4969	P (pCPF5603)	Finished	73	75268
Clostridium perfringens F5603	P (pCW3)	Finished	51	47263
Clostridium perfringens F5603	P (pCPF4969)	Finished	62	70480
Clostridium perfringens SM101	P (1)	Finished	10	12397
Clostridium perfringens SM101	P (2)	Finished	11	12206
Clostridium perfringens str. 13	P (pCP13)	Finished	63	54310

One target only appeared to be conserved in all tested strains, corresponding to the α-clostripain gene. Four genes were shown to be conserved in all strains except for strain SM101. Interestingly, strain SM101 appeared to have the lowest degree of conservation of VirR targets. A search for the corresponding gene sequences in the genome confirmed that they are absent, in agreement with a previous comparative analysis that showed that in this strain several virulence factors and toxins are missing while being present a specific repertoire of genes encoding bacteriocins [80]. On the converse, missing genes in draft genomes cannot be considered as surely absent. Concerning CPE0920 (*virU*) and CPF_1074, corresponding to a regulatory RNA encoding gene and to a gene with unknown function respectively, they have not been identified in some of the genomes, but using their sequences we were able to identify regions with perfect matching using blastn (data not shown) and to locate VirR motifs in their upstream regions (see Table 4.x). Myers et al. [80] showed that purified VirR is able to bind the promoter of CPR_0761 and of CPF_0461. From our analysis it emerged that CPF_0461 in str. ATCC1324 is the ortholog to CPR_0762 in str. SM101, for which too we predicted the presence of a VirR binding motif upstream. This motif is the same attributed to CPR_0761 and whose ability to bind VirR has been tested by Myers et al., 2006. Our comparative analysis, then suggests that the truly regulated gene could be the latter, because of the conservation of the site upstream of its homologs in two other organisms (ATCC3626 and ATCC1324), while we were not able

to find sequences resembling CPR_0761 in any other *C. perfringens* strain by blasting both protein and nucleotide sequences against their genomes. Alternatively, the two genes can also form an operon, with CPR_0761 performing an unknown function.

The accessory VirR regulon

We consider this dataset low confidence for two reasons: first of all this group of genes comprises only one experimentally verified target, i.e. *virT* (CPE0845, [82]) and moreover, all other genes have been found in draft genomes only. The list of all putative targets of VirR is shown in Table 4.x. Notably, JGS1987 is characterized by an expansion of the VirR predicted regulon, while the accessory regulon of ATCC3626, F4969 and SM101 strains is composed of a single gene. The case of *virT*, a regulatory RNA, is particularly interesting. This sRNA implements a negative feed-back loop on some of the VirR targets i.e. *pfoA* and *ccp* [82]. Our analysis showed that *virT* is present in two strains only (strain 13 and strain ATCC3626). We can thus predict that the other strains lack this negative control and express *pfoA* and *ccp* at different levels eventually by using additional regulations. Actually, strains as ATCC 13124 produces large quantities of gangrene-associated toxins [79] and JGS1987 is a Type E strain which, tough containing an enterotoxin gene (*cpe*), did not show enterotoxin production [99]. The relatively large predicted regulon (10 genes) of JGS1987 may contain genes responsible for its peculiar pathogenicity profile. Within such regulon seven genes coding for proteins of unknown function have been found. One of them corresponds to a resolvase/recombinase (AC3_0180) suggesting a possible scenario in which host invasion is linked to gene mobilization. The other two genes with assigned function in the putative regulon of strain JGS1987 include a 2-keto-3-deoxygluconate kinase and a putative lipid A export permease. The first one has been associated with resistance to oxidative stress in *C. perfringens* mutants after transposon mutagenesis [27]. Concerning the putative permease of lipid A, it is known that lipid A is one of the main mediator of bacterial pathogenesis and strongly stimulates inflammation in host tissues [67], so that our prediction is reasonable.

The 'mobile' VirR regulon

Our analysis identified three targets located on plasmids, one coding for ϵ-toxin (pCP8533etx_p28) in plasmid pCP8533etx from strain NCTC 8533B4D,

Table 6.2: The genes that scored positively with respect to the VirR position weight matrix and that were found in at least two genomes. Numbers below each gene name correspond to the score calculated as described in Methods (on a maximum attainable score of 1.52). As described in the text, most of the known VirR targets belongs to this group. * no open reading frame identified in this region but DNA sequence identical to CPE0920; ** no open reading frame identified in this region but DNA sequence identical to CPF_1074.

