Reconstruction, Modeling & Analysis of Haloarchaeal Metabolic Networks

Orland Gonzalez

München, 2009

Reconstruction, Modeling & Analysis of Haloarchaeal Metabolic Networks

Orland Gonzalez

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig-Maximilians-Universität

München

vorgelegt von

Orland Gonzalez

aus Manila

München, den 02.03.2009

Erstgutachter: Prof. Dr. Ralf Zimmer

Zweitgutachter: Prof. Dr. Dieter Oesterhelt

Tag der mündlichen Prüfung: 21.01.2009

Contents

List of Figures

List of Tables

Summary

The metabolic network of an organism refers to the set of biochemical transformations (reactions) that occur within the organism which allows it to maintain life. The network defines what nutrients the organism can use, for example to harvest energy through cellular respiration, what it cannot use, and what it needs for growth. This thesis deals with the metabolic reconstruction of two haloarchaeal species, and on the development of computational models from them.

Extremely halophilic archaea inhabit hypersaline environments, such as salt lakes, salt ponds and marine salterns, which relatively few organisms have been able to adapt to. For the primary subject of this work, *Halobacterium salinarum*, we present a genome-scale metabolic reconstruction that consists of 695 reactions, 485 genes and 548 metabolites. Although semi-automated bioinformatic methods were used during reconstruction, our effort involved a great deal of manual curation so the resulting network is of high quality. We demonstrate that even prior to any formal modeling effort, the network is already valuable because it represents a literature-linked summary of the existing metabolic knowledge. Indeed, the reconstruction led to improvements to the existing genome annotation, and facilitated the formulation of biological predictions that led to further experimental work; for example with respect to the production of pentoses and the biosynthesis of aromatic amino acids.

Using the reconstructed network for *Halobacterium salinarum* as basis, we created a constraints-based model to perform an integrated analysis of energy generation, nutrient utilization and biomass production, during both aerobic and phototrophic growth. We made several unexpected, biologically relevant findings, including: (1) that the organism degrades essential amino acids during respiration, and even under phototrophic conditions, when the cells should already have abundant energy. This is unexpected, particularly in the latter situation, because such behavior can lead to cells terminating their growth earlier than necessary; (2) that carbon incorporation rates were surprisingly low also under

both aerobic and phototrophic conditions. Note that during phototophy, one would typically expect nutrient consumption to be just in quantities that are sufficient for biomass production since energy can already be derived from light rather than from the supplied nutrients. Because of the low incorporation rates, considerable amounts of metabolites were produced and accumulated in the medium, some of which likely resulted from forms of overflow metabolism; and (3) that arginine fermentation actually occurs simultaneously with respiration and photosynthesis, even though cells should already have abundant energy. This is in contrast to the common assumption that the fermentation process only plays a secondary, alternative role. Moreover, our results also show that arginine fermentation can actually produce usable energy in levels that are comparable to the two "primary" modes.

The initially suprising findings mentioned above are consistent with an emerging picture that *Halobacterium salinarum* takes an approach toward growth that favors the here and now, even at the cost of longer-term concerns. We believe that the seemingly greedy behavior actually consists of adaptations by the organism to its natural environments, where nutrients are not only irregularly available but may altogether be absent for extended periods that may span several years. Such a setting probably predisposed the cells to grow as much as possible when the conditions become favorable, so that the probability of survival until the next growth-conducive period is maximized. Consistent with this, the capacity of the halobacteria to survive in adverse conditions for extended periods is well established.

From a more methodological point-of-view, we developed a hybrid, genome-scale flux balance model with both dynamic and constraints-based aspects. The dynamic parts of the model are used to incoporate growth data, such as nutrient consumption and growth rates, to the constraints-based model core in the form of additional constraints. The constraints are derived through linearization of the dynamic parts of the model at specific points during growth, and the linearized values at each time point are considered to be valid for the interval that contains the point. Although a number of similar hybrid models have already been constructed in the past, the systems used for the dynamic components of these models are too simple; in particular they cannot account for transport patterns that change qualitatively during growth. This is a particularly restrictive limitation for this study because, in the experimental growth data that we generated, some amino acids clearly show qualitative changes in their transport patterns during growth. For example, alanine at the start of growth accumulates in the medium, but later on switches to consumption.

To overcome this limitation, we propose and use a new system for dealing with the dynamic aspects of such hybrid models. Briefly, the system that we employ uses two equation forms; one basic definition and another more descriptive one. For each metabolite (nutrient), each of these forms is individually applied using a process that involves solving inverse problems (parameter optimization). From the results of these individual applications, a final choice is made between the two forms based on clear measures of improvement of the second form over the basic definition, or the lack thereof. The process can be easily automated, and as such not only can our proposed system account for qualitatively changing transport patterns, but it can actually detect their presence. Moreover, since qualitative changes in nutrient transport are often the result of, or at least related to, another event, our method can actually uncover relationships between metabolites in the form of coordinated pattern changes. For example, we were able to automatically detect known relationships, such as between arginine and ornithine, and we were also able to uncover possible interaction dynamics, such as competitive inhibition, between other metabolites.

This work is structured into five parts. Chapter 2 is an introduction to the halophilic archaea and their environments. Chapter 3 deals with the reconstruction of the metabolic network of *Halobacterium salinarum* and its uses. Specific aspects of the archaeon's metabolism are also discussed. In Chapter 4, we describe how we derived a constraints-based model from the reconstructed network. In addition, this chapter also contains the results of our analysis of the aerobic growth of *Halobacterium salinarum*. Chapter 5 is similar to the previous chapter, except that the analysis is done for cells grown under phototrophic conditions. Finally, Chapter 6 describes the results of both the metabolic reconstruction and modeling of *Natronomonas pharaonis*.

Zusammenfassung

Das metabolische Netzwerk eines Organismus besteht aus einem Satz biochemischer Transformationen (Reaktionen), welche in dem Organismus vorkommen and ihn am Leben erhalten. Das Netzwerk legt fest, welche Nährstoffe der Organismus nutzen kann, um beispielsweise Energie durch zelluläre Atmung zu gewinnen, was der Organismus nicht nutzen kann und was er zum Wachstum benötigt. Diese Untersuchung befasst sich mit der metabolischen Rekonstuktion zweier haloarchaealer Organismen, und mit der Entwicklung rechnergestützter Modelle für diese.

Extrem halophile Archaeen leben in höchst salzhaltiger Umgebung wie Salzseen, Salztümpeln und marinen Salinen, an die sich relativ wenige andere Organismen anpassen konnten. Für das Hauptthema dieser Arbeit, *Halobacterium salinarum*, beschreiben wir eine genomweite metabolische Rekonstruktion welche aus 696 Reaktionen, 478 Genen und 550 Metaboliten besteht. Obwohl halbautomatische bioinformatische Methoden während der Rekonstruktion genutzt wurden, benötigte unser Vorgehen eine erhebliche Menge manueller Nacharbeit, um ein Netzwerk hoher Qualität zu erstellen. Wir zeigen, dass das Netzwerk bereits vor jeglichem formalen Modellierungsversuch wertvoll ist, weil es eine mit der Literatur verzahnte Zusammenfassung des vorhandenen metabolischen Wissens darstellt. So führte die Rekonstruktion zu Verbesserungen der existierenden Genom-Annotation und erleichterte die Formulierung biologischer Vorhersagen, welche weitere experimentelle Arbeiten nach sich zogen; als Beispiele seien die Herstellung von Pentosen und die Biosynthese aromatischer Aminosäuren genannt.

Unter Benutzung des rekonstruiertenen metabolischen Netzwerks für *Halobacterium salinarum* als Basis, haben wir ein "constraint-basiertes" Modell (Randbedingungs-basiertes Modell) erstellt, um eine integrierte Analyse des Energiemetabolismus, des Nährstoffverbrauchs und der Produktion von Biomasse, sowohl unter aeroben als auch unter phototrophen Wachstumsbegingungen zu erarbeiten. Dabei machten wir mehrere unerwartete biologisch relevante Beobachtungen: (1) der Organismus baut essentielle Aminosäuren ab und zwar

sowohl zu Zeiten, wo er atmet, also auch unter phototrophen Bedingungen, wenn die Zellen eigentlich bereits genug Energie zur Verfügung haben sollten. Das is vor allen bei letzteren unerwartet, denn dieses Verhalten kann dazu führen, dass die Zellen ihr Wachtum früher als notwendig beenden müssen; (2) die Einbauraten für Kohlenstoff waren erstaunlich niegdrig, ebenfalls sowohl unter aeroben als auch unter phototrophen Bedingungen. Es sollte bedacht werden, dass man erwarten würde, dass der Nährstoffverbrauch während der Phototrophie gerade für die Produktion von Biomasse ausreicht, da Energie ja bereits aus Licht gewonnen werden kann und somit zur Verfügung gestellten Nährstoffe dafür nicht eingesetzt werden müssten. Auf Grund der geringen Einbauraten müssen erhebliche Mengen von Metaboliten erzeugt und im Medium angereichert werden, wobei einige davon wahrscheinlich aus Formen des Überlauf-Metabolismus stammen; (3) die Fermentation von Arginin findet gleichzeitig mit Atmung und Photosynthese statt, obwohl die Zellen bereits ausreichend mit Energie versogt sein sollten. Das widerspricht der allgemein anerkannten Meinung, dass Fermentationsprozesse nur eine untergeordnete, alternative Rolle spielen. Zudem zeigen unsere Ergebnisse, dass die Mengen der durch Fermentation von Arginin erhaltenen nutzbaren Energie vergleichbar zu denen sind, die aus den anderen "primären" Modi gewonnenen werden.

Die hier beschriebenen und anfangs unerwarteten Beobachtungen passen zu einem sich abzeichnenden Gesamtbild, nämlich dass *Halobacterium salinarum* bezüglich des Wachstums auf das "hier und jetzt" setzt, selbst auf Kosten längerfristiger Belange. Wir glauben, dass das scheinbar gefrässige Verhalten in Wirklichkeit eine Anpassung des Organismus an seine natürliche Umgebung darstellt, wo Nährstoffe nicht nur unregelmässig verfügbar sind, sondern sogar teilweise für erhebliche Zeiträume, die mehrere Jahre umfassen können, vollständig fehlen. Eine solche Szenerie prädisponiert die Zellen so stark wie möglich zu wachsen, wenn die Bedingungen vorteilhaft sind, so dass die Wahrscheinlichkeit maximiert wird, bis zur nächsten wachstumsfördernden Phase zu überleben. Im Einklang damit hat sich gezeigt, dass Halobacterien fähig sind unter widrigen Umständen über lange Zeiträume zu überleben.

Von einem mehr methodischen Standpunkt haben wir ein hybrides genomweites flux-balance Modell (Fluss-Balance Modell) sowohl mit dynamischen als auch mit constraint-basierten Aspekten entwickelt. Die dynamischen Teile des Modells wurden benutzt, um Ergebnisse von Wachstumsversuchen wie Nährstoffverbrauch und Wachstumsraten in den Kern des constraint-basierten Modells einzufügen und zwar in Form zusätzlicher Randbedingungen. Die Randbedingungen wurden durch Linearisierung der dynamischen Teile

des Modells zu bestimmten Zeitpunkten während des Wachstums gewonnen und die linearisierten Werte an jedem Zeitpunkt für das diesen Punkt enthaltende Intervall als gültig angenommen. Obwohl eine Reihe ähnlicher hybrider Modelle in der Vergangenheit konstruiert worden sind, sind die Systeme, welche für die dynamischen Komponenten der Modelle benutzt wurden, zu sehr vereinfacht, insbesondere können sie keine Transportmuster berücksichtigen, die sich während des Wachstums qualitativ ändern. Das ist eine besonders starke Einschränkung für diese Untersuchung, weil einige Aminosäuren eindeutig qualitative Änderungen des Transportmusters während des Wachstums in unseren experimentellen Messungen gezeigt haben. Beispielsweise reichert sich Alanin am Beginn des Wachstums im Medium an, während später eine Umstellung auf Alanin-Verbrauch stattfindet.

Um diese Beschränkung zu beseitigen benutzen wir ein von uns vorgeschlagenes neues System, um mit den dynamischen Aspekten solcher hybrider Modelle umzugehen. Kurz gesagt nutzt das von uns vorgeschlagene System zwei Formen einer Gleichung; eine elementare Definition und eine weitere, welche anschaulicher ist. Für jeden einzelnen Metaboliten (Nährstoff) ist jede dieses Former individuell angewendet worden unter Nutzung eines Prozesses zur Lösung des mathematischen "inverse problem" (Optimierung der Parameter). Auf Grund der Ergebnisse wird für den jeweiligen Metaboliten eine endgültige Entscheidung zwischen den beiden Formen getroffen, welche darauf beruht, dass die zweite gegenüber der elementaren Form zu einer eindeutigen Verbesserungen führt oder eine solche Verbesserung ausbleibt. Dieser Prozess lässt sich leicht automatisieren und so kann unser vorgeschlagenes System nicht nur sich qualitativ ändernde Transportmuster berücksichtigen, sondern deren Vorhandensein sogar aufspüren. Zudem kann unsere Methode sogar Beziehungen zwischen Metaboliten auf Grund koordinierter Änderungen des Transportmusters aufdecken, denn qualitative Änderungen im Nährstofftransport sind oft das Ergebnis von - oder zumindest bekoppelt an - andere Ereignisse. Beispielsweise haben wir automatisch bereits bekannte Verknüpfungen erhalten, wie die zwischen Arginin und Ornithin und wir konnten bei anderen Metaboliten eine mögliche Interkations-Dynamik erkennen, wie beispielsweise kompetitive Hemmung.

Diese Arbeit ist in fünf Teile geliedert. Kapitel 2 ist eine Einführung zu halophilen Archaeen und ihrer Umwelt. Kapitel 3 befasst sich mit der Rekonstruktion des metabolischen Netzwerks von *Halobacterium salinarum*, sowie dessen Anwendungen. Spezifische Aspekte aus dem Metabolismus dieses Archaeons werden diskutiert. In Kapiel 4 beschreiben wir wie wir das contraint-basierte Modell aus dem rekonstruierten metabolischen Netzwerk

entwickelt haben. Weiterhin enthält dieses Kapiel die Ergebnisse unserer Analyse des aeroben Wachstums von *Halobacterium salinarum*. Kapitel 5 ist ähnlich dem vorangehenden Kapitel, allerdings wurden die Analysen für Zellen gemacht, welche unter phototrophen Bedingungen gewachsen sind. Zuletzt beschreibt Kapitel 6 die Ergebnisse sowohl der metabolischen Rekonstruktion als auch der Modellierung für *Natronomonas pharaonis*.

Chapter 1

Introduction

Metabolism refers to the set of biochemical processes (reactions) that occur within living organisms that allow them to grow and reproduce, to maintain their structures, and to respond to their environments. For example, the catabolic pathways of metabolism are involved with the breakdown of large molecules, and as such define what materials a cell can use for growth and energy. On the other hand, anabolic pathways are concerned with the construction of cellular building blocks from the nutrients provided, and accordingly define what can be synthesized and what needs to be provided externally in order for growth to be possible. The collection of all metabolic pathways within a cell is referred to as its metabolic network. This thesis deals with the reconstruction of the metabolic networks of two extremely halophilic archaeal species, and on the development and use of computational models derived from them. Our aim was for an integrative analysis of bioenergetics, nutrient consumption and growth, under the particularly harsh (hypersaline) environments where the haloarchaea thrive but very few other organisms on Earth have been able to adapt to. To this end, both computational (e.g., modeling) and experimental (e.g., generation of time-series growth data) work were performed. An introduction to the biological subject matter is provided in Chapter 2. An overview of the overall thesis structure is shown in Figure 1.1.

Although most of the tasks (transitions) shown in Figure 1.1 were performed for each of the two haloarchaeal species analyzed in this study, in this introduction the case for *Halobacterium salinarum* is used for illustration. This thesis summarizes interdisciplinary work by the author as it describes both computational modeling and experimental work. Most of the experimental results used for the modeling and simulations were produced by the author in Prof. Oesterhelt's group at the Max Planck Institute of Biochemistry (MPI).
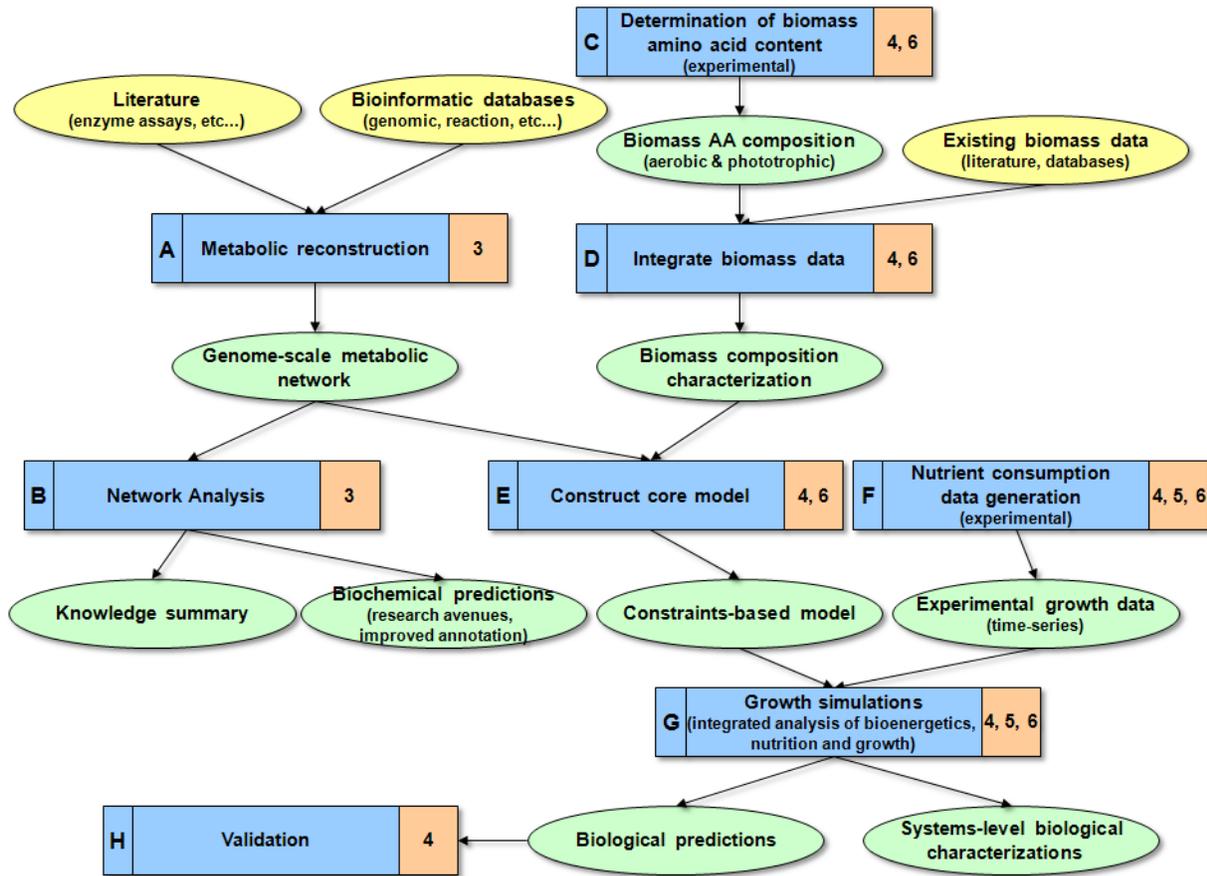
Figure 1.1: Petri net representation of the overall structure of this thesis. Yellow places (ellipses) in the graph represent external entities that were used, while green places represent entities that were produced in this work. The transitions (blue rectangles) correspond to research tasks, which were performed in the thesis work. Adjacent to each transition is an orange box specifying in which chapter(s) the action is described. Single letter codes are associated to each transition so that they can be easily referred to from the main text.

Metabolic reconstruction

Metabolic reconstruction (Figure 1.1, transition A), roughly speaking, refers to the task of assembling the reactions that are known (or believed) to occur within an organism into a network. Clearly, one way of accomplishing this task is through a search of the literature for reactions that have already been experimentally described for *Halobacterium salinarum*. However, a network constructed exclusively from experimental evidence will be incomplete. The problem is that while metabolic pathways are typically conserved, at least in part, across groups of species, direct experimental evidence is often only available for a small subset of model organisms. The difficulty is particularly severe when working

with organisms that are not as popular as, say, *E. coli* and yeast, because typically only a small fraction of metabolic reactions would have been experimentally characterized. This leads to the second main resource for metabolic reconstruction, bioinformatics.

Most biochemical reactions require enzymes under physiological conditions; enzymes are crucial because they allow organisms to drive desirable but otherwise thermodynamically unfavorable reactions by coupling them to favorable ones. This fact is very important for metabolic reconstruction because enzymes are proteins, and proteins are coded for by genes. Therefore, the question of whether a particular reaction occurs within an organism can be indirectly answered by asking whether the required enzyme-coding gene is present. In turn, this latter question can be addressed by matching genes from other organisms that are known to code for the necessary enzyme to the gene complement of the subject organism; this is a bioinformatic problem. For obvious reasons, such a problem transformation is very helpful for organisms with completely sequenced genomes, such as *Halobacterium salinarum*. Full details on the reconstruction procedure used in this study can be found in Chapter 3.

The genome-scale metabolic network for *Halobacterium salinarum* is one of the most significant outputs of this study. Using an existing reconstruction as basis (Falb, 2005), in this thesis, the network was doubled in size to cover nearly 700 reactions. Reconstruction of the network was very laborious because extensive manual curation is necessary in order to ensure high reliability. For example, over 80 unique literature references are associated with the network as experimental evidence, covering about 30% of the reactions (a detailed breakdown of the reactions by evidence is given in Chapter 3). Each of these references had to be reviewed individually. Moreover, given that most of these references are experimental in nature, a nontrivial level of competence with respect to the relevant biochemical subjects had to be acquired first. The necessity for extensive manual curation is further discussed in Chapter 3. The reconstruction for the second haloarchaeal species, *Natronomnas pharaonis*, is described in Chapter 6.

One of the most important uses of reconstructed metabolic networks is to serve as foundation for the development of computational models. However, even prior to any modeling effort, the network is already valuable in that it represents a summary of the current knowledge regarding the metabolism of the organism. For example, by analyzing the network (Figure 1.1, transition B), one could identify interesting regions – such as those associated with missing functions in the network that, based on physiological data, should be present – and subsequently form biological predictions. For *Halobacterium salinarum*,

among the "gaps" that we found are missing initial steps in the sythesis of aromatic amino acids, and the lack of a synthetic pathway for riboses. The first is noteworthy because while aromatic amino acids are necessary components of the biomass, they are known not to be essential for the organism, and as such must be produced *de novo*. Likewise, riboses are precursors of nucleotides (DNA and RNA), and accordingly also essential components of the biomass. Separate research efforts on these two subjects were initiated, the work done by colleagues at the MPI, in part due to the observations made in this study. In addition, the reconstruction effort also lead to improvements/updates to the existing primary annotation of the halophile's genome.

Constraints-based models

The development of dynamic (kinetic) models requires knowledge of the intricate regulatory features of the system in question. Unfortunately, such information is normally difficult to obtain. Indeed, it is often the case that enzyme kinetic parameters are available only for a limited number of model organisms. This, not even considering the fact that K_M values are typically derived from in vitro assays, a condition that casts suspicion on their validity in vivo (Teusink *et al.*, 2000; Theobald *et al.*, 1997; Rizzi *et al.*, 1997). Accordingly, kinetic models of metabolism have mostly been limited to relatively small networks. In part due to the difficulties mentioned, constraints-based models have emerged as successful alternatives. Rather than requiring detailed, difficult to obtain information, constraints-based models need only available physicochemical information such as stoichiometry, reversibility and energy balance (Edwards *et al.*, 2000; Edwards and Palsson, 1999; Ramakrishna *et al.*, 2001). This distinction is particularly relevant because genome-scale metabolic reconstructions, which are performed here, already include a lot of this information. Indeed, each of the two metabolic networks reconstructed in this study is used for the development of the respective constraints-based core model (Figure 1.1, transition E).

In addition to the reconstructed network, the other main component for the development of a constraints-based core model (Figure 1.1, transition E) is a characterization of the biomass of the respective organism. The characterization is used to define a "growth reaction" (Figure 1.1, transition D), which is essentially a pseudoreaction that consumes biomass components (reactants) in the proper ratio and produces a unit of population (such as cells). Note that adding this to the metabolic network and specifying a particular flux to it, for example one derived from observed growth rates, imposes additional constraints on the network; e.g., it defines pathways that need to be activated for the production of the

biomass components. The protein fraction of the biomass was experimentally determined in this study (Figure 1.1, transition C). This involved measurements of individual amino acids at different optical densities (population levels). Values for the rest of the biomass components were derived, and in some instances directly obtained, from existing data. Details on the creation of the growth function are provided in Chapter 4.

Growth simulations using hybrid models

Constraints-based models can be used to perform growth simulations (Figure 1.1, transition G) by integrating appropriate data, such as those that describe nutrient consumption and growth rates. These data are typically enforced as additional constraints. For example, an observed glucose consumption rate of r can be integrated into the model by setting the glucose transport flux to r. With respect to *Halobacterium salinarum*, growth media used for the halophile typically include several amino acids. Accordingly, in this study, we used a growth medium that contains 15 amino acids. Time-series data showing the consumption and production of these amino acids during growth were produced in this thesis (Figure 1.1, transition F). Data sets were generated for both aerobic (Chapter 4) and phototrophic (Chapter 5) conditions, in order to allow for a comparison between the two main bioenergetic modes of *Halobacterium salinarum*. Oxygen consumption was also monitored under aerobic conditions. Growth simulations were performed on the basis of these data.

One limitation of basic constraints-based models is that they are essentially static, i.e., the contraints do not change. While this is fine for continuous cultures where nutrient consumption and growth rates are maintained, it is not applicable to batch cultures where the rates are not constant. One way in which this issue has been addressed is through the use of hybrid models. The idea is to use dynamic models for the system inputs (e.g., nutrients) and outputs (e.g., by-products), and then linearize the system at various points during growth. At each of these specific points, the transport equations are evaluated, and the values are assumed to be valid for an interval containing the point. This approach has been applied to a number of biological systems (Mahadevan *et al.*, 2002; Varma and Palsson, 1994). However, the dynamic models that have been used thus far are very simple in that they assume that transport rates are only proportional to the current population size. This makes them applicable only to certain intervals during growth in batch cultures. The problem is that in batch cultures, nutrient consumption rates are not just a function of the population size; the current growth phase also affects the relevant processes. Moreover,

the dynamic models that have been used thus far cannot account for transport patterns that qualitatively change during growth. For example, in our cultures we observed alanine to initially accumulate in the medium, and then switch to consumption after some time. Previously used models cannot account for such events.

Due to the methodological limitations mentioned above, we propose (and use) a new system for dealing with the dynamic aspect of the hybrid models. Essentially, the new system uses two equation forms; one basic formulation that can adequately describe transport patterns that do not qualitatively change during growth, and another that is better suited for transport patterns that do qualitatively change. Both forms are individually applied to each nutrient, by solving inverse problems (parameter optimization), and the decision to use the more complex form instead of the other for the final model is made based on clear indications of improvement, or the lack thereof. The process can be easily automated. Accordingly, not only can our proposed system account for qualitatively changing transport patterns, but it can actually detect their presence, through the automatic selection between the two equation forms. Moreover, since qualitative changes in the system are often the result of, or at least related to, another event, our method can actually uncover relationships between metabolites in the form of coordinated pattern changes. We demonstrate the value of the system by applying it to the *Halobacterium salinarum* growth data. The method was able to detect an interaction between arginine and ornithine transport, which is an interaction rooted in the well-known arginine fermentation pathway. The results also uncovered possible interaction dynamics between other metabolites, for which we could form biological hypotheses regarding the nature of the interactions (e.g., competitive inhibition). The modeling system is explained in detail in Chapter 4.

Integrative analysis of growth

The growth simulations performed in this study (Figure 1.1, transition G) allowed for an integrated analysis of bioenergetics, nutrient utilization and biomass production. Genome-scale metabolic networks conveniently link these three aspects of cellular growth. For the case of *Halobacterium salinarum*, we made several unexpected, biologically relevant findings, including: (1) that the organism degrades essential amino acids even under phototrophic conditions, when the cells should already have abundant energy. Such behavior can lead to cells terminating their growth earlier than necessary; (2) that carbon incorporation rates were surprisingly low under both aerobic and phototrophic conditions. During phototrophy, one would typically expect nutrient consumption to be just in quantities that

are sufficient for biomass production since energy can already be derived from light rather than from the supplied nutrients. Due to the low incorporation rates, the model predicts that considerable amounts of metabolites are likely produced and accumulated in the medium; and (3) that arginine fermentation occurs simultaneously with, and is in fact comparable in terms of energy output to, respiration and photosynthesis. This is in contrast to the common assumption that the fermentation process only plays a secondary, alternative role to the two main bioenergetic modes. Experimental verification was performed on some of the predictions (Figure 1.1, transition H), including the rapid accumulation of ornithine due to arginine fermentation (Chapter 4).

Chapter 2

The Halophilic Archaea

2.1 Natural Environments

The extremely halophilic archaea are a diverse group of euryarchaeota that inhabit hypersaline environments such as salt lakes, salt ponds and marine salterns. These environments can vary widely in their ionic composition depending on the surrounding geology and general climatic conditions. For example, hypersaline marine lagoons, which originate by the evaporation of sea water, accordingly reflect the ionic composition of sea water. Environments of this type, referred to as thalassohalines, have Na^+ as the dominant cation, Cl^- as the main anion, and have pH which is neutral to slightly alkaline. In contrast, other environments have ionic composition that differ greatly from sea water. For example, in the Dead Sea, divalent cations (Mg^{2+}, Ca^{2+}) are far more abundant than monovalent cations (Na^+) (Ma'or *et al.*, 2006). Such environments are termed athalassohaline. Two examples of hypersaline environments, namely Lake Zug from Wadi Natrun in Egypt and a salt pond in the Arabian desert, are depicted in Figure 2.1.

The Great Salt Lake in Utah, classified as thalassohalinic, has an ionic composition that is essentially a ten-fold concentration of sea water. Given the considerable challenges that extremely high concentrations of ions pose, it is remarkable that life not only survives, but thrives under such conditions. Microorganisms which have been isolated from the Great Salt Lake include photosynthetic algae (two strains of *Dunaliella*) (Van Auken and McNulty, 1973), aerobic archaea (strains of *Halobacterium* sp. and *Halococcus*) (Post, 1977a,b), anaerobic archaea (*Methanohalophilus mahii*) (Paterek and Smith, 1985, 1988), aerobic heterotrophic bacteria (*Halomonas variabilis*, *Pseudomonas halophila* and *Gracilibacillus halotolerans*) (Dobson and Franzmann, 1996; Fendrich, 1988; Waino *et al.*,

1999), and anaerobic bacteria (*Halanaerobium praevalens*, *Halanaerobium alcaliphilum*, *Desulfobacter halotolerans* and *Desulfocella halophila*) (Zeikus, 1983; Zeikus *et al.*, 1983; Tsai *et al.*, 1995; Brandt and Ingvorsen, 1997; Brandt *et al.*, 1999). Similarly, in the athalassohalinic Dead Sea, which, given its exceedingly high concentrations of divalent cations (1.9 M Mg^{2+}, 0.4 M Ca^{2+}), can be hostile even to organisms that are already adapted to the highest salt concentrations, numerous organisms have also been isolated. These include the halophilic archaea *Haloarcula marismortui* (Ginzburg *et al.*, 1970), *Haloferax volcanii* (Mullakhanbhai and Larsen, 1975), *Halorubrum sodomense* (Oren, 1983a) and *Halobaculum gomorrense* (Oren *et al.*, 1995), the aerobic halophilic bacteria *Halomonas halmophila* (Dobson *et al.*, 1990), *Chromohalobacter marismortui* (Ventosa *et al.*, 1989), *Chromohalobacter israelensis* (Rafaeli-Eshkol, 1968) and *Salibacillus marismortui*, the anaerobic halophilic bacteria *Halobacteroides halobius* (Oren *et al.*, 1984), *Sporohalobacter lortetii* (Oren *et al.*, 1987) and *Orenia marismortui*, and even halophilic funji (Buchalo *et al.*, 1998, 1999).

Figure 2.1: Examples of hypersaline environments. Left: Lake Zug from Wadi Natrun, Sahara Desert, Egypt (photograph from Antiquity 77, No 296, June 2003). The amount of water in the lakes of Wadi Natrun varies greatly during the year, and lakes can fuse or disappear. Egyptian craftsmen have used the evaporitic deposits from these lakes for over 6,000 years. *Natronomonas pharaonis* was isolated from lake Gabara of the Wadi Natrun (Soliman and Truper, 1982). **Right:** A salt pond in the Arabian desert. The red color is due to the presence of *Halobacterium salinarum*.

Hypersaline lakes, both natural as well as man-made such as saltern crystallizer ponds, are generally colored red due to the presence of dense communities of halophilic microorganisms (Oren and Rodriguez-Valera, 2001). The red color is due to the C-50 carotenoids, mainly α-bacterioruberin and derivatives (Kushawa *et al.*, 1975; Kelly *et al.*, 1970), of the Halobacteriaceae (archea), the massive amounts of β-carotene which are produced by

Dunaliella (algae) under suitable conditions, and, recently, a carotenoid or carotenoid-like compound from the bacteria *Salinibacter* that differs from archaeal bacterioruberins and algal L-carotene. Although numerous factors contribute to the red colorization, there is consensus that the optical properties of saltern brines are primarily determined by the archaeal community because of the low in vivo optical cross-section of the algal L-carotene (Oren *et al.*, 1992, 1994) and the relatively smaller *Salinibacter* population. Given the massive blooms of extremely halophilic organisms that can be observed through the red colorization of their habitats, salt produced from solar salterns, when not processed further, can easily spoil. A gram of solar salt may contain millions of viable cells that can survive for several years under practical storage conditions (Tansil, 1984).

We mentioned earlier that the characteristics of hypersaline environments are greatly influenced by the surrounding geology and general climatic conditions. Indeed, the ionic composition of these places can vary dramatically from year to year based on these. For example, the Dead Sea, which with a shoreline located at 415 m below sea level is the lowest place on Earth, receives fresh water inflow from flood runoff and from the Jordan river during the rainy season. The year-to-year variability in the amount of these inflows is high; in a very rainy year the amount can be twice as much as the average, and in a particularly dry year the inflow can be negligible. Accordingly, ion concentrations drop as a result of large inflows due to dilution, and increase during periods where there is little or no significant freshwater run (Oren, 2002).

In addition to the relatively immediate consequences that the changing inflows have made on the lake, they have also had a cumulative, long-term effect. The water balance of the Dead Sea has been negative since the beginning of the 20th century (Gavrieli *et al.*, 1999). This, in part, is because of changing climatic conditions, but is also the result of intensified human intervention in the lake's water regime. Significant diversion of the freshwater supply from the Sea of Galilee and the Jordan river started in the early 1950's. As a result, the ionic strength of the lake, particularly in the upper layers, has also steadily increased (Figure 2.2).

2.2   Taxonomy

Members of the taxonomical group referred to now as archaea have long been considered as bacteria (archaebacteria) because of their prokaryotic morphology, circular genomes, and gene organization in operons. However, with the advent of sequencing and other molecular
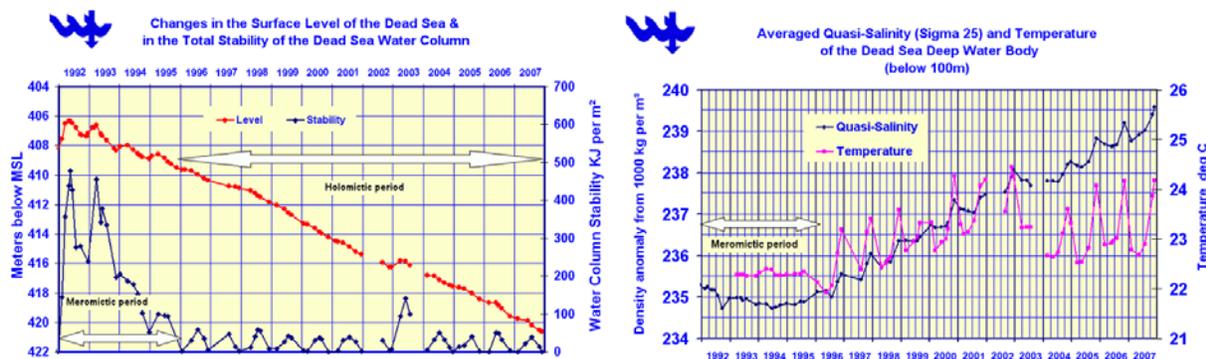
Figure 2.2: Long term changes in the Dead Sea. **Left: Water level of the Dead Sea.** The water balance of the lake has been negative since the beginning of the 20th century (data starting from 1992 reflected in the figure). This decrease in water level was caused by climatic changes and human intervention in the lake's water regime. **Right: Salinity of the Dead Sea.** The decreasing water level of the lake has been accompanied by increasing ionic concentrations. Salnity is expressed as sigma units, indicating the density excess to the standard reference of 1000 kg m^{-3}. The charts were taken from the Israel Oceanographic and Limnological Research (ISRAMAR) website (http://isramar.ocean.org.il/DeadSea/, accessed June 2008).

technologies, the inadequacies of the then existing taxonomic classification scheme were realized. In 1977, Woese and Fox showed that the archaea could clearly be distinguished from the bacteria and the eucaryotes as a third domain of life by applying rRNA phylogeny (Woese and Fox, 1977). The elevation of the archaea to a separate domain was further supported by their distinctive features such as cell membranes containing prenyl side chains that were ether-linked, rather than ester-linked, to a glycerol molecule (sn-glycerol-1-phosphate) that has a steriochemistry different from that found in bacteria. Further work on archaeal genomes revealed that the archaea also share features with the eucarya, and in fact may be closer to them than to the bacteria with which they were orginally grouped together. In this respect we should also mention that the archaea have been proposed as a possible bridge between the other two domains (Eisen, 1999).

Although this study focuses on the haloarchaea, halophilic microorganisms can be found in all three domains of life: Archaea, Bacteria and Eucarya (Oren, 1999). In each of these domains are representative organisms that can grow at the highest salt concentrations. For example, the family *Halobacteriaceae* of the archaea and the unicellular algae of the genus *Dunaniella* in the domain Eucarya can be found in the Dead Sea, which is a NaCl-saturated environment. Likewise, the recently discovered *Salinibacter ruber* and members of the genus *Halorhodospira*, all from the bacterial domain, can be found in saltern crystallizer ponds and soda lakes, respectively. With that said, it is remarkable that halophilic microorganims are often interspersed between non-halophilic relatives in the phylognetic tree (Oren, 2002,

p.23-24). This suggests that adaptation to life in hypersaline environments may have arisen many times during evolution (Falb, 2005, p.5). However, it must be noted that there are some notable cases of phylogenetically coherent groups consisting entirely or almost entirely of halophiles. These are the order *Halanaerobiales* and the family *Halomonadaceae* of the bacteria, and the family *Halobacteriaceae* of the order *Halobacteriales* in the domain Archaea.
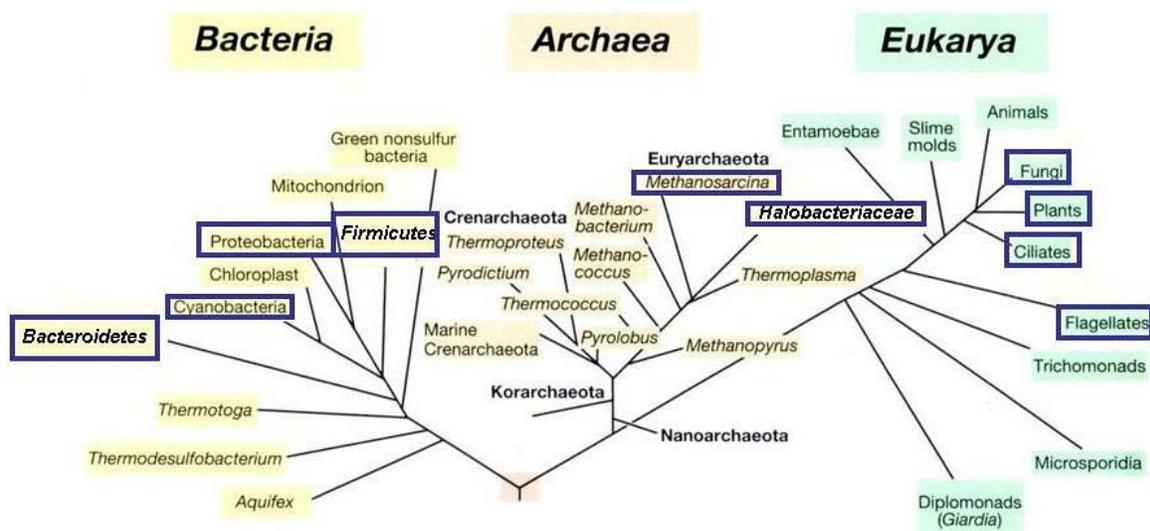
Figure 2.3: Distribution of halophilic microorganisms. Halophilic microorganisms can be found in all three domains of life. In the figure, groups which have at least one halophilic representative are marked with blue boxes. The figure was taken from Oren, 2008.

Currently known halophilic microorganisms occur within the domain Archaea in three families; the *Halobacteriaceae*, the *Methanospirillaceae* and the *Methanosarcinaceae*. This work focuses on the first, which is unlike the other two families as they include members that are not halophilic. Indeed, the *Halobacteriaceae* includes some of the the most salt-requiring and salt-tolerant microorganisms that have been found to date. In part due to the interest generated by its members' strict dependence on high salt concentrations for growth and structural stability, the *Halobacteriaceae* have been the subject of most studies that involve halophilic microorganisms.

The *Halobacteriaceae*, currently the only family within the order *Halobacteriales*, forms a branch within the *Euryarchaeota* near the *Methanonomicrobiales*. Some two decades ago, the diversity within this family was limited to rod-shaped, *Halobacterium*-type cells and to cocci of the genus *Halococcus*, and was therefore considered low. Since that time however, many new isolates exhibiting remarkable diversity have been found. With respect to morphology, the *Halobacteriaceae* now includes flat and extremely pleomorphic cells

(Mullakhanbhai and Larsen, 1975), the species *Haloarcula japonica* which exhibits triangular and trapezoidal conformations (Takashina *et al.*, 1990), and the truly remarkable *Haloquadratum walsbyi* (Figure 2.2) which is flat and perfectly square in shape (Walsbyi, 1980; Bolhuis *et al.*, 2004, 2006; Burns *et al.*, 2007). The family *Halobacteriaceae* (and accordingly also the order *Halobacteriales*) currently has 36 validly described species (Table 2.1).
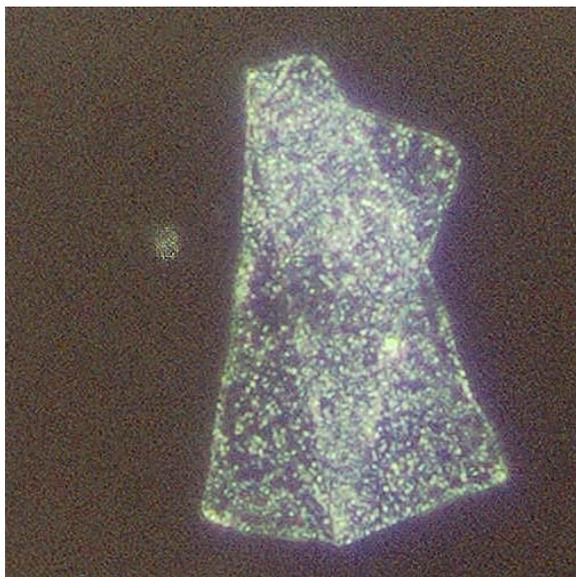
Figure 2.4: Darkfield micrograph of *Haloquadratum walsbyi.* This halophilic archaeon was first described by Walsbyi in 1980, characterizing it as having the form of a thin rectangular sheet (Walsbyi, 1980). *H. Walsbyi* dominates aquatic environments with high concentrations of NaCl and $MgCl_2$, the latter posing additional dessication stress by lowering water activity to extremely low levels. The image was extracted from Bolhuis *et al.*, 2004.

2.3   Physiology and Metabolism

2.3.1   Osmoadaptation

Biological membranes are permeable to water. For this reason, it is impossible for microorganisms living in hypersaline environments to maintain a hypoosmotic intracellular condition with respect to the osmotic pressure of the medium in which it lives. This is true even for halophilic organisms. Given that at least some level of Turgor pressure has to be maintained to allow expansion and growth of cells, various strategies that enable cells to cope with the osmotic stress have evolved. Of the various mechanisms that have been observed, the most common in nature is the accumulation of organic compatible solutes. The term originally coined by Brown and Simpson (Brown and Simpson, 1972), compatible solutes increase the osmotic pressure of the cytoplasm to provide osmotic balance. They are termed *compatible* because they do not significantly disrupt

Table 2.1: Genera and species under the order *Halobacteriales*.

Genus	Species	Genus	Species
Halobacterium	*Halobacterium salinarum*	*Halococcus*	*Halococcus morrhuae*
Halobaculum	*Halobaculum gomorrense*		*Halococcus saccharolyticus*
Halorubrum	*Halorubrum saccharovorum*		*Halococcus salifodinae*
	Halorubrum sodomense	*Natrialba*	*Natrialba asiatica*
	Halorubrum lacusprofundi		*Natrialba magadii*
	Halorubrum coriense	*Natronobacterium*	*Natronobacterium gregoryi*
	Halorubrum distributum	*Halogeometricum*	*Halogeometricum borinquense*
	Halorubrum vacuolatum	*Natronococcus*	*Natronococcus occultus*
	Halorubrum trapanicum		*Natronococcus amylolyticus*
Haloarcula	*Haloarcula vallismortis*	*Haloferax*	*Haloferax volcanii*
	Haloarcula marismortui		*Haloferax gibbonsii*
	Haloarcula hispanica		*Haloferax denitrificans*
	Haloarcula japonica		*Haloferax mediterranei*
	Haloarcula argentinensis	*Natrinema*	*Natrinema pellirubrum*
	Haloarcula mukohataei		*Natrinema pallidum*
	Haloarcula quadrata	*Haloterrigena*	*Haloterrigena turkmenica*
Haloquadratum	*Haloquadratum walsbyi*	*Natronorubrum*	*Natronorubrum bangense*
Natronomonas	*Natronomonas pharaonis*		*Natronorubrum tibetense*

The list was adapted from Oren, 2006.

processes that normally occur within cells. For example, they are very poor enzyme in-
hibitors. Currently known molecules which are used as compatible solutes include betaine
(Imhoff and Rodriguez-Valera, 1984), ectoine (Galinski *et al.*, 1985; Galinski, 1995), amino
acids with different steriochemistry such as Nϵ-acetyl-β-lysine (Sowers *et al.*, 1990) and β-
glutamine (Lai *et al.*, 1991), and even carbohydrates derivatized to reduce reactivity like
glucosylglycerol and α-mannosylglyceramide (Silva *et al.*, 1999; A *et al.*, 2005).

The *Halobacteriaceae* produce no organic solutes. Rather, they use the inorganic ions
that are already in their surroundings to achieve osmotic balance. For example, *Halobac-
terium* pumps large amounts of potassium into the cytoplasm so that the intracellular K$^+$
concentration is considerably greater than the extracellular Na$^+$ concentration. This en-
sures that cells maintain a positive water balance and counteract the tendency to become
dehydrated at high osmotic pressure and ionic strength. The use of such inorganic ions is
radically different from using organic compatible solutes because the ions themselves reach
very high concentrations inside the cells. A total Na$^+$ and K$^+$ concentration reaching as
high as 6.12 M has been reported (Matheson *et al.*, 1976). In contrast, salt concentra-
tions in microorganisms that predominantly use organic compatible solutes remain very
low. This is a very important distinction because the intracellular machinery of halophilic
microorganisms that use nonorganic ions as their solute will further have to be adapted
to a hypersaline cytoplasm. The presence of salts in molar concentrations is in general

devastating to proteins and other macromolecules. Under such conditions, hydrophobic interactions are enhanced, potentially essential electrostatic interactions are interfered with due to charge shielding, and water availability is lowered (Oren, 2002; Madern *et al.*, 2000; Zaccai and Eisenberg, 1991). Indeed, unadapted proteins tend to aggregate or collapse, and often lose function. One way through which the *Halobacteriaceae* have coped is by using large amounts of the acidic amino acids glutamate and aspartate in most of their proteins, and by having only very little of the basic amino acids lysine and arginine. Such a strategy leads to an excess of negative charges on protein surfaces, which can then help to prevent aggregation. Proteins of the *Halobacteriaceae*, with some exceptions, have become so specialized that they denature and lose activity when not in high salt concentrations. For these reasons, members of the family are considered as the hardcore of the halophilic community.

2.3.2 Nutrition and Transport

The members of the *Halobacteriaceae* are varied in their nutritional requirements. In the past, the general consensus was that their demands are complex, to the extent that some could only be met in culture through the inclusion of large concentrations of yeast extract or other rich sources of nutrients. In the case of *Halobacterium salinarum*, then and up to now the most well-studied archeal halophilic species, defined media designed for its isolates typically contained between 10 and 21 amino acids, and in some cases had to be further supplemented with vitamins, nucleosides, and glycerol (Dundas *et al.*, 1963; Grey and Fitt, 1976; Onishi *et al.*, 1965; Shand and Perez, 1999). However, the prevailing view changed when *Haloferax mediterranei* was later isolated because it was found to be able to grow on simple compounds, such as succinate and acetate, as single carbon and energy source (Rodriguez-Valera *et al.*, 1980; Juez and Kushner, 1983). Indeed, such simple growth requirements are now typically found in the genera *Haloferax* and *Haloarcula* (Oren, 2002, p.126), although vitamins and some other supplements can have stimulatory effects.

The *Halobacteriaceae* are basically aerobic heterotrophs. The nutrients which they use come from the other microorganisms present in their hypersaline environments. For example, glycerol and tricarboxylic acid cycle intermediates are excreted by the alga *Dunaliella* and the cyanobacterium *Microcoleus chtonoplastes* (Zvyagintseva *et al.*, 1995), respectively. Since the halophilic archaea are often the most halophilic in their communities, the materials left after the eventual death of other microorganisms as a result of increasing salinity also

become available to them (see Section 2.5). Accordingly, their environments can become quite rich in nutrients. Most halophilic archaea preferentially use amino acids as carbon and energy source (Oren, 2002, p.127). In fact, some organisms such as *Halobacterium salinarum* require at least a number of them in order to grow due to deficiencies in biosynthetic capabilities (Gonzalez *et al.*, 2008). With respect to carbohydrates, simple sugars such as glucose and sucrose are not readily used by all members. In fact, active transport of glucose has been experimentally excluded for *Halobacterium salinarum* (Severina *et al.*, 1991). In contrast, species of *Haloferax* and *Haloarcula* can grow with the hexose as the only carbon source. *Halococcus saccharolyticus*, *Haloferax valcanii* and *Halorubrum saccharovorum* excrete acetate when grown on glucose, although they are able to use acetate as well (Oren, 2002, p.129).

Various amino acid transport systems in the cytoplasmic membrane of *Halobacterium salinarum* have been characterized. Cell envelope vesicles have been shown to accumulate 19 commonly occurring L-amino acids, in response to either a light-induced membrane potential or the establishment of a sodium gradient, or both (MacDonald *et al.*, 1977). The carrier systems were found to be as variable in their properties as those found in other organisms. For example, while some are highly specific to a particular amino acid, others transport several amino acids competitively. Some, like glutamate, are activated by a chemical gradient of sodium only. In contrast, others can function in the complete absence of such a gradient. Many of the amino acid transport systems specifically depend on Na^+ ions. These include leucine (MacDonald and Lanyi, 1975) and glutamate (Birkeland and Ratkje, 1985; Lanyi *et al.*, 1976). The transport of glutamate is unique in that its uptake is driven by the sodium gradient only.

Two transport systems for acetate and propionate have been detected in the alkaliphilic *Natronococcus occultus*. One is a high affinity system that is driven by the Na^+ gradient over the membrane (Kevbrina *et al.*, 1989). Glucose and fructose are also transported by sodium-driven systems in *Haloferax volcanii* (Takano *et al.*, 1995; Tawara and Kamo, 1991). In contrast, neither a Na^+ nor a H^+ gradient alone drives active phosphate transport in *Halobacterium salinarum*. Rather, the process depends on cellular ATP level (Zoratti and Lanyi, 1987). Multi-drug efflux transporters have been described in *Haloferax volcanii* (Miyauchi *et al.*, 1992, 1997).

2.3.3   Motility and Taxis

The *Halobacteriaceae* currently have two mechanisms described for motility: the use of flagella and the production of gas vesicles. In the case of *Halobacterium salinarum*, flagella are right-handed and helical, and can generate forward and backwards directed movement by rotating clockwise and counterclockwise, respectively (Alam and Oesterhelt, 1984). This is quite different from the flagella of bacteria which are left-handed. Cells are predominantly monopolar in the logarithmic phase, and bipolarly flagellated in the stationary phase. Flagellar bundles do not fly apart during the switch from clockwise to counter clockwise rotation or vice versa (Marwan *et al.*, 1991). While flagella-related genes have been identified in many species of the order *Halobacteriales*, including *Haloarcula marismortui* and *Natronomonas pharaonis*, and very likely these use a similar system, the use of flagella is not universal. For some members like *Haloquadratum walsbyi*, there is no indication, whether physiological or genetic, of the existence of an analogous system. An electron microscope image of a *Halobacterium salinarum* cell where the flagellum is readily visible is provided in Figure 2.5-left.
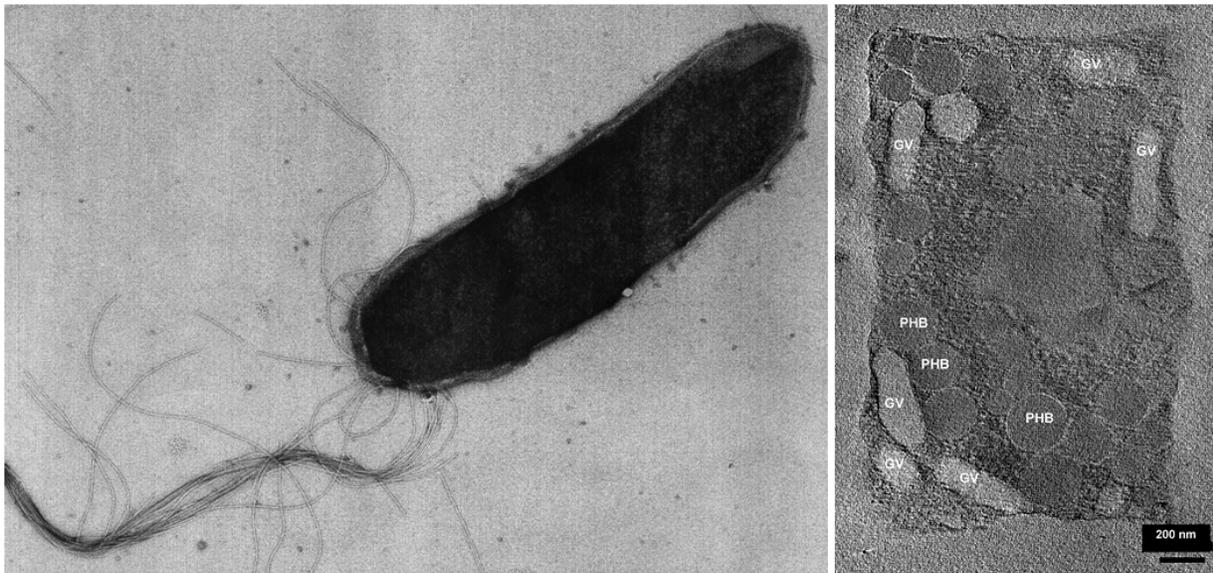
Figure 2.5: Motility structures. (Left:) Electron microscope image of a single *Halobacterium salinarum* cell. The flagellum is readily visible. Unlike their bacterial counterparts, haloarchaeal flagella are right-handed. (Right:) Electron tomographic image of a single *Haloquadratum walsbyi* cell. Gas vesicles (GV), recognizable by their spindle shapes, are found at the borders of the cell. The image was taken from Bolhuis *et al.*, 2004.

Gas vesicles form the second class of motility structures that have been described for the *Halobacteriaceae* (Figure 2.5-right). These are gas-filled subcellullar organelles

that enable cells to float on or near the surface of their aquatic environments by conferring buoyancy. Such a capability can be especially advantageous for halophilic microorganisms because oxygen, which can quickly become limiting in hypersaline environments, is generally more abundant in the upper water levels (Beard *et al.*, 1997). Gas vesicles have cylindrical or spindle-shaped morphology that is determined genetically (Simon, 1981). They have proteinaceous membranes that are extremely stable, rigid, gas permeable, and lipid-free. Gas vesicles have been reported for *Halobacterium salinarum* (Shukla and DasSarma, 2004; Halladay *et al.*, 1993), *Haloferax mediterranei*, *Halogeomtricum borinquense* (Montalvo-Rodriguez *et al.*, 1998), *Halorubrum vacuolatum* (Mwatha and Grant, 1993), and *Haloquadratum walsbyi* (Walsbyi, 1980; Bolhuis *et al.*, 2004).

Halobacterium salinarum was screened for substances that could induce chemotactic behavior (Storch *et al.*, 1999). These include amino acids and derivatives, dipeptides, carbohydrates, central metabolites, inorganic ions, and a few esoteric compounds. Of the over 80 tested, six amino acids, namely leucine, isoleucine, valine, methionine, cysteine and arginine, and several peptides were identified as chemoattractants. Three compounds known to be typical bacterial repellents also elicited phobic responses. In addition to chemotaxis, *Halobacterium salinarum* has also been described to exhibit aerotactic behavior that is mediated by two oxygen-sensing transducers (Hou *et al.*, 2000) as well as phototaxis through sensory rhodopsins (Krah *et al.*, 1994; Bogomolni and Spudich, 1982; Hildebrand and Dencher, 1980).

2.4 Completely Sequenced Genomes

At the time of writing (October 2008), the genomes of six haloarchaeal strains have been completely sequenced. *Halobacterium salinarum* strain NRC-1 was published in 2000 as *Halobacterium* sp. NRC-1 (Ng *et al.*, 2000). The genome of another strain, *Halobacterium salinarum* R-1, also became available at about the same time, but was only published recently (Pfeiffer *et al.*, 2008a). This species is particularly noteworthy as it has emerged as one of the model organisms of the archaea, in part due to the ease with which it is cultured in the laboratory (DasSarma *et al.*, 1995), and also because it is the model organism for producing bacteriorhodopsin. The genome of the Dead Sea isolate, *Haloarcula marismortui*, was published in 2004, and was extensively compared with *Halobacterium* sp. NRC-1. In addition to *Halobacterium salinarum* strain R1, our lab has also completed

sequencing the genomes of *Natronomonas pharaonis* (Falb *et al.*, 2005) and *Haloquadratum walsbyi* strain HBSQ001 (Bolhuis *et al.*, 2006). The genome sequence of *Haloferax volcanii* is available on the University of Maryland Biotechnology Institute (UMBI) website and on the Halolex database (Pfeiffer *et al.*, 2008b), but still awaits publication. Sequencing projects are currently underway for more members of the family *Halobacteriaceae*; these are *Halobaculum gomorrense* (Allers and Mevarech, 2005), the psychrotolerant halophile *Halorubrum lacusprofundi* (Allers and Mevarech, 2005), a second *Haloquadratum walsbyi* strain, C23$^{\text{T}}$ (Dyall-Smith and Oesterhelt, unpublished), a third *Halobacterium salinarum* strain, GRB (Oesterhelt, unpublished), *Natrialba magadi*, and *Natronomonas moolapensis* (Dyall-Smith and Oesterhelt, unpublished), which was previously called isolate 8811.