Gene	**Product**	**Genomes**								**REF**
		ATCC 13124	**Str.13**	**SM101**	**F4969**	**JGS1721**	**JGS1495**	**JGS1987**	**ATCC 3626**	
ccp	**α-clostripain**	CPF_0840 1.52	CPE0846 1.52	CPR_0833 1.52	AC5_0918 1.52	CJD_0991 1.52	CPC_0878 1.52	AC3_1028 1.52	AC1_0991 1.52	Okumura et al., 2008
vrr	**Reg. RNA**	CPF_1204 1.31	CPE0957 1.31		AC5_1228 1.31	CJD_1316 1.21	CPC_1202 1.31	AC3_1326 1.31	AC1_1388 1.31	″
pfoA	**Perfringolysin O**	CPF_0156 1.18	CPE0163 1.18		AC5_0210 1.18	CJD_0196 1.18	CPC_0186 1.18	AC3_0278 1.18	AC1_0175 1.18	″
virU	**Reg. RNA**	CPF_0925 1.20	CPE0920 1.20		* 1.26	CJD_1073 1.26	* 1.20	AC3_1102 1.20	AC1_1131 1.20	″
	Hypothetical	CPF_1074 0.88		CPR_0937 1.11	** 1.03			** 1.03	** 1.03	Myers et al., 2006
	hypothetical	CPF_0461 1.28		CPR_0762 1.38					AC1_0537 1.28	″
	hypothetical				AC5_0209 1.18139196		CPC_0185 1.18139196			
virT	Reg. RNA		CPE0845 1.2977						AC1_0990 1.2977	Okumura et al., 2008

Table 6.3: Strain specific VirR targets Genes identified as VirR targets that are present in a single strain. Strain JGS1987 suggests an expansion of the VirR regulon.

Product	Gene	Score	Strain
2-keto-3-deoxygluconate kinase	AC3_0259	1.264524624	JGS1987
hypothetical protein AC3_0622	AC3_0622	1.166056351	JGS1987
hypothetical protein AC3_A0724	AC3_A0724	1.039200045	JGS1987
hypothetical protein AC3_A0725	AC3_A0725	1.039200045	JGS1987
conserved hypothetical protein	AC3_A0081	1.11347037	JGS1987
resolvase/recombinase	AC3_0180	1.153220148	JGS1987
putative lipid A export ATP-binding/permease protein MsbA	AC3_0181	1.153220148	JGS1987
hypothetical protein AC3_A0587	AC3_A0587	1.344587411	JGS1987
hypothetical protein AC3_0277	AC3_0277	1.18139196	JGS1987
hypothetical protein AC3_A0194	AC3_A0194	1.253557944	JGS1987
hypothetical protein AC1_A0478	AC1_A0478	0.803945302	ATCC 3626
hypothetical protein AC5_A0236	AC5_A0236	1.037739473	F4969
putative metal-dependent hydrolase, membrane-bound	CPR_1028	1.341272427	SM101
hypothetical protein CJD_0545	CJD_0545	0.951310873	JGS1721
hypothetical protein CJD_1387	CJD_1387	1.300913548	JGS1721

in addition with two hypothetical proteins, sharing 98% identity, in pCP8533etx (pCP8533etx_p40) and in pCPF5603 (pCPF5603_50) of strain F5603, respectively. Concerning plasmid pCP8533etx, we noticed that it is also present in the shotgun sequences from ATCC3626 (data not shown based on blastn comparisons) and also in that case we were able to find a VirR motif upstream of the gene encoding ϵ-toxin.