2.5   Dynamics of Blooms

Because the properties of hypersaline environments are tightly linked to their surrounding geology and general climatic conditions, halophilic microorganisms are also affected by these abiotic factors. Perhaps the best characterized example is that of the Dead Sea. Systematic monitoring of the spatial and temporal distribution of the microbial communities in the Dead Sea has been ongoing since 1980, and this has enabled a reasonable understanding of the biological processes in the lake, the main components of its biota, and the dynamics of their appearance and decline (Oren, 2002; Oren and Gurevich, 1995a, p424). As will be discussed in Section 4, the dynamics brought about by abiotic factors can be very important considerations for interpreting the observed behavior of cells and the results of metabolic reconstructions.

The unicellular green alga *Dunaliella parva* is responsible for all of the primary productivity in the Dead Sea. However, no growth of the alga is possible when the salt concentration of the water column is invariably high. Blooms of *Dunaliella* in the Dead Sea occur only after significant dilution of the upper levels by the influx of freshwater. For example, a very rainy winter (1979-1980) preceeded the summer of 1980. The large amounts of rain water that entered the lake at that time initiated a meriomitic episode (Oren and Shilo, 1982), i.e., the stratification of the water column where salinities are significantly lower at the upper levels. This enabled the *Dunaliella* population to grow near the surface. In turn, this allowed the archaeal community to bloom at the expense of the organic material produced by the alga. Although the halophilic archaea do not require the reduction in salinity to grow, their dependence on primary producers nevertheless makes

them subject to the abiotic factors. Given that these blooms of the primary producer can be years in between, growth of the dependent microorganisms can likely become very competitive.

2.6 Motivation

The existence of microorganisms that thrive in physically or geochemically extreme conditions that are detrimental to the majority of life on Earth is a subject that has received much interest. Among these extremophiles are the halophilic archaea, which have become so adapted to their environments that they require hypersaline conditions in order to live. To date, five haloarchaeal species have been completely sequenced, six if the two *Halobacterium* strains are counted separately, and more are currently underway.

Genome sequences are rich sources of information for understanding microorganisms because they define the molecular inventories of the cells. Genome sequences can be mined, for example, using computational methods for gene identification, functional assignment, and metabolic pathway reconstruction. In this study, we perform metabolic reconstruction for two haloarchaeal species, namely *Halobacterium salinarum* and *Natronomonas pharaonis*, and develop and use computational models derived from these. Reconstructed networks by themselves are already valuable because they, by revealing significant portions of the molecular inventories of the microorganisms and organizing them into coherent groups, can already lead to biologically interesting predictions and insights. However, the use of appropriate computational methods can even further expand their utility, for example by allowing growth simulations under specific conditions. Given the ongoing sequencing projects for the *Halobacteriaceae*, not to mention those for halophilic members of other taxonomic groups, comparative genomics will become even more important in the near future for providing insight into life in hypersaline environments. We believe that a considerable fraction of the discoveries that will be made from these sequencing efforts will come from the analysis of metabolic networks, using network reconstruction in concert with methods from systems biology.

Chapter 3

The Metabolism of *Halobacterium salinarum*

This chapter deals with the reconstruction of the metabolic network of *Halobacterium salinarum*. We begin with an introduction to the halophile by describing its physiology, with a particular emphasis on aspects of reasearch interest (Section 3.1). Next, we give a short introduction to metabolism and to the typical steps taken in reconstructing a metabolic network (Section 3.2). Full details on the reconstruction methods used for *Halobacterium salinarum* are provided in Section 3.3, where the importance of extensive manual curation for producing a network with high reliability is also dicussed.

In Section 3.4, we start to discuss the actual results of the metabolic reconstruction. One of the most significant uses of such reconstructed metabolic networks is to serve as foundations for the development of computational models (discussed in succeeding chapters). However, even prior to any modeling effort, each of of these networks is already valuable in that it represents a summary of the current knowledge regarding the metabolism of the respective organism (*Halobacterium salinarum* in this case). Uses of such a network include: (1) by simple inspection of the pathways one could already say a lot about the physiology of the microorganism; e.g., looking at the catabolic pathways that are present reveals what nutrients the organism could use to derive energy. The pathways also define what possible fates supplied nutrients could have, and as such are critical for labeling studies. Detailed discussions are provided for several pathways, including central metabolism (Section 3.5), synthesis and degradation of amino acids (multiple sections), fatty acid metabolism (Section 3.6.6), and nucleotide metabolism (Section 3.7.1). A second use of reconstructed networks is that (2) through a more detailed analysis, one could identify

interesting regions – such as those associated with missing functions in the network that, based on physiological data, should be present –, and subsequently form biological predictions. Two examples of these "knowledge gaps" that we found include missing initial steps in the sythesis of aromatic amino acids (Section 3.7.3), and the lack of a synthetic pathway for riboses (Section 3.7.2). Separate research efforts aimed at elucidating these two subjects were initiated, the work done by colleagues, in part due to the observations made in this study. Other such gaps, which point out regions in the metabolic network where possible novel mechanisms may lie, are also discussed. Finally, (3) the reconstruction effort itself lead to improvements/updates to the existing primary anotation of the halophile's genome.

3.1   The Model Archaeon

Halobacterium salinarum is a model organism for the halophilic branch of the archaea. Although its doubling time is considerably longer than *E. coli*, it is as easy to culture in the lab, which makes it an excellent system for investigating archaeal genetics and functional genomics. *Halobacterium salinarum* is rod-shaped (Figure 3.1, right), motile, and is one of the few species known that can live in saturated salt solutions. It, as well as other Halobacteria, can be found in the Great Salt Lake, the Dead Sea, and other hypersaline environments, such as solar salterns (Figure 3.1, left). They can be easily identified in bodies of water through the reddish hue that they impart due to their bacterioruberins.

3.1.1   Bacteriorhodopsin and Other Retinal Proteins

Halobacterium salinarum has four related retinal proteins in cell membranes; these are bacteriorhodopsin (BR), halorhodopsin (HR), and the two sensory rhodopsins SRI and SRII. These proteins have been the subject of much study, and to a large extent have fueled the sustained interest in the organism. Halorhodopsin is involved in maintaining iso-osmolarity of the cytoplasm during cell growth by transporting chloride ions, nitrate, and other halides into the cell using light energy (Bamberg *et al.*, 1993; Schobert and Lanyi, 1982). Its structure has been resolved at 1.8 Å resolution (Kolbe *et al.*, 2000), which allowed for very detailed descriptions of its mechanisms. The two sensory rhodopsins (SRI and SRII) mediate phototactic behavior which permits cells to avoid harmful blue and UV light, and to accumulate in regions favorable for photosynthesis (Hoff *et al.*, 1997; Spudich, 1998). Finally, bacteriorhodopsin is a light-driven proton pump, which is perhaps the single
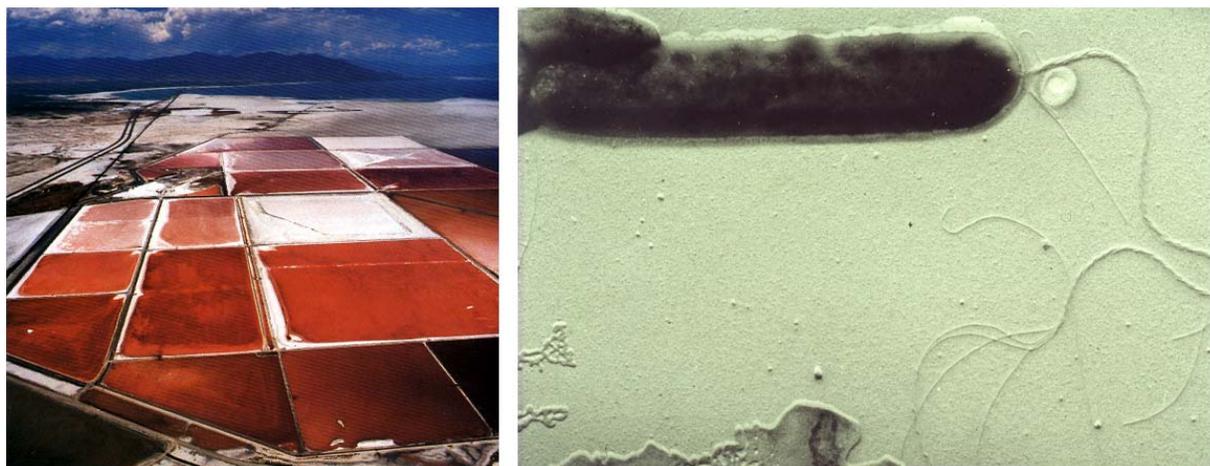
Figure 3.1: Massive Growth of halophilic archaea in a saltern. Left: Massive growth of halophilic archaea can be seen in this aerial view of the "Salterns of Torrevieja" in Costa Blanca, Spain. Seawater is evaporated in these salterns to produce solar salt. The red-purple colour is mainly due to bacterioruberins of the halophilic archaea, including *Halobacterium sali-narum*. **Right:** An EM image of a *Halobacterium salinarum* cell. The flagella, which gives the microorganism motility, is clearly visible.

subject matter that *Halobacterium salinarum* is most well-known for.

BR is currently the only known nonchlorophyll structure that allows photosynthesis (Oesterhelt, 1988). It converts the energy of green/orange light (500 - 650 nm) into an electrochemical proton gradient, which in turn can be used by ATP-synthase to produce energy. For this reason, it is a critical player in cellular metabolism during phototrophy, and is therefore an integral component of this study.

Bacteriorhodopsin is the best understood ion transport protein, and has become a paradigm for membrane proteins in general and transporters in particular (Haupts *et al.*, 1999). The numerous investigations into its atomic structure using X-ray crystallography or electron diffraction at various resolutions (Luecke *et al.*, 1999; Grigorieff *et al.*, 1996, and many others) have resulted in detailed characterizations of the subprocesses involved in proton translocation and to the identification of the relevant structural parts. These include the proton release pathway on the extracellular side, the reprotonation pathway in the cytoplasmic half channel, and the cytoplasmic membrane surface for proton capture and guidance (Oesterhelt, 1988). In addition, a tremendous amount of effort has been made, using a host of biophysical methods, to determine the exact nature of the changes in BR at each step, from the absorption of photons to the vectorial transport of protons out of the cells, and these have lead to the elucidation of the complete catalytic cycle of BR at the molecular level (Haupts *et al.*, 1999).
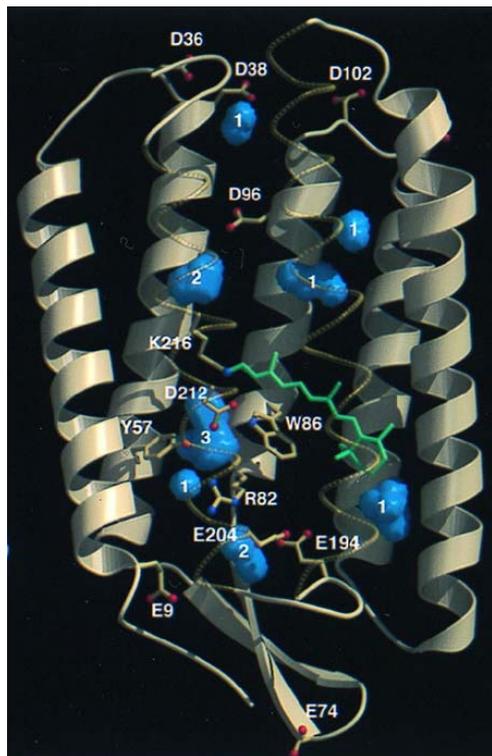
Figure 3.2: The 3-D structure of bacteriorhodopsin. Bacteriorhodopsin (BR) functions as a light-driven proton pump that can build up a proton motif force, which in turn can be used by ATP-synthases to produce ATP. BR is currently the only known nonchlorophyll structure that allows photosynthesis. It contains a retinal molecule that is responsible for absorption of green/orange light (500 - 650 nm). The image was taken from Essen *et al.*, 1998.

3.1.2   Flexible Bioenergetics

Despite the focus on bacteriorhodopsin, photosynthesis is not the only means by which *Halobacterium salinarum* can generate energy. Indeed, the bioenergetic capabilities of *Halobacterium salinarum* are quite extensive. Alternative mechanisms for producing energy that it employs include respiration and the fermentation of arginine (Hartmann *et al.*, 1980; Dundas and Halvorson, 1966). *Halobacterium salinarum* is also one of the few reported organisms that can use potassium gradients for long term energy storage, in a battery-like manner (Wagner *et al.*, 1978).

Oxygen solubility is adversely affected by high salt concentrations and elevated temperatures. Consequently, halophilic microorganisms frequently encounter microaerobic or even anoxic conditions in their hypersaline environments. Despite this, haloarchaea generally still grow heterotrophically using respiration (Müller and DasSarma, 2005). In the case of *Halobacterium salinarum*, cells under laboratory conditions have been described to consume oxygen at about 3-4 times their volume per hour (Oesterhelt and Krippahl, 1973). If their environment becomes microaerobic or anoxic, for example as a result of a high cell density promoted by aerobic growth, then cells can still employ respiration by using alternative terminal electron acceptors, such as dimethyl sulfoxide (DMSO) and trimethylamine

N-oxide (TMAO) (Müller and DasSarma, 2005; Oren and Trüpper, 1990). The respiratory chain of *Halobacterium salinarum* as well as its capacity to ferment arginine are discussed in more detail in Chapter 4.

3.1.3   Industrial Applications

Bacteriorhodopsin exhibits outstanding levels of physical and chemical stability compared to other proteins. This and the fact that its full catalytic cycle has been completely elucidated make BR a viable model for the integration of biological functional systems into technical applications. Bacteriorhodopsin is currently being developed for applications in optical security (Hampp and Neebe, 2006), optical data storage (Yao *et al.*, 2005), and holography (Barnhardt *et al.*, 2004).

The excellent optical properties of BR form the basis of its promising applications. An efficient photoreaction, with a quantum efficiency of 64%, and a high degree of reversibility make it superior to current synthetic organic materials. A characterized photochromism - the reversible transformation between two forms (states) through the absorption of electromagnetic radiation - allows for controlled manipulation of the structure. The photochromism of BR is characterized by two dominant states; the initial B-state and the M-state with maximum absorptions at 570 nm 410 nm, respectively. Exposing the structure to light with a wavelength in the green to red regime causes it to switch its color from purple to yellow. The yellow intermediate returns thermally to the purple-colored state with an adjustable time constant. Where desired, the reverse process can also be photochemically induced by exposing BR to blue light (Hampp and Neebe, 2006). Finally, as BR is a protein, its amino acid sequence can be modified using standard molecular, biological methods. In particular, some amino acid regions can be modified without influencing its function. These are excellent locations for inserting molecular tags that can be used for identification.

3.2   Introduction to Metabolic Reconstructions

3.2.1   Metabolism and Metabolic Pathways

Metabolism refers to the set of chemical processes that occur within a living organism that are necessary for the maintenance of life. These processes allow organisms to grow and reproduce, to maintain their structures, and to respond to their environments. Metabolism

is usually divided into two categories; *catabolism* and *anabolism*. The former has to do with the breakdown of large molecules, for example to harvest energy using cellular respiration. In contrast, anabolism is concerned with the construction of cellular components, such as amino acids and nucleotides, from simpler substrates, often with the use of energy. Accordingly, the metabolism of an organism determines which substances it will find useful and which it will find worthless or even detrimental.

The chemical reactions of metabolism are usually organized into metabolic pathways. These are sequences of biochemical reactions in which a chemical is transformed into another via several intermediates. Under physiological conditions, most biochemical reactions require enzymes, and often need additional cofactors in order to function properly. Enzymes are crucial to metabolism because they allow organisms to drive desirable but thermodynamically unfavorable reactions by coupling them to favorable ones. Given that metabolic pathways can be quite elaborate due to the many chemicals that may be involved, participating compounds are commonly categorized into main or side reactants. This is especially helpful when visualizing pathways as metabolic maps (Figure 3.3), since side reactants can be shown using less visible profiles or even completely omitted. The collection of all metabolic pathways within a cell is referred to as its metabolic network.

3.2.2 Metabolic Reconstruction

The rate at which biological data is accumulated has tremendously increased in recent years. For example, over the past decade the number of sequences stored in the GenBank database has followed an exponential curve with a doubling time of between 12 and 15 months. This is in no small part due to the rapid advances in sequencing technologies (reviewed in Chan, 2005). As of June 1, 2008, 812 genomes have been published, and work is underway for 90 archaea, 1766 bacteria and 936 eukarya.

With the explosion of biological data came the realization of the need for methods that can organize and analyze the accumulating information. In this context, computational methods have come into prominence in biology. Although the words are sometimes used interchangeably, the contributions of the computational sciences to biology can be roughly classified into two broad categories: *bioinformatics* and *systems biology*. Bioinformatics, as defined by the National Institute of Health (NIH) of the USA, is the "research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data". Put more simply, bioinformatics is concerned with
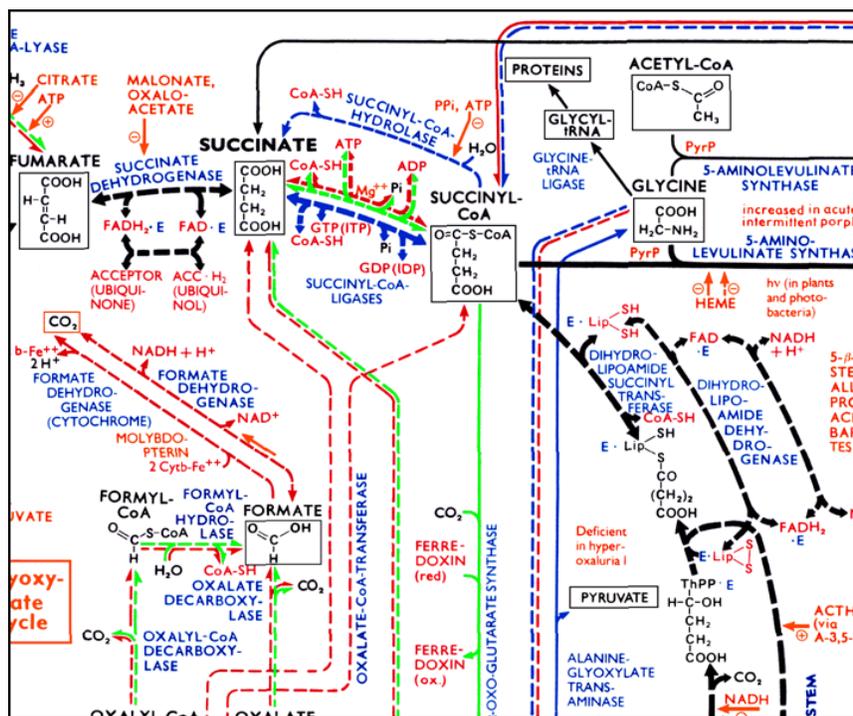
Figure 3.3: A section of the Roche Applied Science "Biochemical Pathways" wall chart showing a segment of the tricarboxylic acid (TCA) and some nearby reactions. Main reactants are indicated using more visible prints, and are shown with their chemical structures provided. Interconversions between these main components are indicated with directed arrows, while side reactants are linked using more acuate arrows. The enzymes that catalyze the reactions are shown in blue. Compounds known to influence enzyme activities are indicated by minus/plus signs, depending on their regulatory effect. Arrow color is used to define the type of organism in which a reaction occurs; black for general pathways, red for prokaryotes, blue for animals and green for plants/yeast.

the development and maintenance of databases and comparative tools for the analysis of biological data. On the other hand, systems biology is concerned with the hypothesis-driven investigation of specific biological problems using computational methods, with the primary goal of discovery and the advancement of biological knowledge.

One important aspect of the study of biological systems is the (partial) reconstruction of cellular networks. This procedure refers to the collection and organization of all cellular constituents and processes which are potentially relevant to a physiology of interest. The reconstruction serves to sort individual components into context, and provides a convenient platform for the visualization and analysis of relevant data. In the case of metabolism, the network components are enzymes, reactions and metabolites. A reconstructed metabolic network can be used to visualize high-throughput data, such as the regulatory factors

of enzyme-genes when the system is subjected to different conditions. Moreover, graph theoretic algorithms, including those for topological analysis, can be applied to the network in order to obtain insights on its overall properties (Albert *et al.*, 2000; Albert, 2005; Jeong *et al.*, 2000; Tanaka, 2005; Stelling *et al.*, 2002). And last but not least, the reconstruction can serve as a solid basis for the development of more sophisticated computational models.

3.3   Methods

3.3.1   Creating the Enzyme Inventory

Obtaining a draft metabolic network for an organism is primarily a bioinformatic problem. The typical steps taken to create such a draft network is provided in Figure 3.4. Given that most metabolic reactions are driven by enzymes, the obvious first step is to obtain an enzyme inventory. For most microorganisms, the primary resource in obtaining this list would most probably be its one-dimensional genome annotation.

Gene Finding

A one-dimensional annotation of a genome, put very simply, refers to the identification of "interesting" regions within the genome, such as those connected to genes, rRNAs, and tRNAs, and the assignment of (putative) functions to their products whenever possible. In small prokaryotes, the identification of genes (Figure 3.4, transition A), or "gene finding", is largely a matter of identifying long open reading frames (ORFs). However, some tricky situations do arise, such as when long ORFs overlap on opposite strands and cause ambiguities as to which are true coding regions (Stein, 2001). Gene finding becomes increasingly tricky as genomes get larger. Currently available gene finding tools include GLIMMER (Salzberg *et al.*, 1998), GENSCAN (Burge and Karlin, 1997), Genie (Reese *et al.*, 2000), GeneMark.hmm (Besemer and Borodovsky, 1999), Grail (Uberacher and Mural, 1991), HEXON (Solovyev *et al.*, 1994), and REGANOR (Linke *et al.*, 2006).

   GC-rich genomes are characterized by a low frequency of stop codons (TAA, TGA, TAG), which leads to a severe overprediction of potential genes. In *Halobacterium salinarum*, the GC content is 68%, and this led to an average of 1.7 additional spurious open reading frames of at least 100 codons for every gene coding for a real protein (Tebbe *et al.*, 2005; Aivaliotis *et al.*, 2007). In addition, problems related to incorrect gene start codon as-
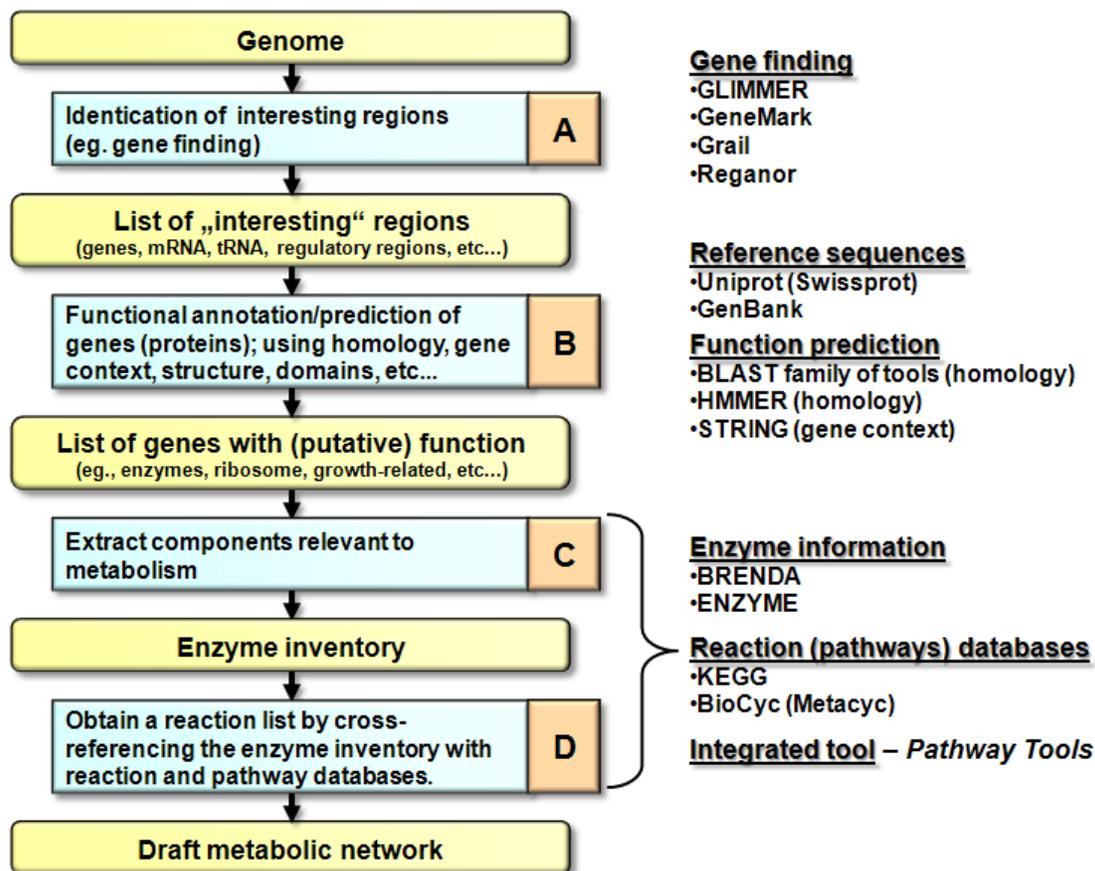
Figure 3.4: Summary of steps leading to a draft network. The yellow curved boxes (places) correspond to intermediate results in obtaining a draft metabolic network for an organism. Processes that link these intermediate results are represented by blue rectangles (transitions). Some popular tools and resources relevant to the transitions are provided to the right. Typically, the first two steps are performed by the people who publish the genome. This is true for *Halobacterium salinarum*, so we only had to use (gene) protein function prediction methods in specific cases. One-letter codes are attached to each transition for easy reference from the main text.

signment are also aggravated in the archaea because ribosome binding sites (Shine-Dalgarno boxes) are poorly conserved and mainly precede genes within transcription units but are not found in the case of leaderless mRNAs (Falb, 2005; Sartorius-Neef and Pfeifer, 2004; Torarinsson *et al.*, 2005). Because of these difficulties, standard gene prediction tools had to be supplemented with methods that exploit features of halophilic proteins, for example their acidic isoelectric point profiles, and with experimental data such as the proteomic identification of coding regions (Falb, 2005). Fortunately, such processed data are typically already available when a microbial genome is published. In this study, we used the expert-validated gene sets for *Halobacterium salinarum* strain R1 (Figure 3.5) and *Natronomonas*

pharaonis available in the Halolex database (Pfeiffer *et al.*, 2008b).
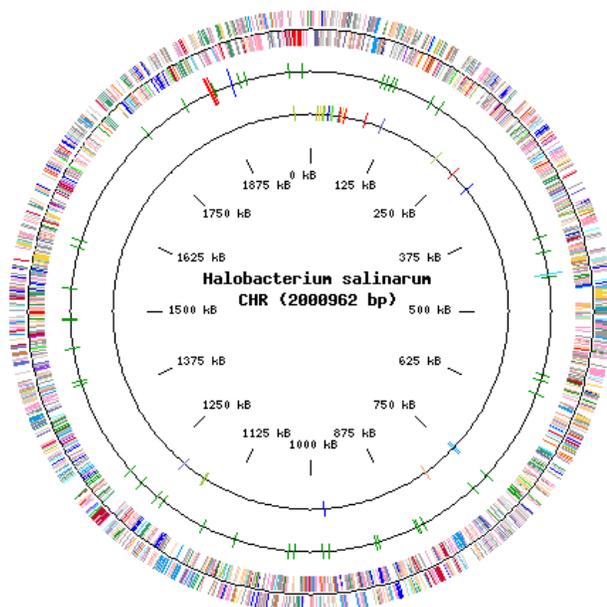
Figure 3.5: A representation of the genome of *Halobacterium salinarum*. Proteins are marked in the outer circle in a strand-specific manner. The central circle shows stable RNAs. Insertion elements are indicated in the innermost circle. The image was taken from the Halolex website (Pfeiffer *et al.*, 2008b).

Gene Function Assignment - Homology

When the hot-spots in the genome have been adequately identified, the next step in the annotation is to obtain a definitive catalogue of the proteins of the organism, where a putative function is assigned to each element whenever possible (Figure 3.4, transition B). To be sure, a rigorous experimental characterization remains the authoritative basis for assigning gene function. However, the extremely rapid pace at which sequences are generated precludes the possibility of performing experiments in each case. In this context, computational methods that require only the sequence for predicting function have been developed, and these, to a degree, have been widely successful.

Two genes are said to be homologous if they share a common ancestry, i.e., they come from the same gene of a common ancestor species. Furthermore, both are *orthologous* to each other if they are from distinct species, and *paralogous* otherwise. This concept of homology carries powerful implications in that it, with certain assumptions, often allows function to be "transferred" from one gene (protein) to another (Bork *et al.*, 1998). For example, a protein that is homologous to a permease very likely also serves a transport-related function. In bioinformatics, homology among proteins and DNA is often concluded on the basis of sequence similarity (Altschul *et al.*, 1990, 1997); if two or more genes have a score that is above a certain threshold using some sequence similarity system, then they

are likely homologous. While not perfect, for instance because sequence similarity can arise from other reasons, such as by chance in short sequences or in different proteins being selected for on the basis of having to bind to a particular structure, the system for determining homology by sequence similarity has been extensively used. Indeed, up to now it still serves as the primary source for annotating genes of newly sequenced organisms with function.

Perhaps the most popular set of tools for comparing sequences at the nucleotide or protein level is the BLAST family of tools (Altschul *et al.*, 1990, 1997; Zhang *et al.*, 2000; Zhang and Madden, 1997; Madden *et al.*, 1996). Typically, new query proteins, such as those obtained from newly sequenced genomes, are each "blasted" against several different databases. At this step, it is important to bear in mind that the accuracy of the results is, to a large extent, determined by the quality of the reference proteins used. For one thing, the nature of transfering functions based on sequence similarity itself dictates that incorrect annotations can also be easily transferred. In this respect, SWISS-PROT and SWISS-PROT TrEMBL (Boeckmann *et al.*, 2003; O'Donovan *et al.*, 2002) have emerged as among the most valuable sequence repositories (Stein, 2001). SWISS-PROT is a curated collection of confirmed protein sequences that have been extensively annotated and cross-referenced with other bioinformatic databases. Annotations include bibliographic references, functional descriptions, biological roles, protein family assignments and, when available, links to structural data.

Supplementing Homology Searches - Gene Context Analysis

Even with the level of success achieved by homology-based methods, it was recognized that further systems that could supplement them had to be developed. Problems associated with relying exclusively on homology include the fact that the stringent requirement of finding for a given query protein a match with known function cannot always be satisfied, and the well-known reality that two proteins with the same function need not necessarily have similar sequences. It was against this backdrop that methods which exploit gene context have emerged as successful complements.

True to their namesake, context-based approaches use contextual information such as gene fusion, the conservation of local gene neighborhoods, and the co-occurrence of genes across genomes. In the first case, proteins encoded by genes with homologs which are fused in another organism tend to be functionaly related (Enright *et al.*, 1999; Snel *et al.*, 2000; Marcotte *et al.*, 1999). Likewise, in the second case, genes which are significantly

encountered as neighbors across genomes, detected by the conservation of either gene order (Dandekar *et al.*, 1998) or genes in a run (Overbeek *et al.*, 1999), also tend to be functionaly related. Finally, in the third case, functionally linked proteins are assumed to be inclined to evolve in a correlated fashion, and as such can be found by comparing their phylogenetic profiles (Pelligrini *et al.*, 1999). While these methods typically do not predict specific functions by themselves, they are used to predict higher level functions, such as the participation of a protein in a particular structural complex or metabolic pathway.

In the case of *Halobacterium salinarum*, protein function assignments based on various homology methods and other approaches are already available in the Halolex database (Pfeiffer *et al.*, 2008b). Accordingly, in this study we only had to use protein function assignment methods in specific cases, such as in instances where new relevant sequences are created or characterized.

3.3.2 Reconstructing the Reaction Network

Protein (gene) function is a central concept that can cause much confusion. For example, an annotation that reads "growth related" tells you about one biological process in which a protein is involved in, but not much else. Because of this, it is often convenient to differentiate between context-independent molecular functions and context-dependent roles. For example, a protein may be a "transporter" that is "involved in drug resistance" because it is used to remove toxic materials inside cells. In addition, several nomenclature and classification schemes that can help organize and standardize annotations have already been developed. These include EC (Enzyme Commission) numbers for enzymes (Fleischmann *et al.*, 2004), TC numbers for transporters (Saier, 2000) and gene ontologies (Harris *et al.*, 2004). Given that proteins with enzymatic function are the most relevant ones in metabolic reconstructions (Figure 3.4, transition C), we devote some time to describe the EC numbering system.

The Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) introduced a classification and nomenclature system of assigning what are called EC numbers. This system is function centric, i.e., identifiers are assigned on the basis of the reaction catalyzed, rather than on structure (sequence). For example, two enzymes receive the same EC number even if they are from different organisms as long as they catalize the same reaction. An EC "number", similar to an IP address, is actually formed by four numbers separated by dots. Each successive number corresponds to a classification that is more specific than the one before it. The first number represents the

Table 3.1: Top-level EC Numbers.

Group	Description	Typical Reaction
EC 1.-.-.-	*Oxidoreductases*; transfer of electrons from one molecule to another.	$AH_2 + B \leftrightarrow A + BH_2$
EC 2.-.-.-	*Transferases*; transfer of a functional group (methyl, acyl, amino or phosphate) from one substance to another.	$AX + B \leftrightarrow A + BX$
EC 3.-.-.-	*Hydrolases*; hydrolysis of a chemical bond.	$AB + H_2O \leftrightarrow AH + BOH$
EC 4.-.-.-	*Lyases*; nonhydrolytic breakdown of various chemical bonds, often forming a new double bond or a new ring structure.	$A{=}B + X\text{-}Y \leftrightarrow \overset{X\ \ Y}{\underset{}{A\text{-}B}}$
EC 5.-.-.-	*Isomerases*; intramolecule rearrangement.	$A \leftrightarrow B$
EC 6.-.-.-	*Ligases*; linking together of two molecules, such as the sugar phosphate backbones of DNA.	$A + B + NTP \leftrightarrow A\text{-}B + NDP + P$

Adapted from Fleischmann *et al.*, 2004.

type of reaction that is catalyzed. For example, oxidoreductases, enzymes that catalyze the transfer of electrons from one molecule to another, are assigned EC numbers that begin with "1". The six current top-level classifications are listed in Table 3.1. Further details on the classification system can be found in Boyce and Tipton, 2000 and in Tipton and Boyce, 2000. While not a perfect system, for example because it cannot always capture the small differences between the enzymes of different organisms that catalyze similar but not identical reactions, particularly with respect to differing substrate and coenzyme specificities, EC numbers can significantly aid the process of moving from a genome annotation to a metabolic network by providing a convenient means of linking relevant databases.

Other useful resources for functional information on enzymes include the BRENDA (Schomburg *et al.*, 2004) and ENZYME (Bairoch, 2000) databases. BRENDA is a comprehensive collection of enzyme and metabolic information based on primary literature. It is classified according to the EC system, which makes it a convenient reference for reconstructions that also use the system. In contrast to other databases, the content of BRENDA is not limited to a specific aspect of enzymes nor is it specific to an organism, although it covers organism-specific information. The database covers data on functional and molecular properties, enzyme names, catalysed reaction, occurrence, sequence, kinetics, substrates/products, inhibitors, cofactors, activators, structure and stability. Finally, perhaps one of the most important features of BRENDA is that its content is continuously updated by manual extraction from publications searched in the literature. Expert curation should help ensure that the database is of high quality. Both the BRENDA and ENZYME databases contain lists of synonyms for enzymes, which can aid the process of

connecting genes annotated with enyzamtic functions to reaction databases.

With the list of the relevant genes and their functional annotations in hand (Section 3.3.1), the next step in the metabolic reconstruction process is to compile the set of putative reactions that occur within the organism (Figure 3.4, transition D). This is usually done by searching the list of annotated genes and proteins against pathway and reaction databases using protein names, EC numbers or any other identifiers. It is at this step where the earlier mentioned enzyme synonym lists can be extremely helpful, particularly if one begins the process of metabolic reconstruction with a pre-existing genome annotation that does not use the EC system. There are several excellent resources that contain comprehensive reaction information that can be conveniently coupled to genome data, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2002) and MetaCyc (Caspi *et al.*, 2006). In the case of *Halobacterium salinarum*, a partial metabolic reconstruction that is based on the KEGG structure already exists (Falb, 2005). Because we used this previous work as the starting point of our reconstruction effort, the resulting network of this study also primarily relies on the conventions of the KEGG framework.

The Kyoto Encyclopedia of Genes and Genomes is actually a collection of databases that are designed to be resources for understanding higher order functional meanings and utilities of organisms from their genomes. These are the GENE database, which contains gene catalogs of completely sequenced genomes as well as some partial ones, the SSDB database, which defines homology relations of protein coding genes in complete genomes, the PATHWAY database, which contains generalized protein interaction networks involving various cellular processes, the LIGAND database, which contains information on chemical compounds and reactions, and several others. In this study, we used the LIGAND database as the basis of reaction definitions, and the PATHWAY database for visualization.

While the primary annotation effort for a genome is focused on the identification and characterization of important regions, such as those associated with protein-coding genes, rRNAs, and tRNAs, the process of network reconstruction, on the other hand, is concerned with the description of the interactions between these components, their products and with other biological elements. In the case of metabolism, the relevant genes are primarily those that code for enzymes, as these determine the interaction links between metabolites (wiring diagram) by defining what reactions are (putatively) possible. Accordingly, a network reconstruction provides an additional layer of information over the primary annotation. In this respect, a reconstructed network can be thought of as a two-dimensional annotation of a genome (Figure 3.6) (Palsson, 2004).
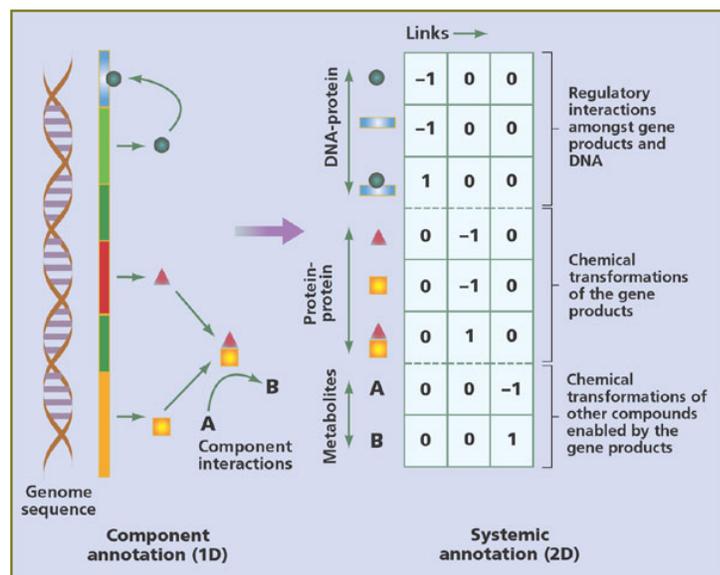
Figure 3.6: 2D Genome Annotation. Genome annotations typically serve as the primary resource for the reconstruction of cellular, including metabolic, networks. In the figure, each state of a biological component is a row in the matrix, and each column represents a chemical transformation among the components. The entries in the matrix are the stoichiometric coefficients of that transformation, or a link in a reaction network. This work focuses on the metabolic components of the network. The image was taken from Palsson, 2004.

3.3.3   The Role of Manual Curation and Literature Search

The procedure described in Figure 3.4 results in a draft metabolic network. Unfortunately, the series of mostly (semi-)automatic steps produces a reconstruction that is neither complete nor accurate (Iliopoulos *et al.*, 2001; Devos and Valencia, 2001, 2000). Manual curation, including a comparison of the network with knowledge from literature, is necessary to correct the errors and to fill in gaps. Some typical sources of errors are described below:

- *Incorrect or insufficient entries in the primary annotation* - Genome annotations are subject to the limitations of the protein function prediction methods that were used in their creation. For example, OE2647F (VNG1148G in the NRC-1 strain) was originally annotated as an "amino acid kinase" ("ornithine carbamoyltransferase"). However, the fact that it is a conserved neighbor of mevalonate kinase (EC 2.7.1.36) suggested that it is somehow involved in the production of isoprenoids. Indeed, the ortholog in *Methanocaldococcus jannaschii* was found to phosphorylate isopentenyl phosphate.

- *Newly characterized reactions (pathways)* - It generally takes some time for databases such as KEGG to include newly characterized reactions (pathways). This is especially true if the reactions are limited to only a few organisms. For example, the classical biosynthetic pathway for isopentenyl pyrophosphate (IPP) through mevalonate proceeds as follows: hydroxymethyl-glutaryl-CoA (HMG-CoA) is produced from acetoacetyl-CoA and acetyl-CoA (EC 2.3.3.10). HMG-CoA is reduced to mevalonate

(EC 1.1.1.34), which is phosphorylated to mevalonate phosphate (EC 2.7.1.36), which is converted to mevalonate diphosphate after a second phosphorylation step (EC 2.7.4.2), which is then finally decarboxylated to IPP (EC 4.1.1.33) (Figure 3.7). However, in *Methanocaldococcus jannaschii*, it was demonstrated that the second phosphorylation step is actually preceded by the decarboxylation step (Grochowski *et al.*, 2006), rather than the other way around. Genetic evidence suggests that *Halobacterium salinarum* likely uses this modified pathway (Falb *et al.*, 2008), which at the time of writing is still not in KEGG. Accordingly, the "noncanonical" reactions had to be manually defined. Another example is the alternative pathway for aromatic amino acid synthesis described in Section 3.7.3.

- *Inconsistencies and ambiguities between protein and function identifiers in different databases* - Even if two databases already share at least one naming convention, problems relating to the nature of the shared convention will still have to be dealt with. In the case of the EC system, several pathway repositories assign partially qualified EC numbers (partial EC numbers) to some reactions. For example, KEGG assigns the partial id of EC 3.1.3.- to the reaction (KEGG id R07280) in which 5-amino-6-(5'-phosphoribitylamino)uracil is hydrolyzed and a phosphate group is removed. Accordingly, one may erroneously assign to this reaction any gene that satisfies the first three components of the EC number, as has been observed in several databases including KEGG (Green and Karp, 2005). The problem is that EC 3.1.3.-, by definition, does not denote a specific reaction. Rather, it only specifies that the relevant enzyme is from a particular group, which in this case is the set of phosphoric monoester hydrolases. Situations like this need to be handled on a case by case basis. With respect to R07280, the two reactions before and after it in the linear segment of the pathway (riboflavin biosynthesis) in which it is involved have adequate genetic evidence in *Halobacterium salinarum*. Accordingly, it is likely that the reaction is also present in the organism. Certainly, the situation only gets worse if enzyme-gene synonym lists have to be used because the pertinent databases do not share naming conventions.

- *Organism-specific cofactor requirements* - Cofactor requirements are often organism-specific and have to be found elsewhere. The BRENDA database is a very convenient resource for this task.

- *Unspecified reaction reversibilities* - Reversibility information is not explicitly included in the KEGG databases. At best, KEGG uses single or bidirectional arrows in

its reference pathway images to indicate that a reaction is reversible or "irreversible", respectively. These visual indicators can sometimes be inconsistent between pathways (images). For these reasons, in this work we had to obtain reversibility information from other sources. In particular, we used organism-specific data from literature as the primary basis. If this was not available, literature data from closely related organisms was used. In the absence of both, we used the assignments made in other reconstructions as well as the annotation by Ma and Zeng (Ma and Zeng, 2003), where the reversibility of each KEGG-defined reaction was assigned based on biochemical principles.
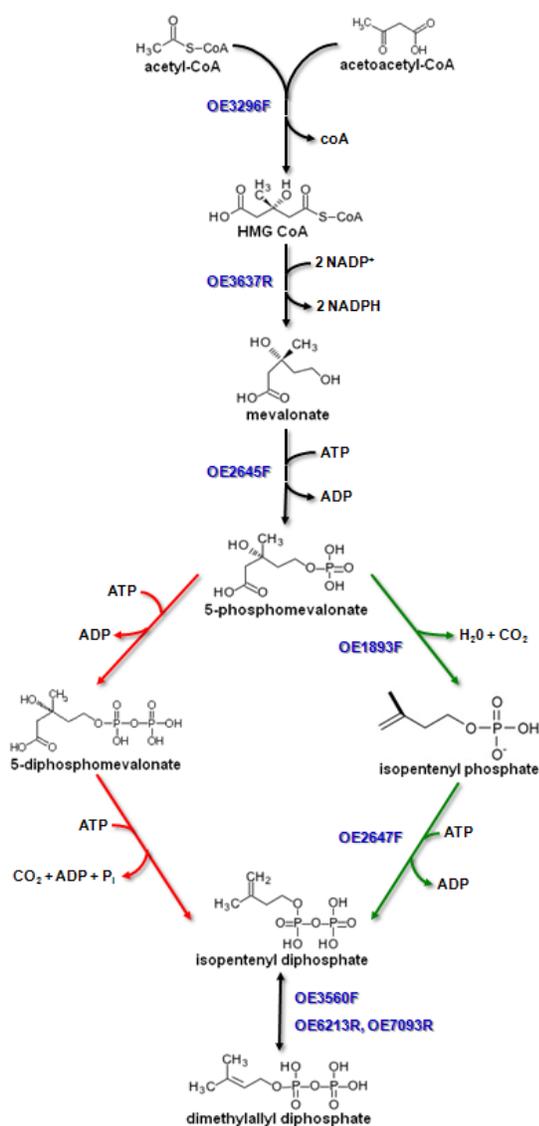
Figure 3.7: 2D Modified mevalonate pathway. One reason why manual curation is crucial to the reconstruction process is that it generally takes some time for databases to include newly characterized reactions (pathways). This is especially true if the reactions are limited to only a few organisms. For example, isopentenyl pyrophosphate (IPP) is the precursor of numerous important cellular constituents, including lipids and retinal. The classical biosynthetic pathway for IPP through mevalonate begins with the production of hydroxymethyl-glutaryl-CoA (HMG-CoA) from acetoacetyl-CoA and acetyl-CoA (EC 2.3.3.10). HMG-CoA is reduced to mevalonate (EC 1.1.1.34), which is phosphorylated to mevalonate phosphate (EC 2.7.1.36), which is converted to mevalonate diphosphate after a second phosphorylation step (EC 2.7.4.2), which is then finally decarboxylated to IPP (EC 4.1.1.33). However, in *Methanocaldococcus jannaschii* it has been demonstrated that the second phosphorylation step is preceded by the decarboxylation step (Grochowski *et al.*, 2006). Genetic evidence suggests that *Halobacterium salinarum* likely uses this modified pathway, which is currently not in KEGG. The red arrows correspond to the reactions of the classical pathway which are not in the modified pathway. The modified steps are indicated in green. Putative enzyme-genes from *Halobacterium salinarum* R1 for each step of the pathway are provided in blue.

3.3.4   Filling the Gaps

Given that genome annotation is never really "completed" but rather evolves as new information is added, the same can also be said of metabolic reconstruction. Nevertheless, some general properties that have to be fulfilled before releasing the first "definite" version of the reconstruction can be defined. First, the metabolic capabilities represented by the reconstructed network should be consistent with the physiology of the organism. For example, if it is known that a particular amino acid, say phenylalanine, is not essential for the microorganism, then a biosynthetic pathway for phenylalanine must be present. Furthermore, if the reconstruction is to be used as the basis of a genome-scale metabolic model with the purpose of yield and flux predictions, such as in this study, the essential pathways should be complete, and the rest of the network be left with as little unresolved gaps as possible.

In this study, we define a pathway gap to mean a specific reaction for which no direct support, genetic or bibliomic, could be found but nevertheless can be present in the organism because the activity of the pathway to which it belongs to has been adequately established. The likelihood of such a reaction being present in the organism is particularly high if the existence of other nearby reactions, such as those before and after it in the case of linear pathways, can be justified. One example of a pathway with gaps in *Halobacterium salinarum* is the biosynthesis of cobalamin.

Cobalamin (coenzyme B12) is a cofactor that is produced by some bacteria as well as some archaea from uroporphyrinogen III (uroIII). Two distinct pathways for the conversion of uroIII to the cobalamin intermediate adenosylcobinamide have been identified from different microorganisms; an anaerobic pathway in *Salmonella typhimurium* (cbi/cob genes) and an aerobic pathway in *Paracoccus denitrificans* (cob genes) (Roth *et al.*, 1996). The major difference between the two pathways, other than the differing conditions under which they are induced, is in the time of cobalt insertion. In the anaerobic pathway, cobalt is inserted early in the reaction sequence using sirohydrochlorin cobaltochelatase (CbiX, EC 4.99.1.3), while in the aerobic pathway, cobalt is inserted relatively later using a different cobaltochelatase (CobN, EC 6.6.1.2). In the case of *Halobacterium salinarum*, uroIII is also likely the precursor of cobalamin (Falb, 2005, p.122-123). Moreover, the gene set of the organism resembles both sets from the two model organisms, and has homologs of the two types of cobalt chelatases, EC 4.99.1.3 (CbiX, OE3221F) and EC 6.6.1.2 (CobN, OE3230F). This is consistent with the fact that *Halobacterium salinarum* can grow both aerobically and anaerobically.

Figure 3.8: Filling the gaps - Coenzyme B12 biosynthesis. The original image was created using KEGG by coloring the "Porphyrin and Chlorophyll Metabolism" reference pathway according to *Halobacterium salinarum* R1 genes. Each box in the figure corresponds to a reaction, where the label is either an EC number or the name of the protein(s) associated with it. Boxes that correspond to an enzyme for which the genome of *Halobacterium salinarum* putatively codes for are filled in green. For obvious reasons, automatically reconstructing the biosynthetic pathway for cobamide coenzyme (Vitamin B12) will leave some "gaps" (indicated in red for the anaerobic pathway). These as well as those from other pathways had to be closed manually before we used the network as the basis for further computational analyses.

Figure 3.8 shows the "Porphyrin and Chlorophyll Metabolism" reference pathway from KEGG, which includes both pathways for cobalamin biosynthesis, colored according to the *Halobacterium salinarum* R1 gene set. Each box in the figure corresponds to a reaction, where the label is either an EC number or the name of the protein(s) associated with it. Boxes that correspond to an enzyme for which the genome of *Halobacterium salinarum* putatively codes for are filled in green. Clearly, an automated reconstruction of the biosynthetic pathway for cobalamin will leave some "gaps" (indicated in red for the anaerobic pathway). These gaps are due to reactions for which the associated enzyme-gene could not be assigned in the organism. Nevertheless, this does not mean that the gaps are really reactions that do not occur in *Halobacterium salinarum*. The fact that most of the other reactions in the pathway have ample genetic support is a strong indication that the pathway is active in the organism. Moreover, that genes for the reactions adjacent to these gaps could be assigned suggest that the gap reactions are likely also present. It is well known that two genes with the same function need not have similar sequences, such as in the case of convergent evolution, which may be the reason the pertinent genes remain unassigned.
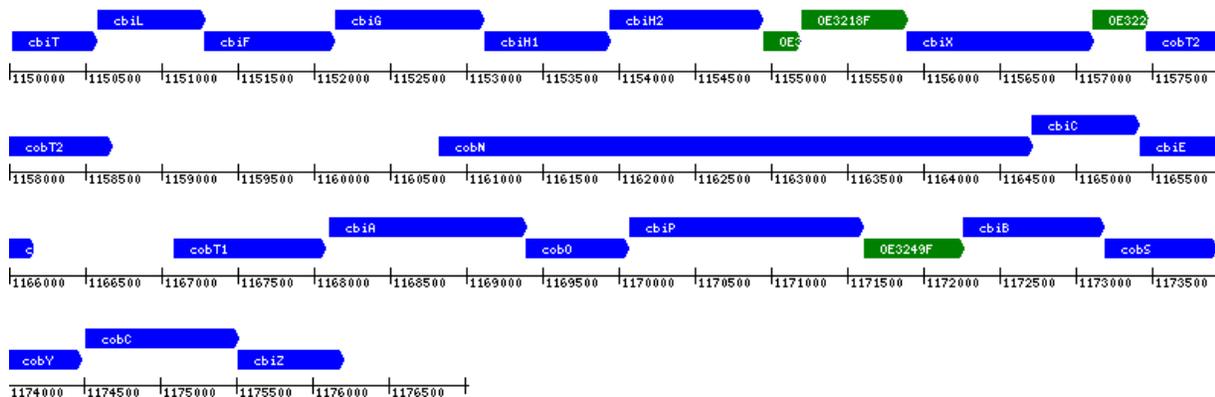
Figure 3.9: The cobalamin cluster of *Halobacterium salinarum*. The figure shows the genome region of *Halobacterium salinarum* where the genes associated with cobalamin synthesis are encoded (blue segments). The green segments denote genes for which no function has been assigned. Given the colocation of these unassigned genes with cobalamin biosynthesis elements and the preservation of this colocation in other microorganisms, it is very likely that these genes are also involved in cobalamin metabolism, and are good candidates for filling up the gaps in the pathway.

The systematic identification of pathway gaps, and network reconstruction in general, can also lead to the recognition of possible gene functional assignments, or details connected to them, that may have been missed by the primary annotation. In connection with the cobalamin biosynthesis example, four genes that are located within the cobal-

amin cluster of *Halobacterium salinarum* currently have unassigned function (Figure 3.9); OE3218F, OE3219F, OE3224F and OE3249F. However, their colocation with cobalamin biosynthesis genes is a good indication that they are also involved in the metabolism of the coenzyme. This possibility is further bolstered by the observation that the gene neighborhood is preserved in other species. For these reasons, the unassigned genes are good candidates for the (specific) gap reactions mentioned previously (Figure 3.8). Another example is nucleotide salvage; purine-nucleoside phosphorylase activity has been reported in the literature for *Halobacterium salinarum* (Stuer-Lauridsen and Nygaard, 1998). However, there is currently no gene that is assigned this function. It is possible that one of the two genes that are currently annotated in the primary annotation as "uridine phosphorylase" (EC 2.4.2.3; OE2311R and OE3603R) is actually a purine-nucleoside phosphorylase (Section 3.7.1).

3.4   Network Overview

The reconstructed metabolic network for *Halobacterium salinarum* includes 485 genes, 695 reactions and 548 distinct metabolites (Table 3.2). The network includes both putative reactions, i.e., those with bioinformatic evidence only, as well as experimentally demonstrated ones. The pathways in the network include central metabolism, biosynthetic and degradative pathways for amino acids, fatty acid degradation, an alternative pathway for shikimate biosynthesis, purine and pyrimidine metabolism, oxidative phosphorylation, and biosynthetic pathways for several vitamins and cofactors. 86% of the ORFs have reliable proteomic identification (Aivaliotis *et al.*, 2007; Klein *et al.*, 2005; Tebbe *et al.*, 2005). More than 99% of the R-1 (Pfeiffer *et al.*, 2008a) genes in the network are directly mapped to the *Halobacterium sp.* NRC-1 genome (Ng *et al.*, 2000).

A total of 571 reactions were taken directly from KEGG, although a significant number of them were modifed to correct errors or achieve balance. The rest of the reactions, 27 internal and 97 transport, were manually defined. The reactions in the network were subdivided into eight highlevel functional categories based on the major metabolic roles of the cell (Figure 3.10). Of these, "Amino acid metabolism" is currently the largest, which may be due to the fact that the organism relies primarily on amino acids for its nutrition. However, there are still a significant number of pathways associated with vitamins and cofactors that are still not completely understood so the majority composition of the amino-acid-metabolism-related reactions may change in the future. Long-chain fatty acid metabolism

Table 3.2: Comparison of genome and reconstruction with other organisms.

	Archaea		Bacteria		
	H. salinarum	*M. barkeri[a]*	*E. coli[b]*	*H. influenzae[c]*	*H. pylori[d]*
Genome statistics					
Length (bp)	2.7Mb	4.8Mb	4.6Mb	1.8Mb	1.7Mb
G-C content	68%	39%	51%	38%	39%
ORF	2867	5072	4288	1743	1590
With assigned function	1179		2656	1011	1091
No assigned function	1688		1632	732	499
Reconstructed metabolic network					
Reactions	695	619	931	461	381
Biochemical conversions	598	531	767	317	292
Transport	97	88	164	144	89
Reactions with assigned ORF (%)	558(80%)	509(82%)	873(94%)	412(89%)	272(71%)
Metabolites	550	558	625	367	332
ORFs (% of ORFs w/ assigned function)	485(41%)		904(34%)	400(40%)	290(27%)

[a]Taken from (Feist *et al.*, 2006).

[b]Taken from (Reed *et al.*, 2003).

[c]Taken from (Schilling and Palsson, 2000).

[d]Taken from (Schilling *et al.*, 2002).
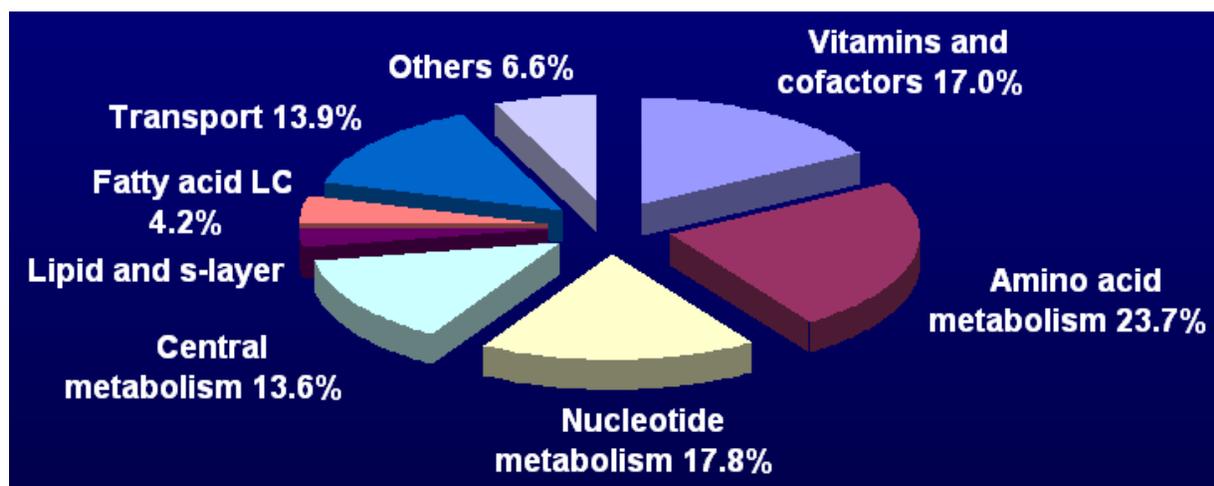
Figure 3.10: The distribution of reactions by category. The reconstructed metabolic network for *Halobacterium salinarum* includes 485 genes, 695 reactions and 548 distinct metabolites. The reactions in the network were subdivided into eight high-level functional categories, of which "Amino acid metabolism" is currently the largest. The "Lipid and s-layer" and "Vitamins and cofactors" categories are likely significantly underrepresented because several pathways belonging to these are yet to be characterized.

was separated from the category involving lipids and other outer coverings, with which they are typically associated, because archaeal lipids are derived from isoprenoids. Some specific pathways are discussed in later sections.