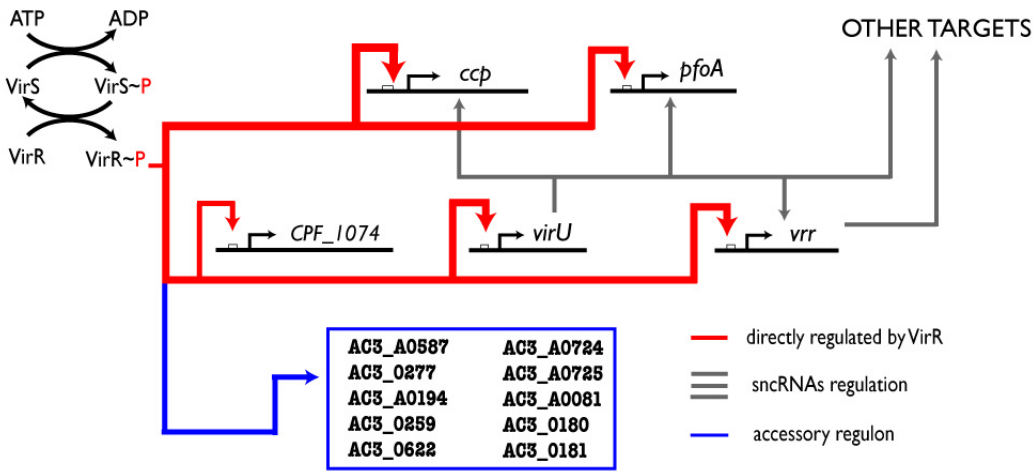

Figure 6.6: Accessory VirR regulon of strain JGS1987 type E

6.3.3 *Plasmid analysis*

Plasmids can be transferred between species, and gene content similarities between plasmids can be used to trace gene flow between different strains. To evaluate evolutionary relationships relating plasmids from *C. perfringens* species, we performed an analysis to quantify the number of genes shared by each pair of plasmids. For this reason, we built the phylogenetic profiles of the proteomes encoded by plasmids in these strains. The phylogenetic profiles for each group of proteins were obtained by comparing all those proteins one against each other with the package Blast2Network [26]. A phylogenetic profile, or phyletic pattern, is represented by a matrix where each row corresponds to a plasmid molecule and each column to a given protein family. The cell at the intersection between row i and column j indicates the presence of a component of protein family j in plasmid i. A phylogenetic profile can be thus interpreted as a graph with two types of nodes: those corresponding to plasmid molecules are connected to nodes of protein families if the corresponding plasmids contains the gene encoding that protein. These matrices can become very large when many plasmids and proteins are involved, so that their analysis and biological interpretation is difficult. A strategy for dimensionality reduction can be through deletion of nodes corresponding to protein families and connection of plasmids directly, through edges that reflect the number of shared protein families. The obtained *hypergraph* is reported in Figure 6.7, where plasmids are connected by links weighted on the basis of the number of common genes. A group of four connected plasmids (i.e. sharing several genes), including pCP8533etx and pCPF5603, was found. This finding is in agreement with previous data showing that plasmids pCPF5603 and pCP8533etx evolved from insertion of mobile genetic elements carrying enterotoxin or *etx* genes, respectively, onto a common progenitor plasmid [78]. This group of plasmids is connected to a second group, composed of three plasmids (plasmid 1, plasmid 2 and pBCNF5603) through a *bridge* represented by pCP13. This implies that pCP13 shares different genes with plasmids from both groups i.e. it may be considered as an evolutionary link between the two groups, one including plasmids with putative VirR targets and the other one with no targets. Plasmids pCP13 and pBCNF5603 seem to have acquired regions from different sources during evolution. Interestingly pCPF5603, belonging to the first group, and pBCNF5603 have been isolated from the same strain, but do not share common regions. The gene encoding the enterotoxin (*cpe*) is only present in pCPF5603 and pCPF4969 and the link from pCP13 to pCP8533etx and pCPF5603 comprises the gene encoding β-toxin. These data confirm and extend the de-

tailed analysis performed by [77]. They observed that plasmids pCPF5603 and pCPF4969 share a region of about 35kb that it is not present in pCP13. From our analysis it emerged that the genes comprised in that region could be conserved also in plasmids pCP8533etx and pCW3.

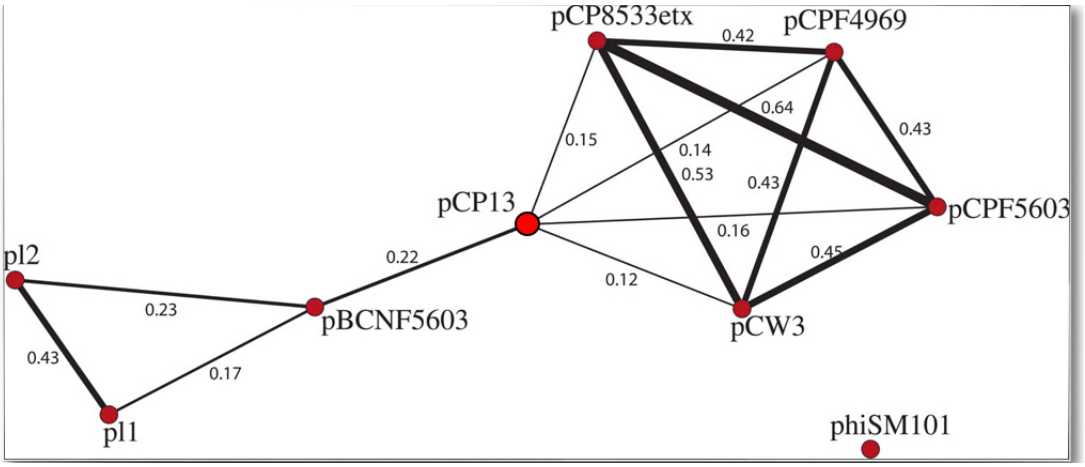

Figure 6.7: Here we applied a basic analysis available in the Blast2Network package (*Brilli et al., 2008*) that consist in comparing different plasmids on the basis of their gene content (*see Methods for a concise description of the methodology*). Values connecting nodes (plasmids) correspond to the percentage of shared genes with respect to the total number of genes of the two plasmids (Jaccard coefficient) and can be considered as a measure of relatedness in terms of evolutionary history (common ancestor) and horizontal transfers and recombination. After this analysis, two groups of plasmids emerged, that are connected through the edge between plasmids pBCNF5603 and pCP13. In the right group of plasmids we identifed some VirR targets. The high similar gene content of some of the plasmids in that group may suggest a high rate of horizontal transfer between different strains, so raising the possibility of the transfer of the VirR targets. Moreover, the connection between the two groups can also suggest that transfers between the two groups of plasmids can happen.

6.4 Conclusions

In this work we exploited experimental information concerning a small number of promoters controlled by VirR to predict the corresponding regulons in all other *C. perfringens* genomes and plasmids available. Our results are in agreement with previous analysis and suggest that the size of

the VirR regulon is quite variable in the analyzed strains as also evidenced by works showing that these strains encode different repertoires of toxin genes. Particularly interesting are the cases concerning *vrr*, *virU* and *virT*, because they encode regulatory RNA that affect gene expression of several other genes. Thus, even at the short phylogenetic distances spanned by these strains (Fig. S2), there could be significant changes in the regulatory cascade initiated by VirR. An event of gain or loss of a VirR target can affect the gene itself only, such as when the event involves a gene coding for a *toxin*, or it can spread downstream of VirR when it involves a regulatory gene, so that also its targets will be affected. As an example consider the regulation exerted by VirR on *virT* in Str. 13 (Fig. 1a). This gene is present only in Str. 13 and in Str. ATCC3626, where it is regulated by VirR. Experiments have demonstrated that *virT* encodes a small RNA able to repress the expression of *ccp* and *pfoA* and all these genes are positively controlled by VirR. The loss/gain of *virT* or of VirR binding sites in its promoter will thus have an impact on its own expression, but this will propagate downstream to *ccp* and *pfoA*.