Most reactions in the network have either bioinformatic, specifically genetic (58%), or bibliomic support (8%), or both (22%) (Figure 3.11). However, given that the reconstruction also served as the basis for a genome-scale computational model, the network also includes a significant number of reactions with neither type of evidence. Transporters and metabolic gaps (Section 3.3.4), mostly in pathways for vitamins and cofactors, figure prominently in this respect (Figure 3.12). In the former, the difficulty is that transport proteins generally have sequences that are similar to each other even if they facilitate for different substrates. This means that it is often difficult to say anything more than that a protein is involved in transport. In the latter, a considerable number of gaps exist in vitamin and cofactor biosynthesis pathways because the archaea, including *Halobacterium salinarum*, commonly employ modified versions of the current model systems for these pathways, which are mostly from bacterial species. For example, genes could not be assigned for several steps in the biosynthetic pathway of coenzyme A, including pantothenate synthetase (EC 6.3.2.1) and its suceeding step, pantothenate kinase (EC 2.7.1.33). Nevertheless, these reactions were included as metabolic "orphans" because coenzyme A is a component of several biomass constituents.
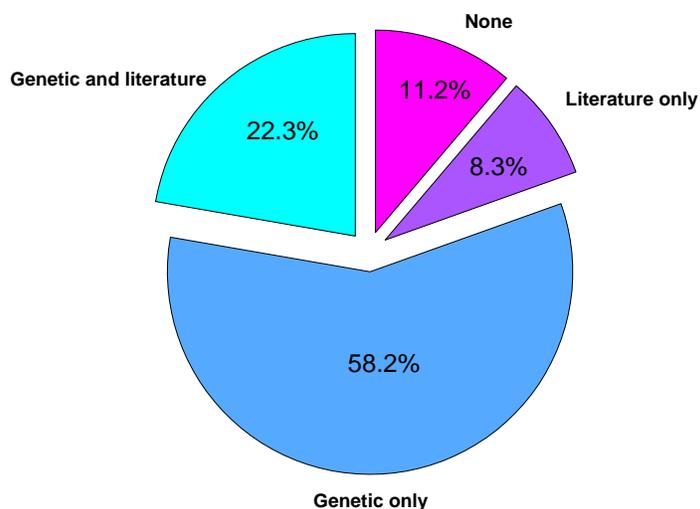
Figure 3.11: Network evidence. Most reactions in the network have bioinformatic or bibliomic support. However, a significant number of reactions with neither type of evidence are also included because they are necessary to complete the network in preparation for computational modeling. A total of 82 unique literature references are associated with the reactions.

Table 3.3: Reactions and enzymes in the network by KEGG defined pathways. A total of 571 reactions were taken directly from KEGG with at most minor modifications. The table lists the KEGG-defined reference pathways that include at least one reaction or enzyme that is in the reconstructed network. The number of enzymes (reactions) that are shared as well as the total in each pathway are indicated. Reactions can belong to several reference pathways. The rest of the reactions in the network, 27 internal and 97 transport, were manually added.

Pathway	Enz.[a]	Rxn.[b]	Pathway	Enz.	Rxn.
Glycolysis / Gluconeogenesis	19/40	13/47	N-Glycan biosynthesis	1/27	1/37
Citrate cycle (TCA cycle)	15/23	15/29	Nucleotide sugars metabolism	8/31	9/35
Pentose phosphate pathway	15/35	12/43	Streptomycin biosynthesis	5/14	2/19
Inositol metabolism	4/6	2/8	Polyketide sugar unit biosynth...	2/5	1/28
Pentose and glucuronate interc...	4/54	3/61	Aminosugars metabolism	5/39	4/45
Fructose and mannose metabolis...	8/62	5/68	Lipopolysaccharide biosynthesi...	2/20	0/21
Galactose metabolism	8/37	4/50	Peptidoglycan biosynthesis	1/17	0/27
Ascorbate and aldarate metabol...	2/32	0/36	Glycerolipid metabolism	6/38	6/42
Fatty acid biosynthesis (path ...	1/14	0/40	Glycerophospholipid metabolism	2/51	1/69
Fatty acid biosynthesis (path ...	3/7	18/25	Glycosphingolipid metabolism	1/28	0/39
Fatty acid metabolism	12/29	30/47	Ganglioside biosynthesis	1/11	0/25
Synthesis and degradation of k...	4/6	4/6	Pyruvate metabolism	23/66	26/70
Biosynthesis of steroids	12/36	18/88	1- and 2-Methylnaphthalene deg...	5/17	0/42
Bile acid biosynthesis	4/26	0/46	Tetrachloroethene degradation	1/5	0/11
Ubiquinone biosynthesis	8/15	5/34	Nitrobenzene degradation	2/18	0/31
Androgen and estrogen metaboli...	2/26	0/58	Glyoxylate and dicarboxylate m...	9/58	9/62
Oxidative phosphorylation	9/16	3/9	1,2-Dichloroethane degradation	1/4	0/4
Urea cycle and metabolism of a...	5/34	6/35	Benzoate degradation via CoA l...	8/38	2/45
Purine metabolism	37/100	64/145	Propanoate metabolism	17/48	13/51
Pyrimidine metabolism	29/62	61/109	Ethylbenzene degradation	1/8	0/14
Glutamate metabolism	22/36	22/36	Styrene degradation	3/18	3/20
Alanine and aspartate metaboli...	19/42	18/37	Butanoate metabolism	19/52	16/53
Tetracycline biosynthesis	1/3	1/5	C5-Branched dibasic acid metab...	3/21	2/22
Glycine, serine and threonine ...	28/59	27/65	One carbon pool by folate	10/24	12/30
Methionine metabolism	11/37	12/39	Methane metabolism	2/27	1/27
Cysteine metabolism	9/23	7/26	Carbon fixation	13/23	14/29
Valine, leucine and isoleucine...	22/33	32/43	Reductive carboxylate cycle (C...	11/13	12/14
Valine, leucine and isoleucine...	5/15	6/24	Thiamine metabolism	5/14	5/16
Lysine biosynthesis	9/29	7/37	Riboflavin metabolism	11/13	13/15
Lysine degradation	13/53	9/56	Vitamin B6 metabolism	2/24	0/38
Arginine and proline metabolis...	16/73	17/86	Nicotinate and nicotinamide me...	13/32	10/35
Histidine metabolism	17/40	15/38	Pantothenate and CoA biosynthe...	12/27	12/28
Tyrosine metabolism	13/70	7/106	Biotin metabolism	1/12	2/10
1,1,1-Trichloro-2,2-bis(4-chlo...	1/10	0/13	Folate biosynthesis	11/25	18/36
Phenylalanine metabolism	6/45	1/53	Retinol metabolism	1/10	1/11
gamma-Hexachlorocyclohexane de...	2/23	0/43	Porphyrin and chlorophyll meta...	26/64	29/89
Benzoate degradation via hydro...	3/48	0/60	Terpenoid biosynthesis	7/15	7/16
Tryptophan metabolism	11/60	5/80	Limonene and pinene degradatio...	5/22	2/59
Phenylalanine, tyrosine and tr...	22/31	22/35	Nitrogen metabolism	16/56	14/70
Novobiocin biosynthesis	6/6	2/35	Sulfur metabolism	12/30	9/30
beta-Alanine metabolism	9/32	6/34	Caprolactam degradation	3/21	0/22
Taurine and hypotaurine metabo...	2/17	1/18	Alkaloid biosynthesis I	3/38	1/46
Aminophosphonate metabolism	1/15	0/15	Alkaloid biosynthesis II	1/13	0/38
Selenoamino acid metabolism	13/22	0/24	Aminoacyl-tRNA biosynthesis	17/21	17/21
Cyanoamino acid metabolism	2/19	1/22	Biosynthesis of siderophore gr...	3/8	1/5
D-Glutamine and D-glutamate me...	1/12	2/11	Biosynthesis of vancomycin gro...	1/1	0/13
Glutathione metabolism	3/27	2/30	Biosynthesis of type II polyke...	2/3	0/14

[a]The number of enzymes in the reference pathway that are also in the reconstructed network over the total number of enzymes in the reference pathway.

[b]The number of reactions in the reference pathway that are also in the reconstructed network over the total number of reactions in the reference pathway.
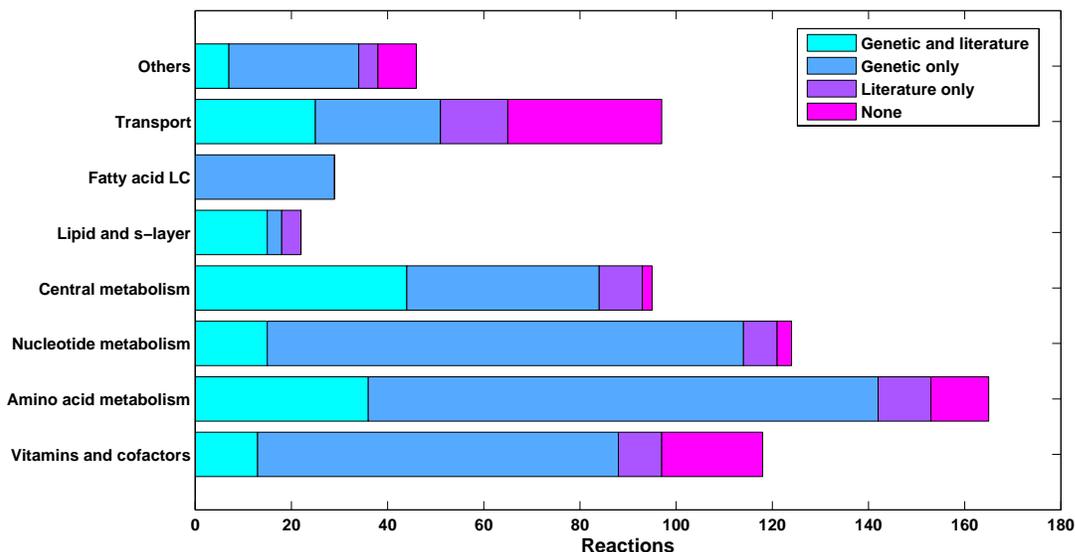
Figure 3.12: Network evidence by category. Most reactions with neither bioinformatic nor bibliomic support is either a transporter or belongs to a vitamin or cofactor biosynthesis pathway. In the former, the difficulty is that transport proteins generally have sequences that are similar to each other even if they facilitate for different substrates. Accordingly, it is often difficult to say anything more than that a protein is involved in transport. In the latter, numerous gaps exist in these biosynthetic pathways because the archaea, including *Halobacterium salinarum*, commonly employ modified versions of the existing reference systems (mostly from bacterial species).

3.5 Central Metabolism

In addition to the improvements that metabolic reconstruction can contribute to the primary annotation, the resulting network is also valuable in that it represents a literature-linked summary of the current knowledge regarding the organism's metabolism. Biological predictions can already be made from it, even prior to any explicit modeling effort. In this section, we discuss metabolic pathways of *Halobacterium salinarum* that are related to central metabolism. Other pathways are covered in later sections.

3.5.1 Glycolysis/Gluconeogenesis

Glycolysis is the sequence of reactions that converts glucose into pyruvate with the concomitant production of ATP. Although the term is often used synonymously with the Embden-Meyerhof (EM) pathway, alternative pathways, such as the Entner-Doudoroff (ED) pathway exist. Glycolysis is a major mode of energy production in many prokaryotes as well as some eukaryotes (e.g., mature erythrocytes, heavily-exercising muscle and fermenting yeast). It is one of the most well-known metabolic pathways.

Carbohydrate-utilizing haloarchaeal species, such as *Haloferax mediterranei* and *Haloarcula vallismortis*, have been reported to grow on fructose, glucose, sucrose and mannitol via a modified EM pathway (Altekar and Rangaswamy, 1992). It was observed that the enzymes of the pathway, i.e., ketohexokinase (EC 2.7.1.3), 1-phosphofructokinase (EC 2.7.1.56) and fructose 1,6-bisphosphate aldolase, were present under all growth conditions. Furthermore, these halophiles are also capable of glucose degradation via the semi-phosphorylated Entner-Douderoff (ED) pathway (Falb, 2005, pp.112). However, not all halophilic archaea are carbohydrate-utilizers. In *Halobacterium salinarum*, reconstruction of all three glycolytic routes, i.e., the EM, ED and pentose phosphate (PP) pathways, showed that it is unlikely that the organism catabolizes glucose (Falb, 2005, pp.112). With respect to the EM pathway, a gene for the key enzyme 6-phosphofructokinase (EC 2.7.1.11) could not be assigned, and neither for the new archaeal types of the enzyme that use ADP as co-substrate instead of ATP. With respect to the semiphosphorylated ED pathway, only a subset of its enzymes, glucose dehydrogenase and 2-keto-3-deoxygluconate kinase, appear to be present. The degradation of other sugars is also unlikely despite the presence two sugar kinase homologs (OE4535F, OE3606R). These findings are consistent with the fact that active transport of glucose has been experimentally excluded for *Halobacterium salinarum* (Severina *et al.*, 1991). The reconstruction of some central metabolic pathways are graphically illustrated in Figure 3.13.

While it is unlikely that glucose and other similar sugars are catabolized by *Halobacterium salinarum*, it has been reported that the exogenous addition of glucose and galactose can stimulate growth of the microorganism to some extent (Gochnauer and Kushner, 1969). One possible explanation for this is that supplying the carbohydrates exogenously removes the burden from the cells of having to produce the necessary amount of sugars for biomass formation; several cellular constituents contain sugar moities. When the BR over-producing strain *Halobacterium salinarum* S9 was fed with $^2H_7$-glucose, deuterium label was found in the sugar moieties of the major glycolipid S-TGA-1. Furthermore, saccharide units were also found attached to halobacterial surface proteins, such as the S-layer protein and flagellins (Sumper, 1987). If the carbohydrates are not supplied, they are likely synthesized via reversal of the EM pathway (gluconeogenesis). Labeling (Ghosh and Sonawat, 1998) and enzyme activity studies (Rawal *et al.*, 1988; D'Souza and Altekar, 1998) have shown a complete and functional reverse EM pathway in *Halobacterium salinarum*. The genes required for the triose part of the pathway could be assigned in the microorganism (Falb, 2005, pp.112).
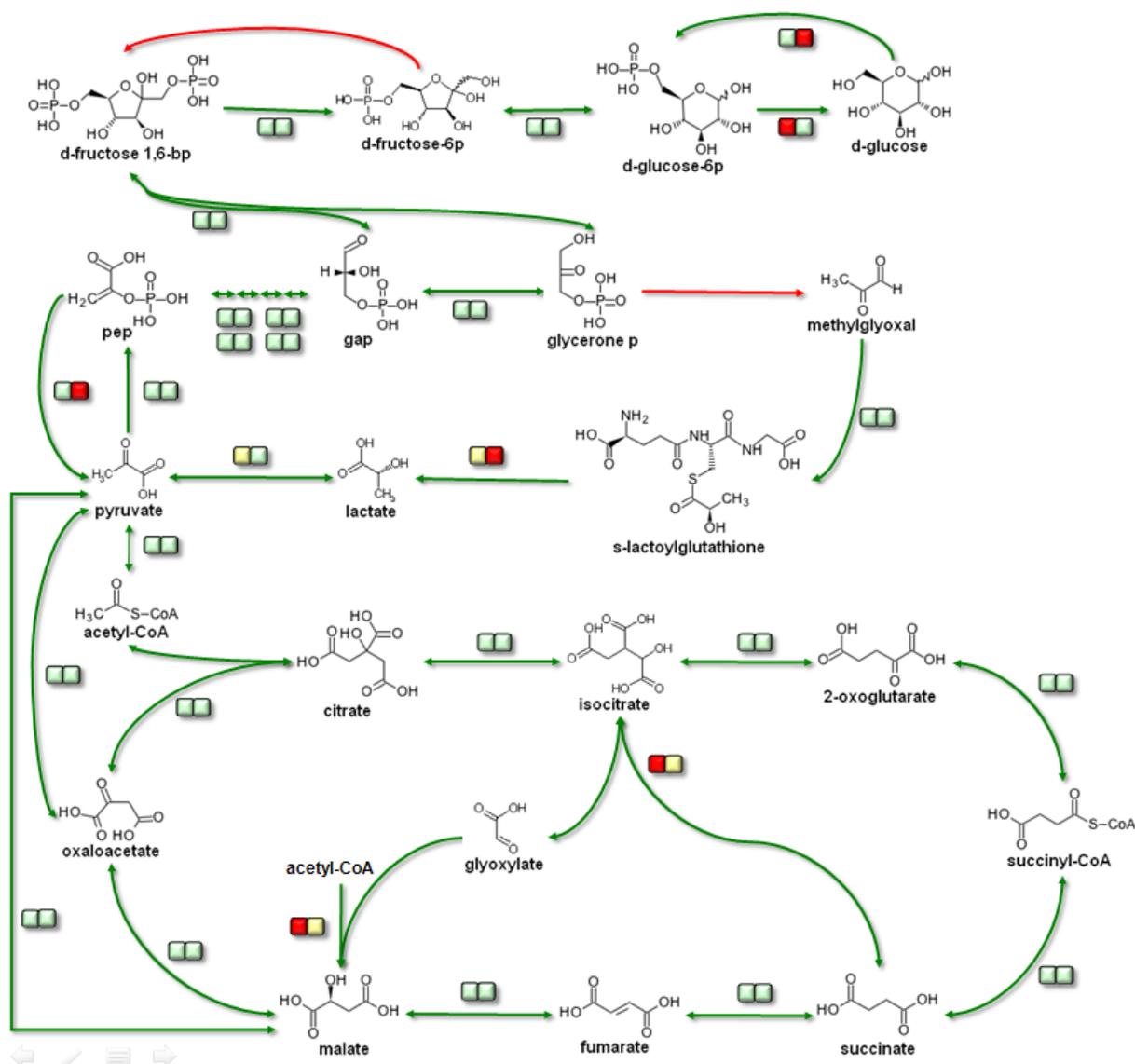
Figure 3.13: EM pathway, TCA cycle and some associated reactions. Green arrows represent biochemical transformations that are present in *Halobacterium salinarum*. Although some reactions likely do not occur in the halophile, they are nevertheless shown (red arrows) to complete the pathways in which they belong. A pair of adjacent boxes nearby each green arrow is used to signify the level of bioinformatic (left) and experimental (right) support associated with the corresponding reaction. Green denotes that the evidence is strong, red that there is none, and yellow indicates an intermediate level. Glycolysis via the EM pathway is not possible in *Halobacterium salinarum* as it lacks the key enzyme 6-phosphofructokinase (EC 2.7.1.11). However, activity of the reverse EM pathway (gluconeogenesis) has been experimentally demonstrated. A complete TCA cycle has been shown for the halophile using labeling studies.

3.5.2   Tricarboxylic Acid Cycle

The tricarboxylic acid cycle (TCA cycle), also known as Krebs' cycle, is a pathway (cycle) involved in the chemical conversion of carbohydrates, fats and proteins into carbon dioxide and water to generate a form of usable energy through respiration. The net effect of the cycle is the oxidation of two carbons to CO_2, with the concomitant production of energy (ATP or GTP) and reducing power (e.g., NADH). The TCA cycle was studied in *Halobacterium salinarum* using ^{13}C nuclear magnetic resonance spectroscopy (Ghosh and Sonawat, 1998), specifically in cells grown in the presence of glucose and alanine. Results were indicative of a complete cycle in the archaeon (Figure 3.13). Isotopomer analysis of glutamate labeling suggested that the flux through pyruvate:ferredoxin oxidoreductase (EC 1.2.7.1) accounted for 90% of the total pyruvate entering the TCA cycle, while pyruvate caboxylase (EC 6.4.1.1) accounted for 10%.

     The results of the labeling studies on the TCA cycle (Ghosh and Sonawat, 1998) also showed that: (1) The malic enzyme played a considerable role, accounting for 7% of the total label that entered the cycle. Approximately 42% of the label went through the flux from α-ketoglutarate to glutamate, and about 41% was recycled. (2) The inclusion of malonate, a competitive inhibitor of succinate dehydrogenase, reduced the cycling of the label in the cycle. However, buildup of the intermediates before the succinate dehydrogenase reaction (EC 1.3.99.1) was not observed. This suggested that either the oxaloacetate pool, which is necessary for condensation with acetyl-CoA in the citrate synthase reaction (EC 2.3.3.1), was depleted or the TCA cycle enzymes existed as a cluster or a metabolon, such that an inhibition of any one enzyme would inhibit the entire cycle. And (3) pyruvate mainly enters the TCA cycle under aerobiosis, while its transamination to alanine is enhanced under anaerobic conditions.

3.5.3   Bypass Pathways

Acetate enters the citric acid cycle as acetyl-CoA, where it is fully oxidized to CO_2. Thus, the cycle, in conjunction with respiration, allows the molecule to be used as an energy source. Indeed, fatty acids are degraded to acetyl-CoA molecules by beta oxidation, and are major sources of energy in some organisms. However, neither the TCA cycle nor its reversal are enough for the (reverse) process of carbohydrate biosynthesis from acetate. This capability is essential for organisms that can grow with acetate as the only carbon source. Rather, the capacity is afforded for by the glyoxylate cycle, which bypasses the

reactions of the TCA cycle where carbon is lost in the form of CO_2. Specifically, isocitrate undergoes cleavage into succinate and glyoxylate via the action of isocitrate lyase (EC 4.1.3.1). Glyoxylate then condenses with acetyl-CoA (EC 2.3.3.9) to yield malate, which can then be converted to phosphoenolpyruvate. The net effect of the cycle is the conversion of two two-carbon molecules (acetate) to a four-carbon compound.

In *Halobacterium salinarum*, activity of the two key glyoxylate cycle enzymes were detected after citrate in the growth medium was replaced by acetate (Aitken and Brown, 1969). However, the observed induction of isocitrate lyase by acetate could not be repeated in a later study (Oren and Gurevich, 1995b). Moreover, of the two strains that were investigated, namely *Halobacterium halobium* R1 and *Halobacterium salinarum* strain 5, isocitrate lyase activity could only be detected in the former. During reconstruction, we were unable to assign genes for both key enzymes in *Halobacterium salinarum* R1, and neither could we for *Halobacterium* sp. NRC-1. We should note that the cycle is not really as critical in *Halobacterium salinarum* as it is in other microorganisms because the archaeon already requires amino acids that enter central metabolism as three- or four-carbon molecules, and as such will not grow with acetate as the sole carbon source anyway. Nevertheless, the glyoxylate cycle enzymes were included in the network (Figure 3.13).

Another route through which D-lactate may be formed is through the methylglyoxal by-pass. This sequence of reactions has been thought of as a bypass for glycolysis, particularly of the EM pathway. In it, methylglyoxal is produced from dihydroxyacetone phosphate (glyerone phosphate) by methylglyoxal synthase (EC 4.2.3.3). Next, methylglyoxal joins with reduced glutathione to form S-lactoylglutathione, through the operation of glyoxalase I (EC 4.4.1.5). Finally, S-lactoylglutathione is hydrolyzed, via the action of glyoxalase II (EC 3.1.2.6), to glutathione and D-lactate. Note that some other -SH molecule could be the coreactant, and not glutathione. Although the role of the methylglyoxal bypass in cellular metabolism is not completely understood, one possibility is that it affords the the formation of acetyl-CoA from dihydroxyacetone phosphate under conditions where the phosphate concentration is low; the normal glycolytic route through glyceraldehde 3-phosphate dehydrogenase (EC 1.2.1.12, EC 1.2.1.13, EC 1.2.1.59) requires an additional orthophosphate molecule. With respect to the halophilic archaea, the presence of methylglyoxal synthase and glyoxalase I have been investigated in several species (Oren and Gurevich, 1995c). Remarkably, only the activity of the latter enzyme was detected in *Halobacterium salinarum*. Consistent with this, we were not able to assign a gene for methylglyoxal synthase during reconstruction (Figure 3.13). Accordingly, the archaeon may have an alternative pathway

that leads to methylglyoxal.

3.6 Catabolic Pathways

Catabolism refers to the set of metabolic pathways that break down molecules into smaller units in order to provide energy (e.g. ATP) and/or to generate metabolic intermediates for use in subsequent anabolic reactions. For example, polysaccharides, nucleic acids and proteins are broken down into monosaccharides, nucleotides and amino acids, respectively. The oxidative degradation of acetate to two CO_2 molecules by the TCA cycle, discussed earlier, is another example of catabolism. In this section, we outline the (putative) catabolic capabilities of *Halobacterium salinarum* with respect to several large molecules. In addition, an oxidative phosphorylation pathway is also proposed.

3.6.1 The Respiratory Chain

During the breakdown of metabolites, such as in the degradation of acetate via the TCA cycle, reducing power in the form of NADH or some other carrier can be produced. This reducing power can be used to produce further ATP through respiration, in addition to the ATP that is already generated by substrate level phosphorylation during the breakdown process itself. During aerobic respiration, electrons are transferred from the carriers (e.g., NADH) to molecular oxygen through a series of redox reactions mediated by an electron transport chain. At specific stages of this process, protons (H+) are pumped out by respiratory complexes, which results in the establishment of a proton motive force (pmf) across the membrane. In turn, the pmf is used by ATP synthase (EC 3.6.3.14) to produce ATP from ADP.

The capacity of *Halobacterium salinarum* to produce energy through respiration has been demonstrated (Sreeramulu *et al.*, 1998; Oesterhelt and Krippahl, 1973). Consistent with this, the halophile has analogs of all five oxidative phosporylation complexes that are found in mitochondria and *E. coli* (complexes I through V, Figure 3.14). However, significant differences from its better-studied counterparts can be discerned. For example, the subunits of complex I that comprise the NADH acceptor module (nuoEFG) could not be assigned. Indeed, it has been experimentally excluded that NADH is oxidized by a type I dehydrogenase in *H. salinarum*. Rather, NADH is oxidized by a non-homologous type II NADH dehydrogenase that also reduces quinones but is incapable of proton translocation (Sreeramulu *et al.*, 1998). Nevertheless, the conservation of eleven complex I subunits

Figure 3.14: Proposed Oxidative Phosphorylation Pathway. *H. salinarum* has analogs of all five respiratory complexes found in mitochondria and *E. coli* (complexes I to V). The boxes beneath each complex represent genes coding for specific subunits, which are often encoded adjacently in the genome (indicated by solid connections or by shared borders). Black broken lines are used to indicate that the connected genes, while not adjacent, are in the same genetic vicinity. The red broken line indicates that the *cbaD* and *hcpB* genes are fused in the archaeon. The proposed pathway has some notable differences from its more well-studied counterparts in *E. coli* or mitochondria. For example, genes coding for cytochrome c, which normally carries electrons to the terminal oxidase, could not be found in *H. salinarum*. Experimental evidence indicates that the function is likely performed by the copper protein halocyanin.

with high levels of sequence similarity make it likely that the complex is functional and translocates protons. Moreover, the lack of the NADH-specific acceptor module and the experimental evidences that NADH is not oxidized by complex I make it likely that the complex I analog actually accepts electrons from another donor molecule, which we left unspecified in Figure 3.14. In addition, the *H. salinarum* pathway is also different in that menaquinone, rather than ubiquinone, is the likely mobile carrier used to shuttle electrons from the complex I analog to the complex III analog. A similar proposal has been made for the closely related organism *Natronomonas pharaonis* (Scharf *et al.*, 1997; Falb *et al.*, 2005).

Two other important differences of the proposed pathway from the systems in mito-

chondria and _E. coli_ are the composition of the complex III analog and the mobile carrier which carries electrons to the terminal oxidase (complex IV analog). The cytochrome _c1_ subunit (_petC_) of _E. coli_ complex III, responsible for transfering electrons to the mobile carrier of the organism, cytochrome c, does not seem to have a homolog in _H. salinarum_. In fact, genes which code for cytochrome c could also not be assigned. Accordingly, we believe that _H. salinarum_ likely uses a different carrier. We propose this to be halocyanin, which is a blue copper protein originally isolated from _N. pharaonis_ (Scharf and Engelhard, 1993). A similar function has been proposed for the molecule in _N. pharaonis_, based on its localization as a membrane protein and its midpoint potential that is consistent with a mobile carrier's (Scharf _et al._, 1997). This proposed function is further supported by the fact that the halocyanin gene, _hcpB_, is fused with the _cbaD_ subunit of complex IV in _H. salinarum_ (Pfeiffer _et al._, 2008a).

Very little is known regarding the stoichiometry of the proton translocating processes in the respiratory chain of _H. salinarum_. Fortunately, while information on the individual components are unavailable, data on the aggregate process of respiration exists. This is very important for modeling because it determines the overall energy production capability of the organism. O_2 pulse experiments indicated an ATP to O_2 ratio of 1:1. Measurements of initial proton uptake during phosphorylation demonstrated a ratio of 10:1 between H^+ and ATP (Hartmann _et al._, 1977). These values are consistent with the experimentally determined photosynthetic stoichiometries of 22 photons per ATP and 2 photons per H^+, given that light inhibits respiration with an observed stoichiometry of 24 photons per O_2 molecule (Hartmann _et al._, 1977; Oesterhelt and Krippahl, 1973). We fixed the stoichiometry of the oxidative phosphorylation pathway in the model according to these values.

3.6.2 Glutamate (C5) Family of Amino Acids

Halobacterium salinarum preferentially utilizes amino acids. In line with this, a wide range of catabolic capabilities for amino acids can be deduced from its genome. With respect to glutamate, uptake of the amino acid has been described to require respiration or a minimum gradient in NaCl (Birkeland and Ratkje, 1985). The uptake process was reported not to be inhibited by a number of acids with similar chemical groups but was inhibited by the D entantiomer. Glutamate can enter central metabolism in one step, specifically through the TCA cycle as α-ketoglutarate, via the action of any one of a number of possible transamination reactions (e.g., EC 2.6.1.1, EC 2.6.1.19, etc...). Numerous such reactions exist because glutamate and α-ketoglutarate are linked to the transfer of amine groups in

several pathways. In addition, it can also be predicted from the gene equipment of *Halobac-terium salinarum* that the halophile degrades glutamate to mesaconate via enzymes of the β-methylaspartate pathway (glutamate fermentation), which are encoded within the *mam* gene cluster (OE4204F-OE4207F) (Falb, 2005, pp.119-120). Mesaconate might further be converted to citramalate and subsequently to pyruvate and acetate as in thermophilic anaerobic bacteria (Plugge *et al.*, 2001). For the reason that glutamine can be converted to glutamate through simple transamination reactions, the possibilities that exist for the latter also apply to it.
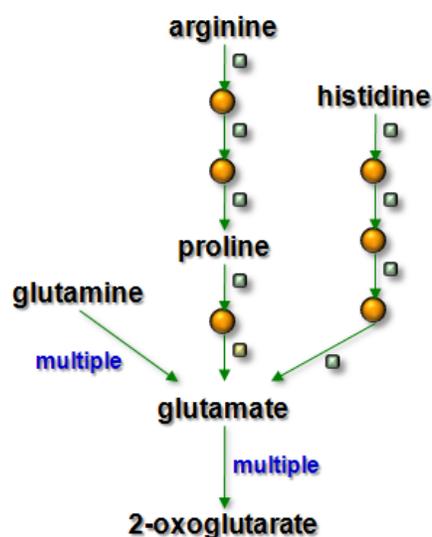
Figure 3.15: Catabolism of the glutamate (C5) family of amino acids. Glutamate can enter central metabolism in one step, specifically through the TCA cycle as 2-oxoglutarate, via the action of any one of a number of possible transamination reactions. In addition, the gene complement of *Halobacterium salinarum* also indicates that the archaeaon can catabolize glutamine, proline, arginine and histidine through conversion to glutamate. The arrows represent reactions and the orange circles intermediate metabolites. Near each reaction is a box that indicates the level of genetic evidence for the biochemical transformation; green denotes strong evidence and yellow that there are several candidate genes for the reaction but the substrate specificity of each is diffucult to determine.

The gene equipment of *Halobacterium salinarum* also indicates the capacity for the catabolism of proline, arginine and histidine, via conversion to glutamate. The first, proline, can be dehydrogenated to 1-pyrroline-5-carboxylate (EC 1.5.99.8; OE3955F), which can then be converted to glutamate after a second dehydrogenation step (EC 1.5.1.12; several possible ORFs). A specific ORF for the second enzyme could not be assigned because the substrate specificities of several dehydrogenases could not be determined using sequence similarity alone. The second, arginine, can be deiminated to citrulline (EC 3.5.3.6; OE5208R), which can be split into ornithine and carbamoyl phosphate (EC 2.1.3.3; OE5205R). The resulting ornithine can then be converted to proline after the removal of an amine group (EC 4.3.1.12; OE4121R, OE2945F). A fermentative pathway for arginine is discussed in the following chapter. Finally, the third, histidine, can be converted to urocanate (EC 4.3.1.3; OE2739F), which can be turned into 4-imidoazolone-5-propanoate (EC 4.2.1.49; OE2734F), which can be hydrolyzed to N-formimino-glutamate (EC 3.5.2.7; OE2738F), which can then be hydrolyzed to glutamate and formamide (EC

3.5.3.8; OE2736F). A summary of the catabolic capabilities for the glutamate family of amino acids is provided in Figure 3.15. In addition to glutamate, active transport has also been demonstrated for glutamaine, proline, arginine and histidine (MacDonald *et al.*, 1977).

3.6.3 Branched-chain Amino Acids

Reconstruction of the catabolic pathways for the branched-chain amino acids was particularly problematic because several of the steps involve enzymes for which genetic evidences are ambiguous. That is, several candidate genes are present for the general functions (e.g., "dehydrogenase"), but it is impossible to assign substrate specificities to them using the available methods. Nevertheless, it is very likely that complete and functional degradative pathways exist for leucine, isoleucine and valine, given the presence of the key enzymes. Indeed, as will be discussed later, numerous indications that leucine and isoleucine are preferred respiratory substrates of *Halobacterium salinarum* were found in the course of this study.

The Na^+-gradient-driven transport of L-leucine and D-leucine has been described in *Halobacterium salinarum*. It was suggested that D-leucine transport is coupled with two moles of Na^+, whereas L-leucine transport is coupled with one mole of Na^+ (Tanaka *et al.*, 2000). Leucine is a chemoattractant of the archaeon (Storch *et al.*, 1999). Degradation of the amino acid proceeds via its conversion to hydroxymethylglutaryl-CoA (hmg-CoA), which can then be split to acetyl-CoA and acetoacetate (EC 4.1.3.4). The former can readily enter the TCA cycle, and the latter can be converted to acetoacetyl-CoA via the action of acetoacetate succinyl-CoA transferase (EC 2.8.3.5; OE5189F or OE4211F). Acetoacetyl-CoA can then be converted to two molecules of acetyl-CoA through the operation of acetoacetyl-CoA thiolase (EC 2.3.1.9; OE3884F).

Similar to leucine, valine and isoleucine are actively transported by *Halobacterium salinarum* (MacDonald *et al.*, 1977), and both are also chemoattractants of the halophile (Storch *et al.*, 1999). The catabolism of isoleucine proceeds by its transformation to 2-methylacetoacetyl-CoA (several steps). With the participation of another CoA molecule, 2-methylacetoacetyl-CoA can be split into propanoyl-CoA and acetyl-CoA through the action of a thiolase (EC 2.3.1.16; OE3884F). Propanoyl-CoA can be carboxylated to (S)-methylmalonyl-CoA (EC 6.4.1.3; OE1939F, OE3175F), and then epimerized to (S)-methylmalonyl-CoA (EC 5.1.99.1; OE1718R). Entry into central metabolism (TCA cycle) can then be completed after transformation to succinyl-CoA (EC 5.4.99.2; OE1721F,

OE2005F, OE1972F). Valine also enters central metabolism through propanoyl-CoA as succinyl-CoA.

3.6.4 Glycine, Serine, Threonine and Alanine

Alanine enters central metabolism through pyruvate. One way through which this can be accomplished is through the action of alanine dehydrogenase (EC 1.4.1.1). The enzyme from *Halobacterium cutirubrum* has been partially purified, and was investigated with respect to enzymatic properties and thermal stability (Kim and Fitt, 1977). However, an ORF for the enzyme in the R1 genome remains to be assigned. Alternatively, alanine can also be converted to pyruvate through transamination reactions with different amine acceptors, such as oxaloacetate and aspartate (EC 2.6.1.1; OE1755F, OE1944R, OE2619F) (Bhaumik and Sonawat, 1994). The reverse reaction, pyruvate to alanine, was reported to occur primarily under anaerobic conditions.
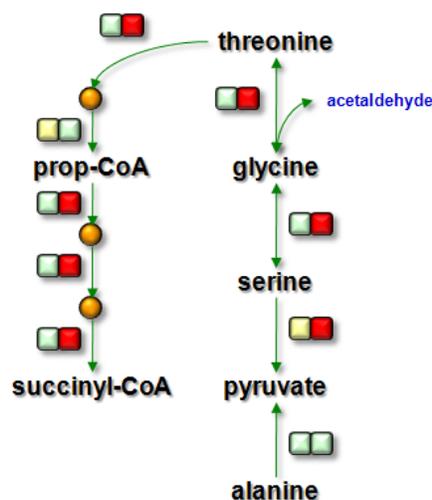
Figure 3.16: Catabolism of glycine, serine, threonine and alanine. Green arrows represent biochemical transformations that are present in *Halobacterium salinarum*. Nearby each reaction is a pair of adjacent boxes; these signify the level of bioinformatic (left) and experimental (right) support associated with the biochemical transformation. Green denotes that the evidence is strong, red that there is none, and yellow indicates an intermediate level.

Similar to alanine, serine enters central metabolism as pyruvate, through the operation of serine deaminase (also called threonine ammonia-lyase) (EC 4.3.1.19; OE3931R). Likewise, glycine also enters as pyruvate, via conversion to serine using the enzyme hydroxymethyltransferase (EC 2.1.2.1; OE3036F). Alternatively, glycine can also be cleaved (several reactions) to donate its methyl group, via tetrahydrofolate, to several biosynthetic pathways. Since the transformation of glycine to serine is reversible, then serine can also be catabolized through the conversion to and cleavage of glycine. Finally, threonine can be converted, through an intial deamination to 2-oxobutanoate (EC 4.3.1.19; OE3931R) and a succeeding reduction step, to propanoyl-CoA, and thus subsequently enter the TCA cycle

as succinyl-CoA. Although genes for the enzyme of the succeeding step, alpha-ketobutyrate-ferredoxin oxidoreductase (EC 1.2.7.2), still await assignment, its activity has already been characterized using enzymatic assays (Kerscher and Oesterhelt, 1977). Since threonine can be reversibly converted to glycine with the release of an acetaldehyde molecule (EC 4.1.2.5; OE4436R), then the possibilities for glycine and serine also exist for threonine, and vice versa, although the reverse direction will require the input of the acetaldehyde molecule. The catabolic possibilities for alanine, serine, glycine and threonine are summarized in Figure 3.16.

3.6.5　Aromatic Amino Acids

As with some eubacteria like *Escherichia coli*, *Halobacterium salinarium* has two types of transport systems for aromatic amino acids: one common system and one specific transporter for each of the three substrates (Lobyreva and Plakunov, 1987). However, the archaeon in addition has a second tyrosine-specific high affinity system, for a total of five transport systems associated with aromatic amino acids. When phenylalanine exceeds [^{14}C]tyrosine by four to sixfold, it inhibits competitively tyrosine transport of the common system. When phenylalanine is provided in 50- to 100-fold molar excess, it inhibits in a non-competitive way the specific high affinity system for tyrosine. In contrast, the specific low affinity system for tyrosine is practically insensitive to phenylalanine (Lobyreva *et al.*, 1994). One possible reason for having multiple transport systens is because an excess in the medium of one aromatic amino acid will result in the uptake inhibition of the other two in the common system. In addition, due to the existence of common regulatory sites in the biosynthetic pathways, an excess of one amino acid in the cells will result in the inhibition of biosynthesis of the other two. In turn, these possibilities can depress growth. The presence of two tyrosine-specific systems was thought to be involved in widening the range of optimal concentrations for the amino acid in the medium.

Tyrosine enters central metabolism as fumarate and acetoacetate (acetoacetyl-CoA) via homogentisate in some mammals, bacteria and fungi. Phenylalanine can share this catabolic route after a multistep conversion to 4-hydroxyphenylpyruvate. Alternatively, gram positive bacteria have been described to degrade tyrosine to pyruvate and succinate, via 2,5-dihydroxyphenylacetate (homoprotocatechuate) rather than homogentisate (Sparnins and Chapman, 1976). In *Halobacterium salinarum*, there is no clear genetic indication that either of these pathways are present. Nevertheless, several observations were made in this study that suggested that the archaeon has the relevant catabolic capabilities

(Chapter 4), possibly through some other means.

3.6.6   Fatty Acids

Although the membrane lipids of *Halobacterium salinarum* consist exclusively of derivatives of the dialkylglycerol diether, 2,3-diphytanyl-sn-glycerol (archaeol), small amounts of fatty acids have been detected in the halophile. Even-numbered chains, such as palmitate and stearate, are bound components of some of the archaeon's membrane proteins (Pugh and Kates, 1994). For example, halorhodopsin assembles into trimers around a central patch consisting of palmitic acid (Kolbe *et al.*, 2000). These findings suggested that a fatty acid synthase (FAS) system might be operative in *Halobacterium salinarum*. The hypothesis is further supported by the finding that whole cells of *Halobacterium cutirubrum* incorporates [^{14}C]acetate into fatty acids (Kates *et al.*, 1968), and by the isolation of a cell-free enzyme system that catalyses the biosynthesis of fatty acids from acetyl-CoA and malonyl-CoA (Pugh *et al.*, 1971). However, fatty acid metabolism is an enigmatic aspect because there is currently no genetic indication for any type of FAS system in the archaeon, and it was reported that only a fraction of the activity remains at the high salt concentrations (> 4 M) which are typically encountered in the cytoplasm of the organism (Pugh *et al.*, 1971).

In contrast to the biosynthetic pathway, all genes for the catabolism of fatty acids through the β-oxidation pathway are present in halobacterial genomes. This pathway consists of a recurring sequence of four steps that continues until the entire chain is cleaved into acetyl-CoA units: (1) oxidation by FAD or some other acceptor via the action of acyl-CoA dehydrogenase (EC 1.3.99.3; several candidate ORFs); (2) hydration of the bond between C-2 and C-3 (from the chain end) using enoyl-CoA hydratase (EC 4.2.1.17; OE3846R, OE1641R); (3) oxidation by NAD+, thus converting the hydroxyl group into a keto group (EC 1.1.1.35; OE2015R, OE2871F, OE3846R); and (4) cleavage of 3-ketoacyl CoA by the thiol group of another CoA molecule (EC 2.3.1.16; OE3884F), which results to acetyl-CoA and a residual chain that is shortened by two carbon atoms. Considering the presence of the requisite genes and the significance of fatty acids given their availability in halobacterial environments, it is somewhat surprising that there has been no report on their oxidation up until recently. Nevertheless, *Natronomonas pharaonis* was found to be able to grow with fatty acids of various lengths as the sole carbon source (especially C14) (Königsmaier, 2006). With respect to *Halobacterium salinarum*, only short-chain fatty acids are possible because medium- and long-chain molecules in high concentrations diminish its growth.

3.7 Biosynthetic Pathways

Most of the biosynthetic pathways for amino acids and cofactors in *Halobacterium salinarum* were covered by a previous reconstruction (Falb, 2005). Accordingly, in this study details are only provided for those that were significantly modified or updated.

3.7.1 Nucleotides, Amino Acids and Lipids

The gene complement of *Halobacterium salinarum* shows the complete set of requisite enzymes for the de novo synthesis of inosine monophosphate (IMP) from ribose 5-phosphate (rib5P) and uridine monophosphate (UMP) from carbamoylphosphate and rib5P. Furthermore, all genes for the subsequent synthesis of purines and pyrimidines required for DNA and RNA synthesis are also present in the genome. However, as is typical of the archaea, there are some significant differences compared to the existing model systems. For example, halophiles have a domain fusion pattern of purine synthesis enzymes that differs from all other organisms (Falb, 2005); genes required for steps 3 (EC 2.1.2.2) and 9 (EC 2.1.2.3) are fused (OE2292F) instead of genes for steps 9 and 10 (EC 3.5.4.10). Moreover, the IMP cyclohydrolase domain (EC 3.5.4.10) that can be found in most organisms is completely missing in halophiles and some other archaea (e.g. *Methanococcus jannaschii* and *Archaeoglobus fulgidus*). This domain is responsible for the conversion of 5-aminoimidazole-4-carboxamide ribonucleotide (ZMP or AICAR) to IMP. Instead, halophiles, including *Halobacterium salinarum*, and some methanogens possess a gene (OE4329F) that encodes an archaeal type of IMP cyclohydrolase which has no sequence similarity to the bacterial and eukaryotic enzymes (the preceeding information are reviewed in Falb, 2005, pp.116). The product of the non-orthologous gene was purified from *Methanococcus jannaschii* (Graupner *et al.*, 2002). Finally, purine-nucleoside phosphorylase activity has been reported in the literature for *Halobacterium salinarum* (Stuer-Lauridsen and Nygaard, 1998), although there is currently no gene that is assigned this function. Accordingly, it is possible that one of the two genes that are currently annotated as "uridine phosphorylase" (EC 2.4.2.3), namely OE2311R and OE3603R, is actually a purine-nucleoside phosphorylase.

As mentioned earlier, membrane lipids of archaea consist of glycerol diether lipids with prenyl side chains. The glycerophosphate (GP) backbone of these molecules is sn-glycerol-1-phosphate (G-1-P), which is the enantiomer of the bacterial and eukaryal counterpart (Nishihara *et al.*, 1999). So far, there has been no exception to this distribution of the GP backbone stereoconfiguration. Stereochemical inversion of exogenously supplied glyc-

Table 3.4: Summary of amino acid metabolism.

AA	Synthesis	Degradation	AA	Synthesis	Degradation
Alanine	yes	yes	Leucine	no	yes
Arginine	no[a]	yes	Lysine	no	yes[b]
Aspartate	yes	yes	Methionine	no	yes
Asparagine	yes	yes	Phenylalanine	yes	probable[c]
Cysteine	yes		Proline	yes	yes
Glutamate	yes	yes	Serine	yes	yes
Glutamine	yes	yes	Threonine	yes	yes
Glycine	yes	yes	Tryptophan	yes	no
Histidine	yes	yes	Tyrosine	yes	probable
Isoleucine	no	yes	Valine	no	yes

[a]Although long considered as essential, some preliminary results reported in this study suggest that arginine synthesis is possible.

[b]Key enzymes of the pathway could not be found but label from the amino acid turning up in isoprenoid molecules indicates degradation.

[c]Key enzymes could not be assigned but uptake rates greatly exceeded biomass incorporation rates.

erol phosphate has been shown to take place during lipid biosynthesis in *Halobacterium salinarum* (Kakinuma *et al.*, 1988; Nishihara *et al.*, 1999). Sn-glycerol-3-phosphate specific dehydrogenase (EC 1.1.99.5; OE3763F, OE3764F, OE3765F) and glycerol kinase (EC 2.7.1.30; OE3762R) for the "common" sterioisomer have been detected in cell-free homogenates of the archaeon (Wassef *et al.*, 1970).

In contrast to the versatility of *Halobacterium salinarum* with respect to amino acid catabolism, as indicated by its gene equipment, the halophile is somewhat limited when it comes to their biosynthesis. While other haloarchaea can grow using single carbon sources, such as *Natronomonas pharaonis* which can grow on acetate, several amino acids (10 or 15) are typically added to synthetic growth media used for *Halobacterium salinarum* (Oesterhelt and Krippahl, 1973; Grey and Fitt, 1976). It is somewhat surprising that up until recently, there has been no effort to systematically determine which of the supplemented amino acids are truly indispensable, and which ones only support growth. Nevertheless, the previous metabolic reconstruction for *Halobacterium salinarum* indicated that six are essential, namely arginine, lysine, methionine, leucine, isoleucine and valine (Falb, 2005, pp.120). This proposed set, except for lysine, fits well with the amino acids that are chemoattractants of the archaeon (Storch *et al.*, 1999).

Arginine has long been held as an essential amino acid for *Halobacterium salinarum*. However, due to modeling results that will be discussed in the succeeding chapter, we proceeded to verify whether ornithine can substitute for arginine in the medium. The

results of the experiments were positive. However, after doing the verification, we also went on to confirm that arginine is indeed essential by trying to grow *Halobacterium salinarum* with neither the amino acid nor ornithine supplied. The results this time were much more surprising because we were able to observe growth (Figure 3.17, left). Although the doubling time was much longer and the maximum cell density reached was just about a fourth of that when either arginine or ornithine is supplied, the density change was enough to preclude the possibility that the observed growth was only due to small amounts of either nutrient introduced during inoculation. Unfortunately, replicates failed to achieve similar results. Nevertheless, it is clear that arginine is critical for good growth in terms of speed and the maximum population size reached.

The possibility that arginine may not be an essential amino acid for *Halobacterium salinarum* prompted us to reevaluate the parts of the primary annotation and the reconstruction that are related to arginine synthesis. From the gene equipment of the halophile, it is clear that all the enzymes (genes) necessary for converting ornithine into arginine are present. Therefore, the only problem is getting to ornithine itself. We were able to find two possibilities in which this may be achieved (Figure 3.17, right): (1) *Halobacterium salinarum* has an ornithine cyclodeaminase (EC 4.3.1.12; OE4121R) that allows ornithine to be broken down into proline and NH_3. Although the reaction was reported in *Clostridium sporogenes* to either be irreversible or to have an equilibrium that is far in the direction of proline (Costilow and Laycock, 1971), a near zero availability of ornithine may be enough to drive the equilibrium in the other direction, considering that all reactions are essentially reversible. On the other hand, proline can already be derived from glutamate. (2) An alternative possibility is to derive ornithine directly from glutamate 5-semialdehyde, rather than through proline, using ornithine aminotransferase (EC 4.3.1.12). Candidate genes for this function exist in the genome, including OE3168R which is currently annotated as a "pyridoxal phosphate-dependent aminotransferase (homolog to acetylornithine aminotransferase)".

3.7.2   The Pentose Phosphate Pathway and Ribose Production

The Pentose Phosphate Pathway (PPP) commonly occurs in bacteria and eukaryotes. Some of its main functions are the production of pentoses, the synthesis of erythrose 4-phosphate, and the generation of NADPH. In the archaea however, the role of the PPP has been an open question. In fact, current information suggest that archaeal members fulfill the PPP functions in diverse ways. For example, pentoses are typically formed through

the oxidative branch of the PPP. However, some archaeal species are believed to utilize the non-oxidative branch (NOPPP) rather than the oxidative branch (Choquet *et al.*, 1994; Yu *et al.*, 1994; Dandekar *et al.*, 1999), and others that lack the NOPPP are believed to utilize a reverse ribulose monophophate pathway (RuMP) (Orita *et al.*, 2006; Kato *et al.*, 2006; Grochowski *et al.*, 2005). Both alternative routes are not present in *Halobacterium salinarum*.

The oxidative branch of the PPP (OPPP) is responsible for the conversion of glucose 6-phosphate to ribulose 5-phosphate. This reaction sequence proceeds via an initial dehydrogenation of the substrate to glucono-1,5-lactone 6-phosphate (EC 1.1.1.49), which is then followed by hydrolysis to gluconate 6-phosphate (EC 3.1.1.31). The pathway is completed with the oxidative decarboxylation of gluconate 6-phosphate to ribulose 5-phosphate (EC 1.1.1.44). No functional oxidative PPP has been described in the archaea yet (Falb, 2005, p.113). In *Halobacterium salinarum*, genes for the first two steps could not be assigned. However, activity of the initial step (EC 1.1.1.49) has been experimentally proven (Aitken and Brown, 1969). The enzyme for the final step seems to be present (EC 1.1.1.44; OE4581F), but it only shows N-terminal domain homology with known sequences. Accordingly, the OPPP remains a possibilty for pentose production in the archaeon (Figure 3.18).

Even if pentoses are not produced through the OPPP, the presence of the final enzyme in the sequence (EC 1.1.1.44) as well as that of ribose-5-phosphate isomerase (EC 5.3.1.6; OE4185F) suggest that the biosynthetic route at least goes through gluconate 6-phosphate. The latter enzyme isomerizes ribulose 5-phosphate to ribose 5-phosphate, which can already be converted to the ribonucleotide precursor phosphoribosyl diphosphate (PRPP) (EC 2.7.6.1; OE4085R). As an alternative to the OPPP, gluconate 6-phosphate can be produced through the direct phosphorylation of gluconate (EC 2.7.1.12) (Figure 3.18). In support of this possibility, labeled glucose has been shown to be metabolized to gluconate by *Halobacterium salinarum* cells (Sonawat *et al.*, 1990). If this is indeed the route for pentose synthesis, then it would explain the seeming presence of an incomplete semi-phosphorylated EntnerDoudoroff pathway in *Halobacterium salinarum*, particularly if the gene currently assigned as 2-keto-3-deoxygluconate kinase (EC 2.7.1.45) actually facilitates the missing step.

A structurally novel phosphopentomutase was recently characterized in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. Searching for a homologous gene in *Halobacterium salinarum* returned OE2318R, which was orginally annotated as "phosphohexomutase (phosphoglucomutase (EC 5.4.2.2), phosphomannomutase (EC 5.4.2.8))".

The phosphopentomutase activity (EC 5.4.2.7) together with 2-deoxyribose 5-phosphate aldolase (EC 4.1.2.4; OE3616F) allows the interconversion of deoxyriboses with acetaldehyde and glyceraldehyde 3-phosphate (Figure 3.18), and accordingly provides a link between central metabolism and deoxypentoses. This pathway presents another possibility for producing the ribonucleotide precursor PRPP, specifically if *Halobacterium salinarum* has the capability of bringing back deoxyriboses to the level of riboses. However, we should note that even if this pathway were in the archaeon, the presence of the final steps from 6-phospho-gluconate to PRPP make it likely that the latter can still be derived from 6-carbon sugars.

3.7.3   Biosynthesis of Aromatic Amino Acids

The biosynthetic pathways for aromatic amino acids are well established in bacteria and eukaryotes. In the archaea however, analysis of several genomes, including *Halobacterium salinarum*, has revealed examples where essential enzymes could not be assigned based on sequence searches (Higuchi *et al.*, 1999). Particularly, genes could not be found for the first two steps of the shikimate pathway, which itself comprises the early steps in the biosynthesis of aromatic amino acids. The first step in this established pathway is an aldol-like condensation of erythrose 4-phosphate with phosphoenolpyruvate to form 3-deoxy-D-arabino-2-heptulosonate 7-phosphate (DAHP), via the action of 2-dehydro-3-deoxyphosphoheptonate aldolase (EC 2.5.1.54). DAHP is then converted into 3-dehydroquinate (DHQ) by 3-dehydroquinate synthase (EC 4.2.3.4). Consistent with the apparent absence of the genes, biochemical and labeling experiments (Fischer *et al.*, 1993; Tumbula *et al.*, 1997) suggested that a noncanonical series of reactions is responsible for either the biosynthesis of DHQ or its precursor, erythrose-4-phosphate. It was recently shown that the former is true in *Methanocaldococcus jannaschii*, and that erythrose-4-phosphate is not at all involved in the pathway (White, 2004).

The first step in the recently elucidated pathway is a transaldolase reaction between L-aspartate semialdehyde and 6-deoxy-5-ketofructose 1-phosphate to form compound I. This step is catalyzed by the product of gene MJ0400. Next, compound I is oxidatively deaminated to 3,7-dideoxy-D-threohepto-2,6-diulosonic acid through the action of the MJ1249 gene-product. Finally, the result of the previous step is cyclized to DHQ, also via the action of the MJ1249 enzyme. A search for orthologs of MJ0400 and MJ1249 in the *Halobacterium salinarum* genome returned OE1472F and OE1475F, respectively. Both genes are in the tryptophan operon, which is a good indication that at least a similar path-