The prediction of VirR targets in the genome of strain JGS1987 revealed the presence of 10 specific putative targets that could be important for the peculiar characteristics of this strain.

On an evolutionary perspective, we noticed that once one gene have been found to be regulated by VirR in one genome, it is either regulated by VirR in other genomes or it is lost. This suggests that many of these genes are useful only when controlled by VirR, and also in this case, that their function is not essential for pathogenesis. Then we can imagine that after loss of the VirR binding site these genes are rapidly deleted from the genome; alternatively the deletion concomitantly involves the gene and its promoter.

Bibliography

[1] L. Addadi, S. Raz, and S. Weiner. Taking advantage of disorder: amorphous calcium carbonate and its roles in biomineralization. *Adv Mat*, 15:959-970, 2003.

[2] S. V. D. Anderson, J. Appanna, T. Huang, and Viswanatha. A novel role for calcite in calcium homeostasis. *FEBS Lett*, 308:94-96, 1992.

[3] A. Angnihotri and L. Hungwen. Enoyl-coa hydratase: reaction and mechanism, and inhibition. *Bioorganic Med Chem*, 11:9–20, 2003.

[4] F. Arigoni, P. A. Kaminski, H. Hennecke, and C. Elmerich. *Mol. Gen. Genet.*, 225, 514–520, 1991.

[5] W. Ba-Thein, M. Lyristis, K. Ohtani, I. T. Nisbet, H. Hayashi, J. I. Rood, and T. Shimizu. The virR/virS locus regulates the transcription of genes encoding extracellular toxin production in Clostridium perfringens. *Journal of bacteriology*, 178(9):2514–20, May 1996.

[6] S. S. Bang, J. K. Galinat, and V. Ramakrishnan. Calcite precipitation induced by polyurethane-immobilized bacillus pasteurii. *Enzyme Microb Technol*, 28(404–409), 2001.

[7] C. Barabesi, A. Galizzi, G. Mastromei, M. Rossi, E. Tamburini, and B. Perito. Bacillus subtilis gene cluster involved in calcium carbonate biomineralization. *Journal of bacteriology*, 189(1):228–35, Jan. 2007.

[8] C. Barabesi, E. Tamburini, G. Mastromei, and B. Perito. *Mechanisms of microbial calcium carbonate precipitation. Art, Biology, and Conserva-*

tion: Biodeterioration ofWorks of Art, volume 472–485. Metropolitan-Museum of Art, New York, 2003.

[9] S. Baskar, R. Baskar, L. Mauclaire, and J. A. McKenzie. *Current science*, 90 (1), 58 – 64, 2006.

[10] D. Bazylinski. Controlled biomineralization of magnetic minerals by magnetotactic bacteria. *Chem Geo*, 132:191-198, 1996.

[11] D. A. Bazylinski and R. B. Frankel. Biologically controlled mineralization in prokaryotes. *Rev Mineral Geochem*, 54:217-247, 2003.

[12] D. A. Bazylinski and B. M. Moskowitz. Microbial biomineralization of magnetic iron minerals: Microbiology, magnetism and environmental significance. *Rev Mineral*, 35:181-223, 1997.

[13] L. A. Bedzyk, K. W. Escudero, R. E. Gill, K. J. Griffin, and F. E. Frerman. *J. Biol. Chem.*, 268, 20211–20217, 1993.

[14] K. Ben Chekroun, C. Rodriguez-Navarro, M. Gonzalez-Muñoz, J. Arias, G. Cultrone, and M. Rodriguez-Gallego. *Journal of Sedimentary Research*, 74 (6), 868-876, 2004.

[15] N. Ben Omar, M. Martínez-Cañamero, M. González-Muñoz, J. Arias, and F. Huertas. *Chemosphere*, 30 (12), 2387-2396, 1995.

[16] M. J. Berridge, M. D. Bootman, and P. Lipp. Calcium-a life and death signal. *Nature*, 395:645-648, 1998.

[17] T. J. Beveridge. Role of cellular design in bacterial metal accumulation and mineralization. *Annu. Rev. Microbiol.*, 43:147–171., 1989.

[18] J. F. Binstock and H. Schulz. Fatty acids oxidation complex from escherichia coli. *Methods Enzymol*, 71(C):403–411, 1981.

[19] E. Boquet, A. Boronat, and A. Ramos-Cormenzana. Production of calcite (calcium carbonate) crystals by soil bacteria is a general phenomenon. *Nature*, 246:527–529, 1973.

[20] M. A. Borowitzka. Morphological and cytological aspects of algal calcification. *Intl Rev Cytology*, 74:127-160, 1982.

[21] M. A. Borowitzka, A. W. D. Larkum, and C. E. Nockolds. A scanning electron microscope study of the structure and organization of the calcium carbonate deposits of algae. *Phycologia*, 13:195-203, 1974.

[22] Z. L. Boynton, G. N. Bennett, and F. B. Rudolph. Cloning, sequencing, and expression of clustered genes encoding -hydroxybutyryl-coenzyme a (coa) dehydrogenase, crotonase, and butyryl-coa dehydrogenase from clostridium acetobutylicum atcc 824. *J. Bacteriol.*, 178:3015–3024., 1996.

[23] M. M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of proteyn-dye binding. *Anal Biochem*, 72:248–254, 1976.

[24] O. Braissant, G. Cailleau, M. Aragno, and E. P. Verrecchia. Biologically induced mineralization in the tree milicia excelsa (moraceae): its causes and consequences to the environment. *Geobiology*, 2: 59–66., 2004.

[25] M. Brilli, A. Mengoni, M. Fondi, M. Bazzicalupo, P. Liò, and R. Fani. Analysis of plasmid genes by phylogenetic profiling and visualization of homology relationships using Blast2Network. *BMC bioinformatics*, 9:551, Jan. 2008.

[26] M. Brilli, A. Mengoni, M. Fondi, M. Bazzicalupo, P. Lió, and R. Fani. Analysis of plasmid genes by phylogenetic profiling and visualization of homology relationships using blast2network. *BMC Bioinformatics*, 9:551–, 2008.

[27] V. Briolat and G. Reysset. Identification of the *Clostridium perfringens* genes involved in the adaptive response to oxidative stress. *J Bacteriol*, 184:2333–2343, 2002.

[28] E. M. Bryan, B. W. Beall, and C. P. J. Moran. A sigma e-dependent operon subject to catabolite repression during sporulation in bacillus subtilis. *J Bacteriol*, 178:4778–4786, 1996.

[29] P. Cacchio, R. Contento, C. Ercole, G. Cappuccio, M. P. Martinez, and A. Ledepi. Involvement of microorganisms in the formation of carbonate speleothems in the cervo cave (l'aquila-italy). *Geomicrobiology*, 21:497– 509., 2004.

[30] P. Cacchio, C. Ercole, G. Cappuccio, and A. Ledepi. Calcium carbonate precipitation by bacterial strains isolated from a limestone cave and from a loamy soil. *Geomicrobiology*, 20:85–98., 2003.

[31] J. W. Campbell and J. E. J. Cronan. Escherichia coli fadr positively regulates transcription of the fabb fatty acid biosynthetic gene. *J Bacteriol*, 183:5982–5990.

[32] J. W. Campbell and J. E. J. Cronan. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu Rev Microbiol*, 55:305–332, 2001b.

[33] S. Castanier, G. Le Metayer-Levrel, G. Orial, J. F. Loubiere, and J. P. Perthuisot. *Bacterial carbonatogenesis and applications to preservation and restoration of historic property*, volume Of microbes and art: the role of microbial communities in the degradation and protection of cultural heritage. Plenum Publishers, New York, NY., 2000.

[34] S. Castanier, A. Maurin, and J. P. Perthuisot. A trial to get dolomite in freshwater. *Geobios*, 23:121–128, 1990.

[35] H. S. Chafetz and C. Buczynski. Bacterially induced lithification of microbial mats. *Palaios*, 7: 277–293, 1992.

[36] D. Chen and R. P. Swenson. *J. Biol. Chem*, 269, 32120–32130, 1994.

[37] J. Cheung, M. Awad, S. McGowan, and J. Rood. Functional analysis of the virsr phosphorelay from clostridium perfringens. *PLoS One*, 2009.

[38] J. K. Cheung, B. Dupuy, D. S. Deveson, and J. Rood. The Spatial Organization of the VirR Boxes Is Critical for VirR-Mediated Expression of the Perfringolysin O Gene , pfoA , from Clostridium perfringens. *Society*, 186(11):3321–3330, 2004.

[39] J. K. Cheung and J. I. Rood. The VirR response regulator from Clostridium perfringens binds independently to two imperfect direct repeats located upstream of the pfoA promoter. *Journal of bacteriology*, 182(1):57–66, Jan. 2000.

[40] R. E. Crick. *Origin, Evolution, and Modern Aspects of Biomineralization in Plants and Animals.*, volume Intl Symposium on Biomineralization 198. Plenum Press, New York, 1989.

[41] S. Date and J. Peregrin-Alvarez. Phylogenetic profiling. *Methods Mol Biol*, 453:201–16, 2008.

[42] A. Decho. *Continent. Shelf. Res.*, 20, 1257-1273, 2000.

[43] C. C. DiRusso. Primary sequence of the escherichia coli fadba operon, encoding the fatty acid-oxidizing multienzyme complex, indicates a high degree of homology to eucaryotic enzymes. *J Bacteriol*, 172(11):6459–6468, 1990.

[44] S. Douglas and T. J. Beveridge. Mineral formation by bacteria in natural microbial communities. *FEMS Microbiol. Ecol.*, 26:79–88., 1998.