way is in operation in the halophile. Experimental effort aimed at validating the presence of the pathway, including the confirmation of the reported substrates, was initiated in part due to the reconstruction. We should note that evidence for the existence of the alternative pathway in *Halobacterium salinarum* was produced more than a decade ago, but obiously was not recognized as such, in a study that involved labeling of bacteriorhodopsin residues (Engelhard *et al.*, 1990). They described a then unexplained specific labeling of tryptophan by ^{13}C-aspartate. The labeling pattern they observed is consistent with the alternative precursor pathway proposed in *Methanocaldococcus jannaschii*. Finally, we should note that the menaquinone precursor chorismate is also derived from shikimate. Given that quinones are responsible for shuttling electrons between some respiratory complexes, then the proposed pathway is also relevant for oxidative phosphorylation.
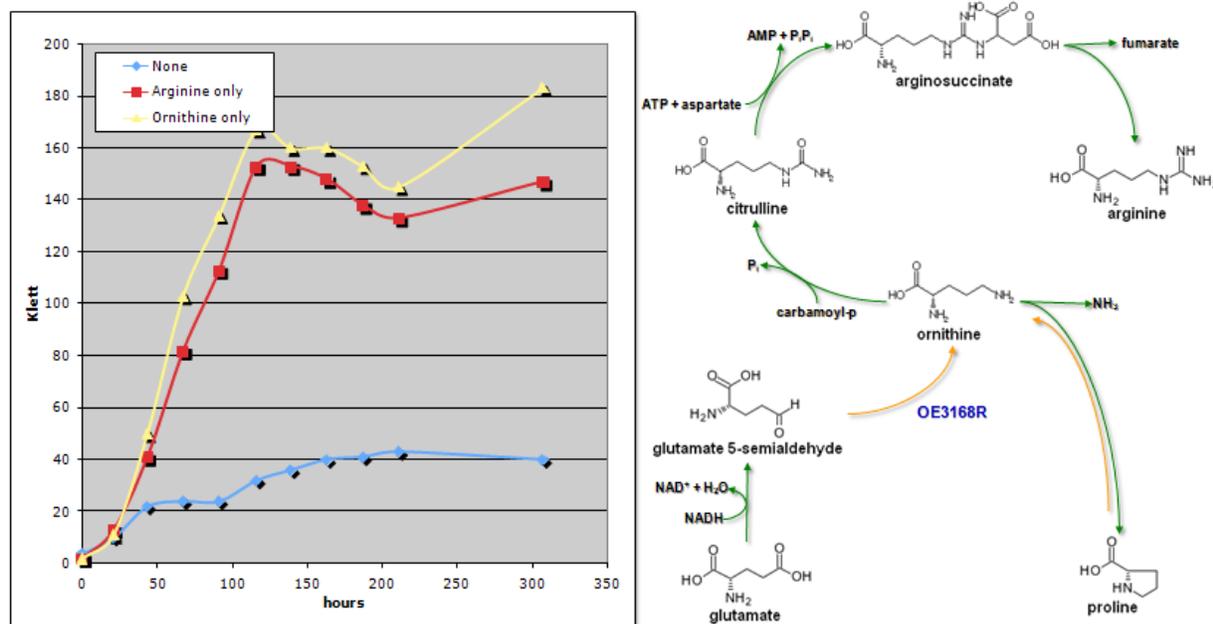
Figure 3.17: Essentiality of arginine. It has been a long-held assumption that arginine is an essential amino acid for *Halobacterium salinarum*. However, in one of our experiments, we were able to observe growth in a culture that had neither arginine nor ornithine included in the medium (left image, blue diamonds). This prompted us to reevaluate the parts of the primary annotation and the reconstruction that are related to arginine synthesis. From the gene equipment of the halophile, it is clear that all the genes necessary for converting ornithine to arginine are present (right image), and therefore the only problem is getting to ornithine itself. One possibility for achieving this is through a reversal of ornithine cyclodeaminase (EC 4.3.1.12; OE4121R) that normally allows ornithine to be broken down into proline and NH_3. An alternative possibility is to derive ornithine directly from glutamate 5-semialdehyde using ornithine aminotransferase (EC 4.3.1.12), for which candidate genes exist, including OE3168R that is currently annotated as a "pyridoxal phosphate-dependent aminotransferase (homolog to acetylornithine aminotransferase)".
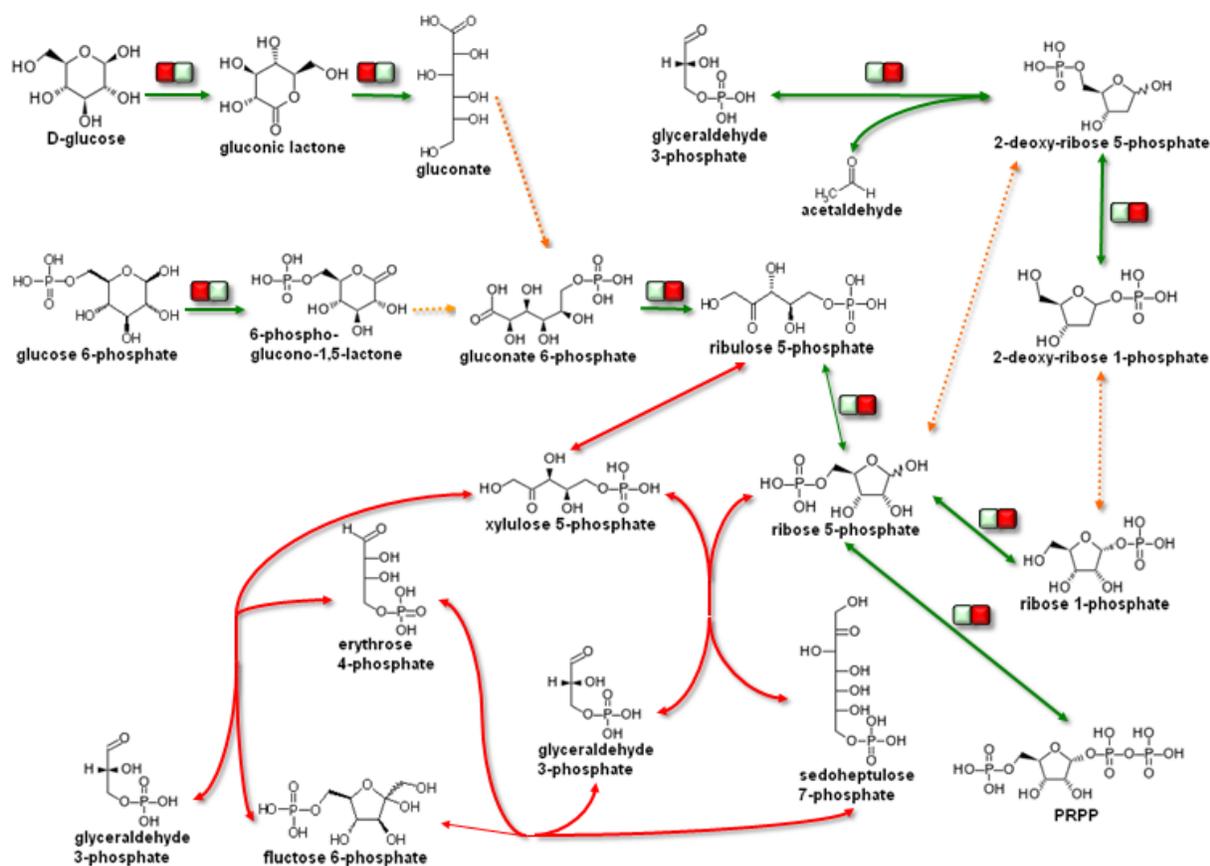
Figure 3.18: The Pentose Phosphate Pathway and ribose production. Green arrows represent biochemical transformations that are present in *Halobacterium salinarum*. Although some reactions do not occur in the halophile, they are nevertheless shown (red arrows) to complete the sequences to which they belong. A pair of adjacent boxes nearby each green arrow is used to signify the level of bioinformatic (left) and experimental (right) support associated with the corresponding reaction. Green denotes that the evidence is strong, red that there is none, and yellow indicates an intermediate level. As with all archaea, the pentose phosphate pathway is "broken" in *Halobacterium salinarum*. Possible reactions that can complete the production of the ribonucleic acid (RNA, DNA) precursor phosphoribosyl pyrophosphate (PRPP) are indicated in orange.
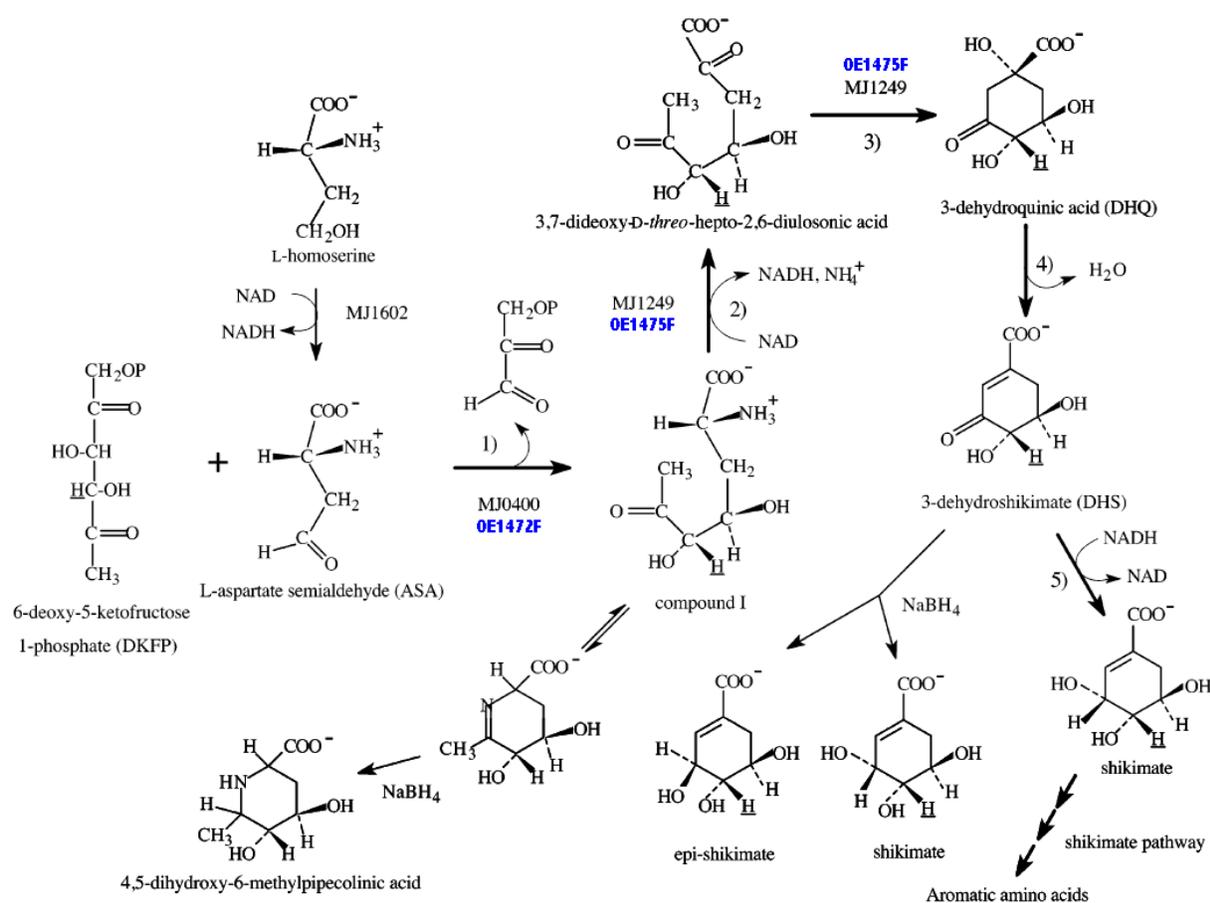
Figure 3.19: Biosynthesis of aromatic amino acids. It was recently shown that 6-deoxy-5-ketofructose 1-phosphate and L-aspartate semialdehyde are the precursors of aromatic amino acids in *Methanocaldococcus jannaschii*, rather than erythrose 4-phosphate and phosphoenolpyruvate. A search for orthologs of the two genes involved, MJ0400 and MJ1249, in the *Halobacterium salinarum* genome returned OE1472F and OE1475F (shown in blue in the figure), respectively. That orthologous genes exist in the halophile, and that both genes are in the tryptophan operon, is a strong indication for a similar pathway in *Halobacterium salinarum*. In part due to the reconstruction, experimental effort aimed at confirming the pathway, including the reported substrates, is currently underway. The figure was adapted from White, 2004.

Chapter 4

Systems Analysis of *H. salinarum* Aerobic Growth and Bioenergetics

This chapter deals with the development of a computational model from the metabolic reconstruction described in the previous chapter and on the use of the model for the analysis of growth under aerobic conditions. We begin with an introduction to dynamic models of metabolism, by describing some of the more popular formalisms currently used (Section 4.1). In Section 4.1.5, we describe the practical difficulties that are typically encountered in developing such dynamic models, which have limited their application to mostly small networks. We then move on to the constraints-based modeling framework, which circumvents these difficulties, in exchange for detail. The models used in this study are hybrids that have both kinetic (dynamic) and constraints-based components.

Although hybrid models with both kinetic (dynamic) and constraints-based aspects have been constructed in the past, the dynamic components of these models are too simple to be applicable to our problem at hand. In particular, previous approaches cannot account for transport patterns that change qualitatively during growth. This is a particularly restrictive limitation for this study. To overcome this, we propose a new system for defining the dynamic aspects of the hybrid models that not only can account for qualitatively changing transport patterns, but can actually detect their presence (Section 4.2.1). Moreover, our method can uncover relationships between metabolites, which is demonstrated by automatically detecting the relationship between arginine and ornithine that is defined by the arginine fermentation pathway.

This chapter also contains: (1) a description of how we characterized the *Halobacterium salinarum* biomass, including an experimental determination of the protein fraction (Sec-

tion 4.3); (2) the growth data we generated for the aerobic condition (Sections 4.2.3 and 4.2.4); and (3) results of our aerobic simulations (Sections 4.4 to 4.8.2).

4.1 Introduction to Metabolic Models

Considerable effort has been devoted to the mathematical description of metabolic function. In this section, several formalisms that have been developed for this purpose are reviewed. We begin by introducing kinetic modeling using Michaelis-Menten-type rate laws and two canonical modeling systems. Then, we describe the fundamentally different constraints-based framework, which, unlike the others, does not require kinetic information to be available. This distinction is significant because kinetic data are typically difficult to obtain. Accordingly, although constraints-based models are less detailed, much larger networks can be handled.

4.1.1 Chemical Kinetics

Chemical kinetics is the study of rates of chemical processes. In contrast to thermodynamics, chemical kinetics does not focus on energy levels, but rather on the temporal aspects of a reaction. For example, how fast does a reaction proceed? What factors affect the speed of the reaction? Certainly, the answers to these questions involve thermodynamic considerations at a deeper level. However, kinetic studies center directly on metabolite concentrations and fluxes.

In the overwhelming majority of studies, spatial features are ignored by chemical kinetic models, and it is implicitly assumed that all participating metabolites are available in a homogenous mix. A typical biochemical rate function relates the velocity of a reaction to the concentrations of the chemical species involved. For example, in the simple case of the first-order degradation of a metabolite X that does not involve an enzyme, the chemical reaction rate can be written as

$$\frac{dX}{dt} = \dot{X} = -kX \tag{4.1}$$

where k is a positive rate constant that represents the turnover per time unit. Equation 4.1 has the solution

$$X = X_0 e^{-k \cdot t} \tag{4.2}$$

where X_0 is the initial concentration of the substrate (that is, $X(t_0)$). Equation 4.2, which allows the direct evaluation of the concentration of X as a function of time, could be

derived analytically in this example, but this is not always the case. In kinetic studies, differentiation is almost always with respect to time (as in Equation 4.1). Factors that can affect a reaction's speed include the nature, physical state and concentration of the reactants, as well as temperature and the presence of catalysts.

4.1.2 Michaelis-Menten-type Rate Laws

In the early 1900's, Leonor Michaelis and Maud Menten proposed a reaction scheme for enzyme-catalyzed reactions. They postulated that the substrate S and the catalyzing enzyme E are in fast equilibrium with their complex (ES), which would then dissociate to yield either the product P and the free enzyme or the substrate and the free enzyme. The system can be mathematically written as

$$\dot{S} = k_{-1}(ES) - k_1 S \cdot E \tag{4.3}$$

$$(\dot{ES}) = k_1 S \cdot E - (k_{-1} + k_2)(ES) \tag{4.4}$$

$$\dot{P} = k_2(ES) \tag{4.5}$$

where k_1, k_{-1} and k_2 are rate constants. The assumption is that the laws of chemical kinetics apply; i.e., that all species are in a homogenous mixture, and that the concentrations are high enough to justify averaging what are essentially stochastic processes.

To obtain the familiar algebraic form of the Michaelis-Menten rate law, further assumptions are necessary (Torres and Voit, 2002, p.14). These are: (1) the total enzyme concentration E_T is constant (i.e., $E_T = E + (ES) = c$, where c is a constant); (2) the total substrate concentration S_T is much larger than the total enzyme concentration (i.e., $S_T >> E_T$); and (3) the concentration of the intermediate complex is constant (i.e., $(\dot{ES}) = 0$), which implies the earlier postulation that the processes forming and destroying (ES) are much faster than the overall conversion of S into P. With these assumptions, using equations 4.4 and 4.5 allows us to write the rate of product formation as

$$\dot{P} = \frac{V_{max} S}{K_M + S} \tag{4.6}$$

where $K_M = (k_{-1} + k_2)/k_1$ and $V_{max} = k_2 E_T$. This rate law form, along with many generalizations, has been immensely successful in the characterizaion of enzymes and the analysis of simple pathways in vitro. However, because of the numerous assumptions made in their derivations, care must be exercised in their use. For example, the supposition of a

constant total enzyme level makes these rate laws inapplicable to systems where regulation of the enzyme is of interest.

4.1.3   Biochemical Systems Theory

Biochemical systems theory (BST) is a canonical modelling framework that is based on ordinary differential equations (ODE). In it, biochemical processes are represented using power-law expansions in the variables of the system (Savageau, 1969a,b, 1970; Voit, 2000). It is based mathematically on Taylor's theorem; specifically by executing Taylor linearization in log space. A typical BST model will have the form

$$\dot{X}_i = \sum_{k=1}^{p} \gamma_{ik} \prod_{j=1}^{n+m} X_j^{f_{ijk}} \tag{4.7}$$

where n is the number of dependent variables, m is the number of independent variables, p is the number of processes affecting X_i, and γ_{ik} and f_{ijk} are parameters which are estimated at a specific operating point. As a testament to the generality of the formalism, differential equations from different contexts, including those that are recognized as among the most general, can be recast into equivalent BST forms (Irvine and Savageau, 1990). Accordingly, the system can handle complex behaviors such as oscillations, saturation and chaos, which are abundant in biology.

One very attarctive feature of BST models is that all parameters have well-defined biological interpretations. For example, the parameter f_{ijk} describes how a metabolite X_j affects the process k, which in turn affects metabolite X_i. Specifically, a value greater than zero for the parameter indicates a positive influence on the process, while a value below zero a negative influence, such as in the case of inhibition. Note that if f_{ijk} were zero, then $X_j^{f_{ijk}} = 1$, which means that metabolite X_j essentially drops out from the term for k. This property makes the formalism particularly useful for network inference (Kimura *et al.*, 2005; Gonzalez *et al.*, 2007; Liu and Wang, 2008). Past applications of BST to metabolic networks include glycolysis in *Lactococcus lactis* (Voit, 2006), sphingolipid metabolism in *S. cerevisiae* (Alvarez-Vasquez *et al.*, 2004), and human red blood cell metabolism (Ni and Savageau, 1995).

4.1.4 Linlog Kinetics

Metabolic control analysis (MCA) is a framework that was developed for the rational re-design of metabolic pathways (Kacser and Burns, 1973; Heinrich and Rapoport, 1974). In it, control aspects are quantified using coefficients that describe how changes in enzyme or extracellular effector concentrations affect fluxes and intracellular metabolite concentrations. These coefficients are defined relative to a reference state, which is typically the steady state. Accordingly, the practical applications of MCA are necessarily limited, because the theory is only valid for infinitesimal changes where linearization is a good approximation. One way of getting around this problem is to develop kinetic models from which control coefficients can be calculated or the optimal distribution of enzyme levels determined using numerical optimization. The linlog formalism is one such approach; specifically, it is an attempt to integrate kinetic modeling and MCA into one framework (Visser and Heijnen, 2002).

Similar to BST, the linlog framework uses approximative non-mechanistic rate equations. The main advantage of these approaches is that although they are not as detailed, the number of parameters that have to be estimated is decreased. Accordingly, the experimental effort needed for their identification is reduced as well. The argument is that metabolic re-design does not require detailed mechanistic models because of homeostasis constraints. In linlog, a reaction rate will typically have the form

$$v = e \left(a_0 + \sum_{j=1}^{n} a_j \ln x_j \right) \tag{4.8}$$

where e is the enzyme level, n is the number of metabolites (species) that affect the reaction, the x's are metabolite concentrations, and the a's are parameters. Clearly, the reaction rate is proportional to the enzyme level, and the effect of the metabolites is a linear sum of the (nonlinear) logarithms of their concentrations. Although developed in the context of metabolic redesign, linlog can evidently be also used for metabolic modeling in general.

4.1.5 Constraints-based Models

All of the formalisms that have been discussed so far are for the development of kinetic models. In order for any one of them to be applicable, knowledge of the intricate regulatory features of the system in question is required. Unfortunately, such information is normally difficult to obtain. Indeed, it is often the case that enzyme kinetic parameters are available

for only a limited number of model organisms. This, not even considering the fact that K_M values are typically derived from in vitro assays, a condition that casts suspicion on their validity in vivo (Teusink *et al.*, 2000; Theobald *et al.*, 1997; Rizzi *et al.*, 1997). Accordingly, kinetic models of metabolism have mostly been limited to relatively small networks.
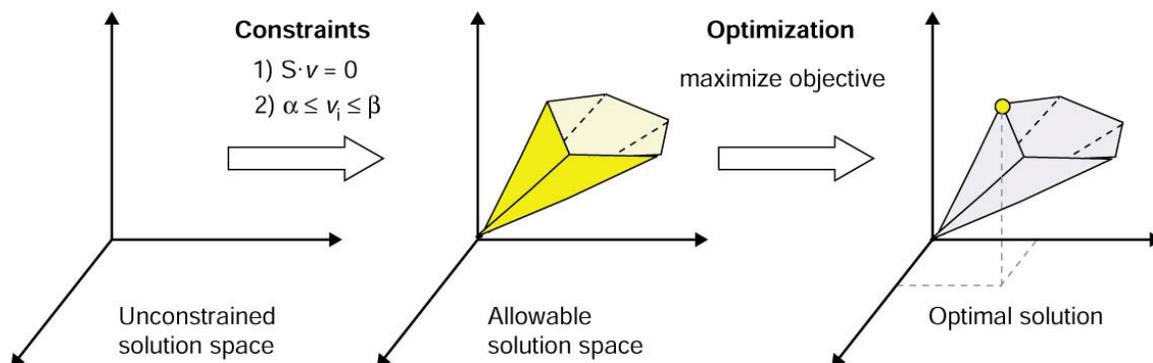
Figure 4.1: Principles of constraint-based modeling. As the name suggests, the idea of constraints-based modeling is to begin with a large, unconstrained solution space, and then proceed by narrowing it down through the addition of constraints. A three-dimensional flux space for a given metabolic network is depicted here. Without any constraints the fluxes can take on any real value (left). After application of stoichiometric, thermodynamic and enzyme capacity constraints, the possible solutions are confined to a region in the total flux space (center), termed the allowable solution space. Any point outside of this space violates one or more of the applied constraints. Given that it is typically the case that even with all constraints in place the solution space is still large, particular solutions can be identified using linear programming by introducing cellular objective assumptions, such as optimal ATP or biomass production. The figure was taken from Reed *et al.*, 2003.

In part due to the difficulties mentioned above, the constraints-based framework has emerged as a successful alternative to kinetic models. Rather than requiring detailed information that can be difficult to obtain, constraints-based models need only generally available physicochemical information such as stoichiometry, reversibility, energy balance, and, when available, reaction velocities (Edwards *et al.*, 2000; Edwards and Palsson, 1999; Ramakrishna *et al.*, 2001). As the name suggests, the principle of the framework is to begin with a large solution space, and then proceed by narrowing it down through the addition of constraints (Figure 4.1). To illustrate, consider a metabolic network with three reactions. Without any further information, the flux through each reaction can be any real number. However, if it is known that all three reactions are irreversible, then we can add the constraint that the flux through each must be greater than or equal to zero. Similarly,

if the exact flux through one of the reactions is known, for example through experimental measurement, then an equality constraint can be enforced for that reaction. Finally, given that it is typically the case with real networks that even with all the constraints in place the solution space is still large, particular solutions can be identified using linear programming by introducing cellular objective assumptions, such as optimality with respect to ATP or biomass production. This procedure is referred to as flux balance analysis.
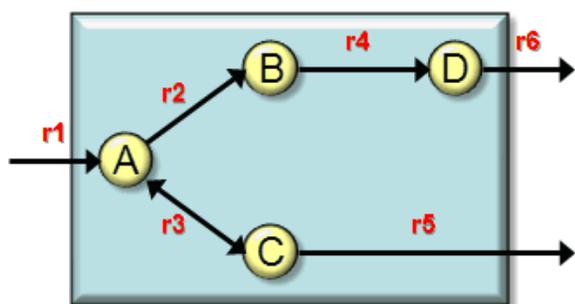
Figure 4.2: A small illustrative network. It is composed of the four metabolites: A, B, C and D; and the six reactions r1, r2, r3, r4, r5 and r6. The box is used to differentiate between inter- and intracellular space.

Since the metabolic network for *Halobacterium salinarum* has 695 reactions (Section (3.4), it is not very convenient for illustrating the development of a stoichiometric matrix, which is central to constraints-based models. For this purpose, we use the small network depicted in Figure 4.2. The network is essentially a branching pathway consisting of four metabolites (A, B, C and D) and six reactions (r1 through r6). It can be conveniently represented as the stoichiometric matrix

$$
\mathbf{S} = \begin{array}{c} \\ A \\ B \\ C \\ D \end{array}
\begin{array}{cccccc}
r1 & r2 & r3 & r4 & r5 & r6 \\
\left[\begin{array}{cccccc}
1 & -1 & -1 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & -1
\end{array}\right]
\end{array}
\tag{4.9}
$$

where each row corresponds to a metabolite and each column to a reaction. The entries of $\mathbf{S}$ are the stoichiometric coefficients that define the relationships between the reactions and compounds. A positive value for s_{ij} indicates that compound i is produced in the left to right direction of reaction j, while a negative value indicates that it is consumed. For example, column two of $\mathbf{S}$ is defined as $[-1\,1\,0\,0]^T$ indicating that A is consumed and B is produced by reaction r2.

As discussed in the the previous chapter, metabolism is responsible for converting the available nutrients into usable energy and biomass (growth). In the case of *Halobacterium*

salinarum, defined media used for the archaeon typically consist of amino acids. Therefore, in order for growth to be possible, its metabolism should be able to convert the amino acids into the unsupplied components of its biomass, such as nucleotides, vitamins, and other cofactors. The processes involved in these conversions can easily be translated into constraints by writing material balance around the metabolic network, using its stoichiometric matrix $\mathbf{S}$ as in

$$\mathbf{S} \cdot \boldsymbol{v} = \boldsymbol{b} \tag{4.10}$$

where $\boldsymbol{v}$ is a vector of reaction rates, and $\boldsymbol{b}$ is a vector containing the net metabolite concentration changes. Note that unlike in kinetic models where reaction velocities are directly computed from rate equations, metabolic fluxes are the unknown quantities that need to be determined in flux balance models.

In order to simulate growth, a pseudoreaction that is often called the "growth function" is added to the network. Formally, it is defined as a reaction where the reactants are cellular constituents, and the product is a unit of biomass. The stoichiometric coefficients used for the reactants are determined with respect to the unit of biomass used. For example, if the product is defined to be one gram of biomass, and it is known that this quantity on average contains a moles of alanine, then a is used as the stoichiometric coefficient of alanine. The growth function is added to the stoichiometric matrix $\mathbf{S}$, just like any other reaction in the network. Although terms relating to energy maintenance are often integrated into the growth reaction, this was not done in this study because we investigate systems where growth rates vary with time.

As mentioned earlier, it is often the case with real networks that the available constraints are not enough to determine a unique solution; i.e, the systems are frequently underdetermined. Nevertheless, particular solutions can be obtained using linear programming, using the assumption that cells optimize their metabolism with respect to certain objectives, such as growth and energy production. For example, the linear programming problem can be posed as

$$\text{Maximize } f = v_{obj} \tag{4.11}$$

Subject To

$$\mathbf{S} \cdot \boldsymbol{v} = \boldsymbol{b} \tag{4.12}$$

$$-\infty \leq v_j \leq +\infty \text{ where } r_j \in R_{rev} \tag{4.13}$$

$$0 \leq v_j \leq +\infty \text{ where } r_j \in R_{irrev} \tag{4.14}$$

$$v_j = c_j \text{ where } r_j \in R_{known} \tag{4.15}$$

where *obj* is the index of the objective reaction (e.g., the growth reaction), R_{rev} is the set of reversible reactions, R_{irrev} is the set of irreversible reactions, R_{known} is the set of reactions for which fluxes are already known, and c_j is the known flux through each reaction in R_{known}. Note that metabolic flux models are based on the separation of time scales between cellular growth rates and metabolic transients. That is, metabolism typically has transients that are shorter than a few minutes, while growth is typically measured in hours. Accordingly, metabolic fluxes are assumed to be in a quasi-steady state relative to growth.

The mechanisms by which metabolic flux distributions are chosen is a complex interplay of regulatory events at different levels. More often than not, only a limited subset of these are known in detail. Therefore, one may be surprised at the attempt of determining a unique solution for considerably underdetermined systems by just using an objective function. Nevertheless, the argument is that although the regulatory mechanisms for the most part are unknown, metabolic networks, because of evolutionary pressures, are adapted to work in concert to achieve "optimal metabolism", so that the survivability of cells are enhanced. It is expected that wild-type microorganism strains have metabolic phenotypes that are defined by a tendency to optimize their growth rates, at least in the environments where they are found.

To investigate the extent to which optimality principles can describe the operation of metabolic networks, Schuetz and coworkers systematically evaluated flux balance predictions using 11 objective functions against actual in vivo fluxes. This was done in *Escherichia coli* using data derived from ^{13}C-based flux analysis (Schuetz *et al.*, 2007; Sauer, 2006). They reported that although no single objective adequately described the flux states under all conditions tested, they were able to identify two sets that allowed for biologically meaningful predictions without the need for further, potentially artificial constraints. Specifically, they found that growth on glucose in oxygen or nitrate respiring batch cultures is best described by nonlinear maximization of the ATP yield per flux unit, and that linear maximization of the overall ATP or biomass yield achieves the highest predictive accuracy in continuous cultures under nutrient scarcity. It was concluded that the identified optimality principles reflect, to some extent, the evolutionary selection of metabolic network regulation that realizes the various flux states.

4.2   Methods

This study includes both computational and experimental work; a considerable fraction of the data that were used for computational analysis were generated here. The computational and experimental methods that were employed are described in this section.

4.2.1   Metabolite Consumption and Production Equations

The model for the consumption and production of metabolites had to satisfy the following: (1) allow for multiple metabolites in the medium; (2) account for the possible (dynamic) interactions between the supplied metabolites; and (3) account for changes in transport rates/modes that may be brought about by time-varying growth rates. The first requirement is because *Halobacterium salinarum* requires relatively rich media consisting of amino acids, unlike other microorganisms such as *Escherichia coli* that can live on a single carbon substrate. The second is because preliminary inspection of the nutrient time-course data revealed that the utilization patterns of some metabolites change during growth; i.e., they exhibit distinct modes. And the third is because our simulation of batch-grown cultures goes past the logarithmic phase of growth.

For each nutrient X_i with a transport pattern that does not change significantly, we used a simple model consisting of three terms:

$$\dot{X}_i = k_{X,i}X_{i,t} + k_{\gamma,i}\gamma_t + k_{\dot{\gamma},i}\dot{\gamma}_t \tag{4.16}$$

where $\dot{X}_i$ is the uptake rate of nutrient i, γ_t is the population size at time t, $\dot{\gamma}_t$ is the current growth rate defined as $\frac{d\gamma}{dt}(t)$, and $k_{X,i}$, $k_{\gamma,i}$ and $k_{\dot{\gamma},i}$ are optimizable parameters. The rationale of the construction is that the production or consumption of the metabolite depends on the availability of the metabolite, the population size and the current growth rate.

As mentioned earlier, inspection of the nutrient time-course data indicated that the transport patterns of some metabolites qualitatively changes during growth. For example, ornithine, at some point, switches from being produced and accumulated in the medium to being steadily consumed by cells. Such changes may be brough about by interactions with other metabolites, or by general changes in metabolism. For these nutrients we used

a simple extension of Equation (4.16) that allows for two modes:

$$\dot{X}_i = \begin{cases} k_{1,X,i}X_{i,t} + k_{1,\gamma,i}\gamma_t + k_{1,\dot{\gamma},i}\dot{\gamma}_t, & \text{if } t < t_{i,b} \\ k_{2,X,i}X_{i,t} + k_{2,\gamma,i}\gamma_t + k_{2,\dot{\gamma},i}\dot{\gamma}_t, & \text{if } t \geq t_{i,b} \end{cases} \tag{4.17}$$

where $k_{1,X,i}$, $k_{1,\gamma,i}$, $k_{1,\dot{\gamma},i}$, $k_{2,X,i}$, $k_{2,\gamma,i}$, $k_{2,\dot{\gamma},i}$ and $t_{i,b}$ are optimizable parameters. Similar to the basic form, the consumption or production of each metabolite still depends on the availability of the metabolite, the population size and the current growth rate. However, the new parameter $t_{i,b}$ now separates time into two intervals, and the distinct set of parameters for each of these allows the utilization pattern to qualitatively change in moving from the first to the second.

The use of either Equation 4.16 or Equation 4.17 was not chosen *a priori*. Rather, the choice was made based on clear indications of improvement or the lack thereof. Specifically, for each nutrient X_i, parameters were estimated for both equation forms by solving inverse problems (parameter optimization) using experimental data. Systematically defined sets of initial guesses were used in solving the inverse problems (parameter estimation). Based on the results, the more sophisticated form of Equation 4.17 was used in the final model, in favor over the simpler equation form, only if it resulted in a significantly better fit; i.e, Equation 4.17 was used only if it led to a substantially reduced residual error, which is a measure of the (dis-)agreement between model and data, when compared to that obtained using Equation 4.16. Making the choice between the two forms in this way minimizes the possibility of overfitting. In this respect, it is reassuring that although Equation 4.17 is clearly better for some metabolites such as alanine and ornithine, it hardly makes any improvement for others such as leucine and isoleucine that seem to be staples of the cells (see results). In cases where Equation 4.17 is used, the optimized parameter $t_{i,b}$, intuitively, corresponds to a point near where the transport pattern changes qualitatively. Because the choice between the two equation forms can be done automatically, the computational system we employ not only accounts for changes in metabolite modes, but can actually lead to their identification, and, subsequently, to the recognition of the biological processes behind them. The borders ($t_{i,b}$) need not be equal for all X_i's.