[45] E. R. DuPlessis, R. J. Rohlfs, R. Hille, and C. Thorpe. *Biochem. Mol. Biol. Int.*, 32, 195–199, 1994.

[46] S. Dupraz, M. Parmentiera, B. Meneza, and F. Guyota. Experimental and numerical modeling of bacterially induced ph increase and calcite precipitation in saline aquifers. *Chem Geol*, 265: 44–53., 2009b.

[47] B. E. *Biomineralization.* Wiley-VCH Verlag GmbH, Weinheim Germany, 2000.

[48] H. L. Ehrlich. *Geomicrobiology*. 1996.

[49] H. L. Ehrlich. Microbes as geologic agents: their role in mineral formation. *Geomicrobiol. J*, 16:135–153, 1999.

[50] G. Finocchairo, I. Colombo, B. Garavaglia, C. Gellera, G. Valdameri, N. Garbuglio, and S. DiDonato. *Eur. J. Biochem.*, 213, 1003–1008, 1993.

[51] J. C. Fong and H. Schulz. Purification and properties of pig heart crotonase and the presence of short chain enoyl-coa hydratases in pig and in guinea pig tissues. *J Biol Chem*, 252(2):542–547, 1977.

[52] D. Fortin, F. G. Ferris, and T. J. Beveridge. Surface-mediated mineral development by bacteria. *Rev. Mineral. Geochem.*, 35:161–180, 1997.

[53] Y. Fujita, H. Matsuoka, and K. Hirooka. Regulation of fatty acid metabolism in bacteria. *Mol Microbiol*, 66(4):829–839, 2007.

[54] U. K. Gollapudi, C. L. Knutson, S. S. Bang, and M. R. Islam. A new method for controlling leaching through permeable channels. *Chemosphere*, 30: 695–705, 1995.

[55] J. E. González-Pastor, E. C. Hobbs, and R. Losick. Cannibalism by sporulating bacteria. *Science*, 301:510–513, 2003.

[56] M. Gubler, T. Zurcher, and H. Hennecke. *Mol. Microbiol.*, 3, 141–148, 1989.

[57] M. G. N. Hartmanis and S. Gatenbeck. Intermediary metabolism in clostridium acetobutylicum: levels of enzymes involved in the formation of acetate and butyrate. *Appl Environ Microbiol*, 47 (6):1277–1283, 1984.

[58] M. Husain and D. J. Steenkamp. *J. Bacteriol.*, 163, 709–715, 1985.

[59] M. Inui, M. Suda, S. Kimura, K. Yasuda, H. Suzuki, H. Toda, S. Yamamoto, S. Okino, N. Suzuki, and H. Yukawa. Expression of clostridium acetobutylicum butanol synthetic genes in escherichia coli. *Appl Microbiol Biotechnol*, 77(6):1305–1316, 2008.

[60] L. L. Jiang, A. Kobayashi, H. Matsuura, H. Fukushima, and T. Hashimoto. Purification and properties of human d-3-hydroxyacyl- coa dehydratase: medium chain enoyl-coa hydratase is d-3- hydroxyacyl-coa dehydratase. *J Biochem*, 120(3):624–632, 1996.

[61] P. A. Kaminski, F. Norel, N. Desnoues, A. Kush, G. Salzano, and C. Elmerich. *Mol. Gen. Genet.*, 214, 496–502, 1988.

[62] J. L. Kirschvink and J. W. Hagadorn. *A grand unified theory of biomineralization.*, volume Biomineralization, p 139-149. Wiley-V. C. H. Verlag GmbH, Weinheim, Germany, 2000.

[63] A. Knoll. Biomineralization and evolutionary history. *Rev Mineral Geochem*, 54:329-356, 2003.

[64] K. O. Konhauser. Bacterial iron biomineralisation in nature. *FEMS Microbiol Rev*, 20:315-326, 1997.

[65] U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage t4. *Nature*, 227:680–685, 1970.

[66] B. S. C. Leadbeater and R. Riding. *Biomineralization in Lower Plants and Animals.* The Systematics Association Special Volume. Clarendon Press, Oxford, 1986.

[67] V. Lee and O. Schneewind. Protein secretion and the pathogenesis of bacterial infections. *Genes Dev*, 15:1725–1752, 2001.

[68] H. A. Lowenstam. Minerals formed by organisms. *Science*, 211:1126-1131, 1981.

[69] H. A. Lowenstam and L. Margulis. *Calcium regulation and the appearance of calcareous skeletons in the fossil record.*, volume The Mechanisms of Biomineralization in Animals and Plants, p 289-300. Tokai University Press, Tokyo, 1980.

[70] H. A. Lowenstam and S. Weiner. *On Biomineralization*. Oxford University Press, New York, 1989.

[71] S. Mann. Mineralization in biological systems. *Struct Bonding*, 54:125-174, 1983.

[72] S. Mann. *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry.* Oxford University Press, New York, 2001.

[73] J. Marmur. A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J. Mol. Biol.*, 3:208–217, 1961.

[74] M. Marvasi, P. T. Visscher, B. Perito, G. Mastromei, and L. Casillas-Martínez. Physiological requirements for carbonate precipitation during biofilm development of Bacillus subtilis etfA mutant. *FEMS microbiology ecology*, 71(3):341–50, Mar. 2010.

[75] H. Matsuoka, K. Hirooka, and Y. Fujita. Organization and function of the ysia regulon in bacillus subtilis involved in fatty acids degradation. *J Biol Chem*, 282(8):5180–5194, 2007.

[76] T. McConnaughey. 13c and 18o isotopic disequilibrium in biological carbonates: I. patterns. *Geochim Cosmochim Acta*, 53:151-162, 1989.

[77] K. Miyamoto, D. Fisher, J. Li, S. ayeed, S. Akimoto, and B. McClane. Complete sequencing and diversity analysis of the enterotoxin-encoding plasmids in *Clostridium perfringens* type a non-food-borne human gastrointestinal disease isolates. *J Bacteriol*, 188:1585–98, 2006.

[78] K. Miyamoto, J. Li, S. Sayeed, S. Akimoto, and BAMcClane. Sequencing and diversity analyses reveal extensive similarities between some epsilon-toxin-encoding plasmids and the pcpf5603 *Clostridium perfringens* enterotoxin plasmid. *J Bacteriol*, 190:7178–88, 2008.

[79] R. Mollby, , and T. Holme. Production of phospholipase c (α-toxin), haemolysins and lethal toxins by *Clostridium perfringens* types a to d. *J. Gen. Microbiol*, 96:137–144, 1976.

[80] G. Myers, D. Rasko, J. Cheung, J. Ravel, R. Seshadri, R. DeBoy, Q. Ren, J. Varga, M. Awad, L. Brinkac, S. Daugherty, D. Haft, D. odson, R. Madupu, W. Nelson, M. Rosovitz, S. Sullivan, H. Khouri, G. Dimitrov, K. Watkins, S. Mulligan, J. Benton, D. Radune, D. Fisher, H. Atkins, T. Hiscox, B. Jost, S. Billington, J. Songer, B. McClane, R. Titball, J. Rood, S. Melville, and I. Paulsen. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. *Genome Res*, 16:1031–40, 2002.

[81] K. Okumura, K. Ohtani, H. Hayashi, and T. Shimizu. Characterization of genes regulated directly by the VirR/VirS system in Clostridium perfringens. *Journal of bacteriology*, 190(23):7719–27, Dec. 2008.

[82] K. Okumura, K. Ohtani, H. Hayashi, and TShimizu. Characterization of genes regulated directly by the virr/virs system in *Clostridium perfringens*. *J Bacteriol.*, 190:7719–27, 2008.

[83] H. O'Neill and G. Mayhew, S. G. Butler. Cloning and analysis of the genes for a novel electron-transferring flavoprotein from megasphaera elsdenii. expression and characterization of the recombinant protein. *J. Biol. Chem.*, 273:21015–21024, 1998.

[84] B. C. Orsbur, S. B. Melville, and D. L. Popham. Etfa catalyses the formation of dipicolinic acid in clostridium perfringens. *Mol. Microbiol.*, 75:178–186, 2010.

[85] H. Paerl, T. Stepp, and R. Reid. *Environ. Microbiol.*, 3 (2), 123-130, 2001.

[86] M. Pellegrini, E. Marcotte, M. Thompson, D. Eisenberg, and T. Yeates. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*, 96:4285–8, 1999.

[87] A. Pentecost. *Geomicrobiology Journal*, 4, 285–298, 1985.

[88] M. Perego. *Integrational vectors for genetic manipolation in Bacillus subtilis*, volume p. 615–622. American Society for Microbiology, Washington, DC, 1993.

[89] B. Perito, L. Biagiotti, S. Daly, A. Galizzi, P. Tiano, and G. Mastromei. *Of microbes and art: the role of microbial communities in the degradation and protection of cultural heritage-Bacterial genes involved in calcite crystal precipitation*. Plenum Publishers, New York, NY, 2000.

[90] M. G. Reese. *Computational prediction of gene structure and regulation in the genome of Drosophila melanogaster*. PhD thesis, UC Berkeley/University of Hohenheim, 2000.

[91] M. G. Reese. Application of a time-delay neural network to promoter annotation in the drosophila melanogaster genome. *Comput Chem*, 26(1),51-6, 2001.

[92] M. Rivadeneyra, J. Prraga, R. Delgado, A. Ramos-Cormenzana, and G. Delgado. *FEMS Microbiol. Ecol.*, 48, 39- 46, 2004.

[93] M. A. Rivadeneyra, G. Delgado, M. Soriano, A. Ramos-Cormenzana, and R. Delgado. Biomineralization of carbonates by marinococcus albus and marinococcus halophilus isolated from the salar de atacama (chile). *Curr Microbiol*, 39: 53–57, 1999.

[94] D. L. Roberts, F. E. Frerman, and J. J. Kim. *Proc. Natl. Acad. Sci. U. S. A.*, 93, 14355–14360, 1996.

[95] C. Rodriguez-Navarro, M. Rodriguez-Gallego, K. Ben Chekroun, and M. Gonzalez-Munoz. Conservation of ornamental stone bymyxococcus xanthus-induced carbonate biomineralization. *Appl Environ Microb*, 69: 2182–2193, 2003.

[96] T. Ryuichi, S. Kenichiro, and H. Akihito. Fluorescence detection of calcium-binding proteins with quinoline ca-indicator quin2. *ANALYTICAL BIOCHEMISTRY*, 254, 126–131, 1997.

[97] K. Sato, Y. Nishina, and K. Shiga. *J. Biochem.*, 114, 215–222, 1993.

[98] S. Sato, T. Christopher, C. T. Nomura, H. Hideki Abe, Y. Doi, and T. Tsuge. Poly[(r)-3-hydroxybutyrate] formation in escherichia coli from glucose through an enoyl-coa hydratase-mediated pathway. *J Biosci Bioeng*, 103(1):38–44, 2007.

[99] Y. Sawires and J. Songer. *Clostridium perfringens*: insight into virulence evolution and population structure. *Anaerobe*, 12:23–43, 2006.

[100] D. Schüler and R. Frankel. Bacterial magnetosomes: Microbiology, biomineralization and biotechnological applications. *Applied Microbiol Biotech*, 52:464-473, 1999.

[101] T. Shimizu, K. Shima, K.-i. Yoshino, K. Yonezawa, T. Shimizu, and H. Hayashi. Proteome and Transcriptome Analysis of the Virulence

Genes Regulated by the VirR / VirS System in Clostridium perfringens. *Society*, 184(10):2587–2594, 2002.

[102] K. Simkiss and K. Wilbur. *Biomineralization. Cell Biology and Mineral Deposition.* Academic Press, Inc., San Diego, 1989.

[103] R. J. Smith. Calcium and bacteria. *Adv. Microb. Physiol.*, 37:83–133, 1995.

[104] D. J. Steenkamp and M. Gallup. *J. Biol. Chem.*, 253, 4086–4089, 1978.

[105] S. Stocks-Fischer, J. K. Galinat, and S. S. Bang. Microbiological precipitation of caco3. *Soil Biol Biochem*, 31: 1563–1571, 1999.

[106] H. Tamao, T. Tsuge, T. Fukui, T. Iwata, K. Miki, and Y. Doi. Crystal structure of the (r)-specific enoyl-coa hydratase from aeromonas caviae involved in polyhydroxyalkanoate biosynthesis. *J Biol Chem*, 278(1):617–624, 2003.

[107] B. M. Tebo, W. C. Ghiorse, L. G. van Waasbergen, P. L. Siering, and R. Caspi. Bacterially mediated mineral formation: Insights into manganese(ii) oxidation from molecular genetic and biochemical studies. *Rev Mineral*, 35:225-266, 1997.

[108] C. Thorpe. *Chemistry and Biochemistry of Flavoenzymes*, volume Mu 2, pp. 471–486. CRC Press, Inc., Boca Raton, FL, 1991.

[109] P. Van Cappellen. Biomineralization and global biogeochemical cycles. *Rev Mineral Geochem*, 54:357-381, 2003.

[110] Visone. Visone: analysis and visualization of social networks [http://visone.info/].

[111] H. Von Knorre and W. Krumbein. *Microbial Sediments*, volume Microbial Sediments, pp. 25– 31. Springer, Berlin, 2000.

[112] L. A. Warren, P. A. Maurice, N. Parmar, and F. G. Ferris. Microbially mediated calcium carbonate precipitation: implications for interpreting calcite precipitation and for solid-phase capture of inorganic contaminants. *Geomicrobiol J*, 18: 93–115, 2001.

[113] N. J. Watmough, J. Kiss, and F. E. Frerman. *Eur. J. Biochem.*, 205, 1089–1097, 1992.

[114] P. Westbroek. *Biological metal accumulation and biomineralization in a geological perspective.*, volume Biomineralization and Biological Metal Accumulation. Westbroek P, de Jong EW. Publishing Co, Dordrecht, Holland, p 1-11, 1983.

[115] D. Whitworth and P. Cock. Evolution of prokaryotic two-component systems: insights from comparative genomics. *Amino Acids*, 37:459–466, 2009.

[116] A. N. Wipat, S. C. Carter, B. J. Brignell, K. Guy, J. Piper, P. T. Sanders, C. R. Emmerson, and C. R. Harwood. The dnab-phea (256 degrees-240 degrees) region of the bacillus subtilis chromosome containing genes respon- sible for stress responses, the utilization of plant cell wallls and primary metabolism. *Microbiology*, 142:3067–3078, 1996.

[117] D. T. Wright and A. Oren. Nonphotosynthetic bacteria and the formation of carbonates and evaporites through time. *Geomicrobiol J*, 22: 27–53., 2005.

[118] G. A. Zavarzin. Microbial geochemical calcium cycle. *Microbiology*, 71:1–17, 2002.

Declaration

Results shown in this thesis come from works done in the Lab. of Microbiology headed by Prof. Giorgio Mastromei and at the Dept. of Biomedical technology and Sciences (DiSTeB) of the university of Cagliari, in the Lab of Prof. Enrico Sanjust.

Florence, December 2010

ANTONIO FRANDI