4.2.2 Hybrid Genome-scale Flux Balance Model

At particular intervals during growth, we obtained fluxes that are consistent with the observed nutrient depletion/accumulation and biomass formation by solving the linear

program

$$\text{Maximize } f = v_{obj} \tag{4.18}$$

Subject To

$$\mathbf{S} \cdot \boldsymbol{v} = \boldsymbol{0} \tag{4.19}$$

$$-\infty \leq v_j \leq +\infty \text{ where } j \in (R_r \cup E_f) \tag{4.20}$$

$$0 \leq v_j \leq +\infty \text{ where } j \in R_i \tag{4.21}$$

$$-\infty \leq v_j \leq 0 \text{ where } j \in E_u \tag{4.22}$$

$$v_j = \beta_j \text{ where } j \in E_m \tag{4.23}$$

where $\boldsymbol{v}$ is a vector of fluxes defining the flux v_j through each reaction j, v_{obj} is the objective function, R_r is the set of reversible internal reactions, R_i is the set of irreversible internal reactions, E_f is the set of exchange fluxes associated with ubiquitous metabolites, and E_m is the set of exchange fluxes that correspond to experimentally measured nutrients. The set of ubiquitous compounds include CO_2, H_2O, Na^+, Cl^- and H^+. For each reaction $j \in E_m$, the value β_j is the appropriate evaluation of the corresponding uptake equation (4.16 or 4.17) for the interval. In this work, all of the supplied carbon and energy sources, except for citrate, were included in E_m. In addition, the set also includes ornithine, which is initially accumulated in the medium, and oxygen. Biomass production is treated in a similar manner by fixing the flux of the growth (pseudo-)reaction to observed rates. To account for the possibility that cells produce and accumulate certain metabolites in the medium, such as in the case of overflow metabolism, the set of one-way exchange reactions E_u was introduced. It includes central metabolites such as acetate, pyruvate, malate and GAP, nucleotides, and some sugars such as glucose. Unless otherwise specified, energy production was used as the objective function.

4.2.3   Culture Conditions and Sample Preparation

Strain *Halobacterium salinarum* R1 (DSM 671) cells were grown in chemically-defined media, with composition defined in Table 4.1. Preparatory cultures were grown in 100 ml flasks containing 35 ml of the medium to a cell density of ≈100 Klett (1 OD), from which 1 ml inoculants were taken to start the next culture. This was done repeatedly to adapt cells to the growth conditions. All cultures were prepared in flasks which had side arms to measure turbidity (cell density) via a Klett photometer, and were carried out in duplicates.

Cell suspensions were shaken at 105 rpm at 40°C in the dark. At specific points, samples were taken from the cultures so that 14-18 samples were collected over the growth period, and these were stored at 4°C. To separate the cells from the medium, the samples were centrifuged for five minutes at 15,000 rpm, using a SS34 rotor. Pellets were resuspended in 500 μl basal salt (medium without the amino acids) and spun down as before. Amino acid analysis was performed on both the pellets and the original supernatants, using an Amino Acid Analyzer (Biotronik LC3000).

Table 4.1: Composition of chemically-defined medium[a]

Description	Value	Description	Value
NaCl	4 M	L-Glycine	1.0 mM
KNO$_3$	1 mM	L-Alanine	2.5 mM
KCl	27 mM	L-Arginine	3.5 mM
MgSO$_4$ * 7H$_2$O	81 mM	L-Isoleucine	3.4 mM
Sodium Citrate * H$_2$O	1.7 mM	L-Methionine	1.3 mM
K$_2$HPO$_4$	0.42 mM	L-Proline	0.9 mM
KH$_2$PO$_4$	0.58 mM	L-Phenylalanine	0.8 mM
FeSO$_4$ * 7H$_2$O	11.6 μM	L-Serine	5.8 mM
CuSO$_4$ * 5H$_2$O	0.2 μM	L-Threonine	4.2 mM
MnCl$_2$ * 4H$_2$O	1.8 μM	L-Tyrosine	1.1 mM
ZnSO$_4$ * 7H$_2$O	1.5 μM	L-Valine	2.5 mM
Na$_2$MoO$_4$ * 2H$_2$O	0.1 μM	L-Lysine	2.1 mM
L-Aspartate	3.0 mM	Thiamin	16.5 μM
L-Glutamate	13.5 mM	Folate	11.5 μM
L-Leucine	6.1 mM	Biotin	2.1 μM

[a]pH was adjusted to 7.2

4.2.4   Oxygen Consumption

The amount of dissolved oxygen in the medium (solution) was continuously monitored using the "Fibox 3-trace v3, fiber-optic oxygen meter" from Precision Sensing GmbH (Regensburg, Germany). Under conditions identical to those of the aerobic cultures, argon was blown into a flask that contains only the medium. When the amount of dissolved oxygen in the solution dropped to zero, the argon source was removed, and the flask was equilibrated with air. We used the subsequent nonlinear rise of oxygen from 0% to 100% saturation to characterize oxygen dissolution kinetics. Specifically, the oxygen transfer rate

(OTR) was defined as

$$OTR = \frac{dx}{dt} = k\left(x_{max} - x\right) \tag{4.24}$$

where x is the amount of dissolved oxygen, and x_{max} is the maximum value of x, which indicates saturation. The data and the model fit are shown in Figure 4.3. With the appropriate parameters for Equation (4.24), we then calculated the oxygen consumption rate of a culture at time t using

$$consumption(t) = k\left(x_{max} - x\right) + m \tag{4.25}$$

where m is the current rate at which x is changing in the culture. For the sake of completeness, we note that Equation (4.24) has the analytical solution

$$x = x_{max} - \left(x_{max} - x_0\right)e^{-kt} \tag{4.26}$$

where $x(0) = x_0$ is the initial condition.
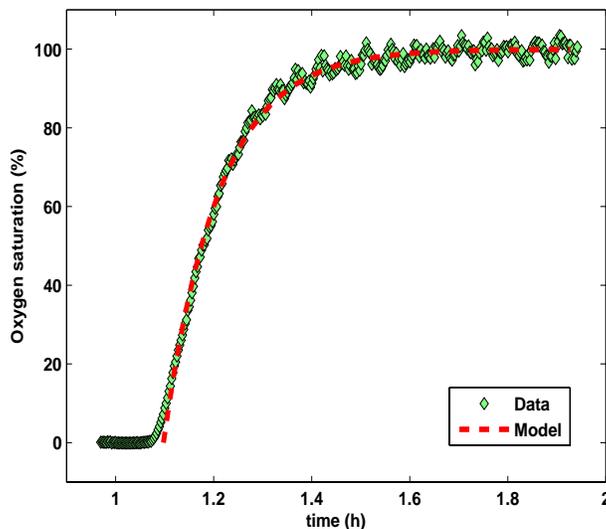
Figure 4.3: Dissolution kinetics of oxygen. Under conditions identical to those of the aerobic cultures, argon was blown into a flask that contained only the medium. When the amount of dissolved oxygen in the solution dropped to zero, the argon source was removed, and the flask was equilibrated with air. The graph shows the velocity at which oxygen subsequently dissolved into the cell-free solution. This was used to calibrate oxygen consumption measurements.

4.3   Determination of Biomass Composition

The growth function described in Section 4.1.5 stoichiometrically correlates a unit of biomass with the average composition of cells. Say that this pseudoreaction corresponds to the production of one cell and that it is associated with column j of the stoichiometric matrix $\mathbf{S}$, then the value s_{ij} for each compound i that is a biomass consituent should reflect how much of i is present in a cell on average. For example, there are about 186.3 attomoles of alanine in a cell. If i_{ala} is the index for the amino acid in $\mathbf{S}$, then $s_{i_{ala},j} = -186.38$. Other units of biomass are also possible, such as dry weight or optical density. Although the growth reaction is typically used as the objective function in determining particular flux distributions during flux balance analysis, in this study it is often also treated as a constraint that is equated to observed growth rates. The approximate biomass composition of *Halobacterium salinarum* is summarized in Table 4.2. We describe how we determined the individual components in the following subsections.

4.3.1   Amino Acids

Amino acids are generally present in cells in two forms; as protein residues or as free metabolites. Given that proteins are ubiquitous components of cells, amino acids generally constitute a considerable fraction of the biomass. For *Halobacterium salinarum*, the cellular concentration of each was experimentally determined by performing amino acid analysis on samples taken at different optical densities. The measurements for four distinct cultures that were grown and processed separately are illustrated in Figure 4.4. The values are inclusive of protein residues and free metabolites. Due to experimental limitations, the data for aspartate and glutamate are already inclusive of asparagine and glutamine, respectively, and the cysteine and tryptophan content could not be measured.

The cellular concentration of each amino acid shows a reasonably good linear correlation with the optical density (red curves in Figure 4.4). This supports the idea of an average cell, which is an assumption that is made by most genome-scale flux balance models that simulate growth. The total value of 381.4 μg/OD·ml that we determined, which corresponds to 49% of the cellular dry weight or 73% of the total organic mass, is consistent with measurements of total protein content that indicated a value of approximately 400 μg/OD·ml (Koch, 2005). However, we should note that the cultures represented in Figure 4.4 are not the ones that were used for computational analysis; these were only used for the initial determination of amino acid content. Although the conditions used are gener-

Table 4.2: Average cellular biomass composition.

Compound	Amount (μg/ODml)	Compound	Amount (μg/ODml)
Amino Acids[a]		Nucleotides[b]	
Alanine	22.5 ± 5.2	dAMP	0.6
Arginine	26.9 ± 5.6	dTMP	0.6
Aspartate	42.0 ± 9.2	dGMP	1.3
Asparagine	9.6	dCMP	1.1
Cysteine	2.9	AMP	15.7
Glutamate	77.8 ± 23.2	UMP	14.7
Glutamine	13.4	GMP	34.9
Glycine	15.9 ± 2.9	CMP	31.1
Histidine	15.1 ± 3.0		
Isoleucine	13.1 ± 3.0	S-Layer non AA	
Leucine	24.1 ± 5.0	Glucosamine	0.6
Lysine	12.3 ± 2.5	Galactosamine	0.6
Methionine	5.4	Glucose	2.2
Phenylalanine	13.4 ± 2.6	Galactose	2.2
Proline	18.5 ± 4.5		
Serine	13.9 ± 2.8	Membrane	
Threonine	19.3 ± 4.2	Archaeol	20.0
Tryptophan	7.7		
Tyrosine	6.3 ± 3.0	Others[c]	
Val	21.4 ± 4.9	ATP	2.0

[a]Amino acid content was measured from samples taken at different optical densities.

[b]The exact ratio between DNA and RNA is not known. Nevertheless, the total is consistent with a 610 fg cellular dry weight, after subtracting the other major components of the biomass.

[c]Cells also contain other compounds that may be essential to survival but presumably contribute only little to the biomass. These metabolites were added to the growth function by assuming that each was 0.01% of the total organic mass. These are siroheme, FAD, NAD+, NADP+, tetrahydrofolate (THF), thiamin, coenzyme-A, menaquinone, molybdenum cofactor, and cobamide coenzyme.
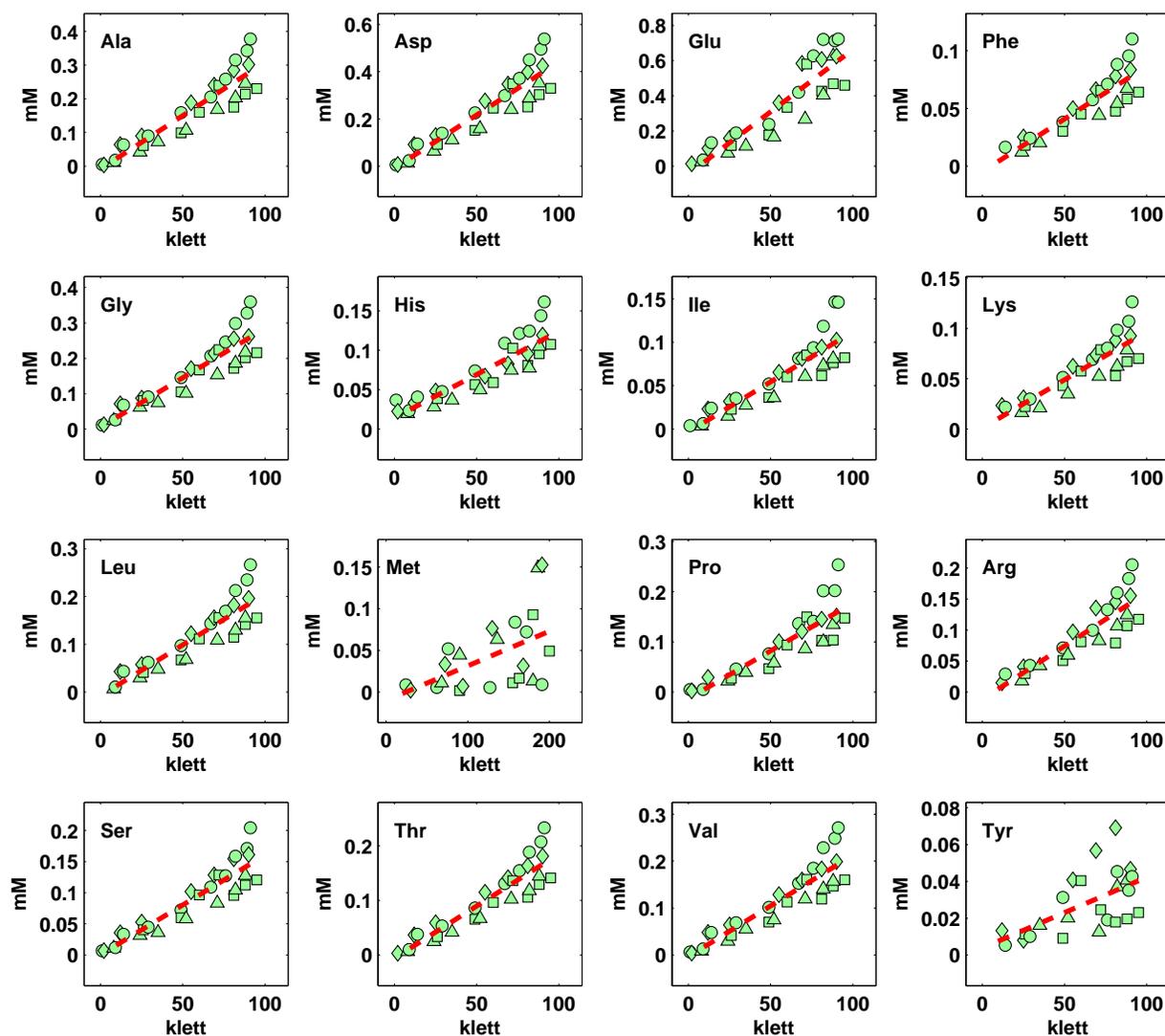
Figure 4.4: Amino acid composition of the *Halobacterium salinarum* biomass. The amino acid content of cells was measured at different optical densities. The values shown for each amino acid represent the total amount, including protein residues and free metabolites. Different filled shapes are used to represent 4 distinct cultures that were grown and processed separately. Red broken lines are used to show the best linear fits. Due to experimental limitations, the data for aspartate and glutamate are already inclusive of asparagine and glutamine, respectively, and the cysteine and tryptophan content could not be measured.

ally similar, there are some differences, including the fact that these preparations were not inoculated from cells that were as well-entrained. Amino acid analysis was repeated for the cultures that were actually used for the growth simulations.

4.3.2   Nucleic Acids

Nucleic acids are macromolecules composed of chains of monomeric nucleotides, which are universal in living things. These molecules carry genetic information or are parts of structures within cells. The most common nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The former contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules is the long-term storage of information. In *Halobacterium salinarum*, DNA is organized into a main chromosome and four megaplasmids. We determined the dAMP, dTMP, dGMP and dCMP stoichiometric coefficients of the growth function by counting directly from the DNA structures, assuming one copy of each.

Although very similar to DNA, RNA is different in important structural details. For one, while DNA is usually double stranded in cells, RNA can typically be found as single strands. In addition, as the names suggest, RNA nucleotides contain ribose, while DNA contains deoxyribose. RNA is produced from DNA through a process called transcription. One of its main functions is the production of proteins. Specifically, a type of RNA called messenger RNA (mRNA) carries information from DNA to structures called ribosomes. These ribosomes can read mRNAs and translate the information they carry into proteins. RNAs are also used for regulatory purposes, and are utilized by some viruses to replicate their genomes by reverse transcription. Ribonucleic acids were approximated to comprise 20% of a cell's organic mass, and the individual copy numbers of AMP, UMP, GMP, and CMP were computed by weighting the total RNA content based on the chromosome's 68% GC composition.

The exact ratio between DNA and RNA in *Halobacterium salinarum* is not known. Nevertheless, the current total nucleic acid content in the growth function is consistent with a 610 fg cellular dry weight, after subtracting the other major components of the biomass (proteins, lipids, salt, etc...). With respect to the computational model (Section 4.2.2), the ratio influences the fluxes through only a very small part of the network because DNA and RNA share most of the steps of their biosynthetic pathways.

4.3.3 Surface Layer Glycoproteins

Rather than employing a true cell wall, *Halobacterium salinarum* uses what is called a surface-layer as its outermost covering (Mescher and Strominger, 1976). This monolayer covers the surface of the archaeon in a hexagonal macromolecular pattern. It consists of subunits of a glycoprotein that is essential for maintaining the rod shape of the microorganism. The cell surface glycoprotein (CSG) contains three types of protein-linked saccharides (Lechner and Sumper, 1987): (1) a single high molecular weight saccharide, which is composed of repeats of sulfated pentasaccharides, that is connected to the polypeptide chain via an Asn-GalNac linkage (Paul *et al.*, 1986); (2) about 10 sulfated oligosaccharides that are connected to the polypeptide via Asn-Glc linkages (Wieland *et al.*, 1983); and (3) about 20 disaccharides (Glc-Gal) which are O-glycosidically attached to threonine residues of the polypeptide chain (Mescher and Strominger, 1976). Although the contribution of the polypeptide backbone of the CSG molecule to the biomass is already accounted for in the growth function (Section 4.3.1), the carbohydrate units had to be included separately. A schematic of the CSG is shown in Figure 4.5.
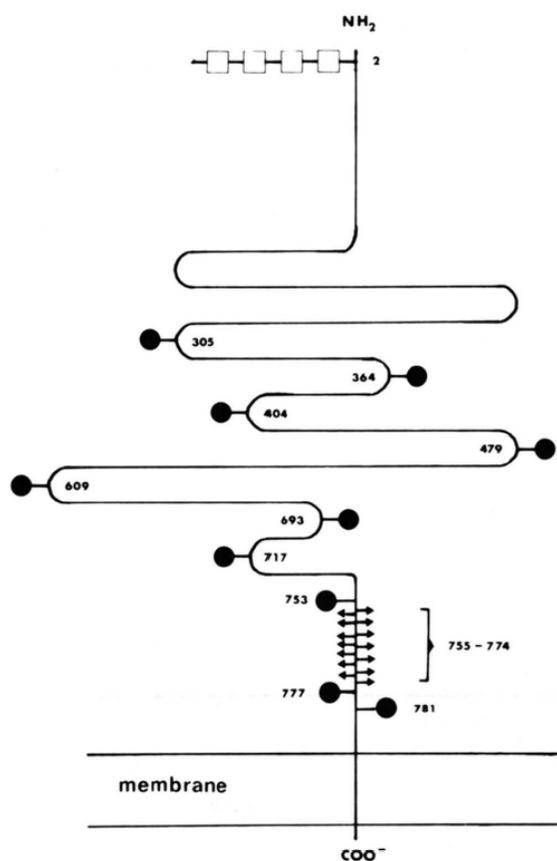
Figure 4.5: Schematic representation of the cell surface glycoprotein. Numbering indicates amino acid positions in the polypeptide chain (black solid curve). The unfilled boxes represent the repeated saccharide units. The filled circles indicate the N-linked oligosaccharides. The arrows represent the 0-linked Glc-Gal disaccharides. The image was taken from Lechner and Sumper, 1987. The carbohdrate and sulphur units are the relevant molecules for the definition of the growth function since the polypeptide chain should already be accounted for by the amino acid determination of the biomass (Section 4.3.1).

The CSG accounts for approximately 50% of the protein and all of the nonlipid carbohydrates in the surface layer. We estimated its cellular copy number using structural data from the closely related archaeon *Haloferax volcanii* (Kessel *et al.*, 1988); secifically by fitting as much of the p6 lattice arrangement over the average surface area of *Halobacterium salinarum* cells ($9.99 \times 10^{-8} \text{cm}^2$). We then calculated the individual copy numbers of the non-protein components using charaterization data of the CSG (Lechner and Sumper, 1987; Mescher and Strominger, 1976). Since there is still no assembly pathway for the complete structure, its components were included separately in the growth function.

4.3.4   Membrane Lipids

Members of the family *Halobacteriaceae* have archaeal-type lipids that are based on isoprenyl chains attached to glycerol (Kamekura, 1993, 1998; Kamekura and Kates, 1988). The diether core lipid is mostly 2,3-di-*O*-phythanyl-*sn*-glycerol (C_{20}, C_{20}), although some species in addition contain the asymmetiric 2-*O*-sesterterpanyl-3-*O*-phytanyl-*sn*-glycerol (C_{25}, C_{20}) (De Rosa *et al.*, 1982; Dyall-Smith, 1995; Matsubara *et al.*, 1994) and/or 2,3-di-*O*-sesterterphanyl-*sn*-glycerol in different amounts (De Rosa *et al.*, 1983). Asymmetiric core lipids (C_{25}, C_{20}) were also detected in *Halobacterium halobium* IAM13167, which may be a *Halobacterium salinarum* strain (Morita *et al.*, 1998).

For the sake of simplicity, and considering that complete biosynthetic pathways for archaeal lipids have not yet been completely elucidated, only archaeol (2,3-di-*O*-phythanyl-*sn*-glycerol) was used in the growth function. We assumed that lipids and proteins each occupy 50% of the total cellular surface area. Accordingly, we approximated the lipid's copy number by dividing the surface area of a cell by the area occupied by the molecule ($65A^2$), since the membrane is a bilayer.

4.4   Consumption and Production of Nutrients

Consumption and production rates were modeled using differential equations of the form of either Equation 4.16 or Equation 4.17 (see Section 4.2.1), with the rational that the production or consumption of a metabolite depends on the availability of the metabolite, the population size, and the current growth rate; note that growth-rate is time-varying in batch cultures. The piecewise extension of Equation 4.16, Equation 4.17, was necessary because several metabolites exhibited distinct modes over the growth period. For example, alanine switches from production to consumption. After obtaining parameters for a

Table 4.3: Comparison of transport equation forms.

Nutrient	Basic	Piecewise	Error reduction
Alanine	0.3946	0.0300	92.4%
Aspartate	0.0135	0.0133	0.8%
Glutamate	0.4211	0.3947	6.3%
Phenylalanine	0.0009	0.0009	1.0%
Glycine	0.0075	0.0043	43.2%
Isoleucine	0.0171	0.0171	0.2%
Lysine	0.0245	0.0227	7.3%
Leucine	0.0599	0.0593	1.0%
Methionine	0.0144	0.0112	22.4%
Proline	0.0367	0.0167	54.5%
Serine	0.1810	0.1299	28.2%
Threonine	0.0803	0.0575	28.4%
Valine	0.0196	0.0128	35.0%
Tyrosine	0.0087	0.0082	6.6%
Ornithine	2.3873	0.0325	98.6%

The transport rate of each compound was modeled using a differential equation of the form of either Equation (4.16) or Equation (4.17). The latter equation form, which is a piecewise version of the former, was necessary because the uptake patterns of several nutrients qualitatively change during growth. Columns 2 and 3 above list the best (lowest) residual error values for Equation (4.16) and Equation (4.17), respectively. The rightmost column indicates how much the error is reduced by using the piecewise version over the simpler form.

particular nutrient using both equations, we used the piecewise form only if it resulted in at least a 10% residual error reduction in order to minimize the possibility of overfitting.

The best (lowest) residual error values we were able to obtain for each nutrient using the two equation forms are listed in Table 4.3. Inspection of the results clearly shows that the improvement gained in using the piecewise form is nearly negligible for a number of metabolites, including aspartate, isoleucine, and leucine. In fact, it is very likely that the slight decreases in the relevant residual errors are just due to overfitting to the noise in the data. For the three amino acids mentioned, residual error values decreased by only 0.8%, 0.2%, and 1%, respectively. Considering that these same three amino acids, after arginine, are also the ones with the highest uptake rates, it would seem that they are the preferred or primary metabolites of *H. salinarum* during aerobic growth, at least among the nutrients supplied. Unlike most of the other provided metabolites, their uptake patterns remained constant, and their consumption rates high, up until their depletion. The experimental nutrient utilization data and the corresponding model simulations are shown in Figure 4.6.

In contrast to the situation with aspartate, leucine and isoleucine, results showed (Table 4.3) that the piecewise transport equation form is clearly superior over the basic definition (Equation 4.16) for ornithine, alanine and proline. For the latter amino acids, residual
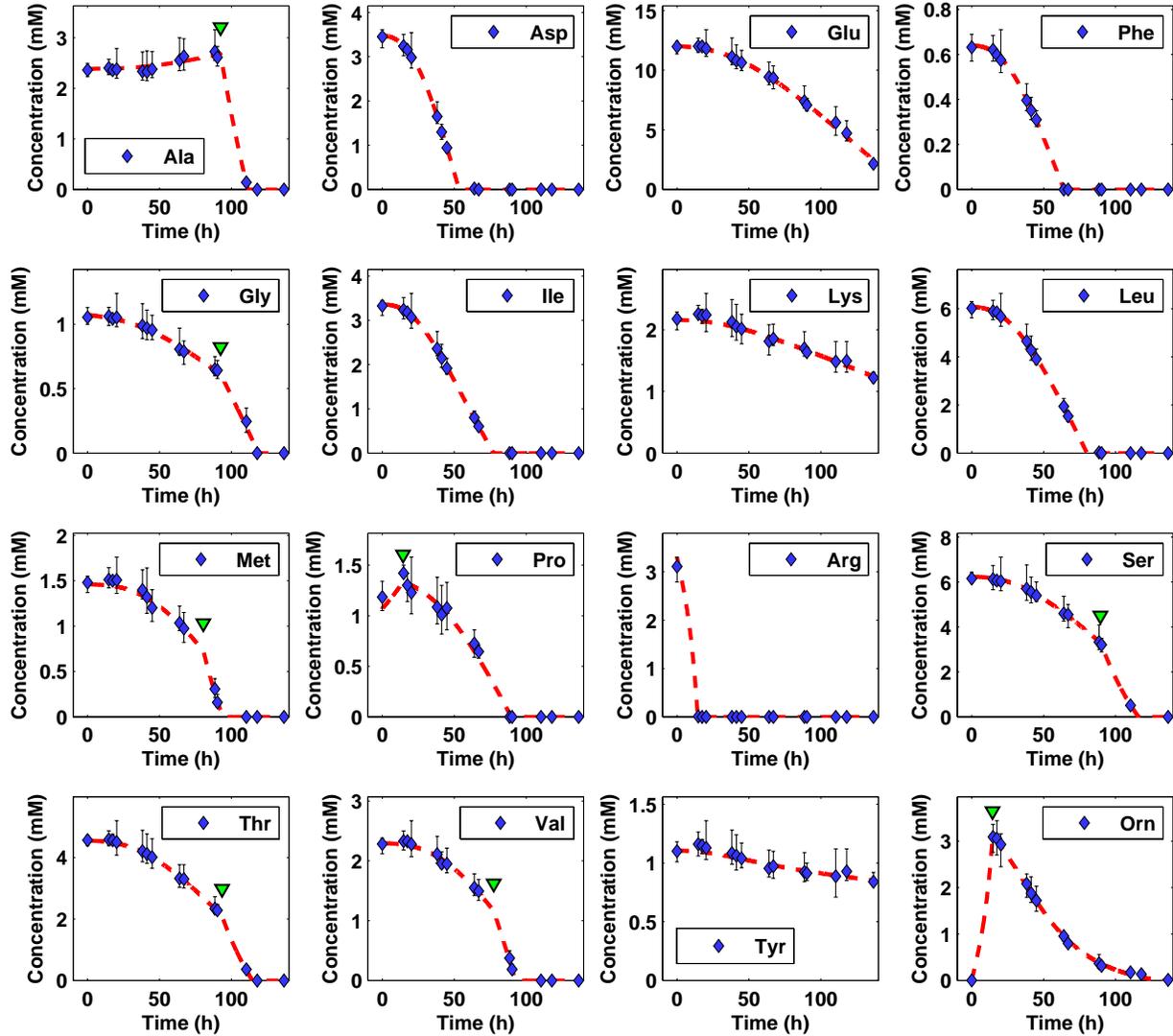
Figure 4.6: Nutrient consumption and production data from aerobically grown cells. Experimental data are shown using diamonds (average) with error bars provided. Model simulations are illustrated using red broken curves. The transport patterns of several metabolites qualitatively change during growth. For example, alanine and ornithine switch from production to consumption. For such metabolites, the piecewise Equation (4.17) was used to model utilization, and the corresponding $t_{i,b}$ parameter value, which represents a point near where the qualitative change occurs, is indicated with an inverted green triangle.

errors dropped by 98.6%, 92.4% and 54.5%, respectively. This was an indication that the corresponding transport processes exhibit distinct modes. Indeed, transport of ornithine, alanine and proline each exhibited a switch from production to consumption. Again, we stress that we did not decide *a priori* to use the piecewise over the basic form of the transport equation for any metabolite. Rather, the choices were made based on clear indications of improvement. Therefore, not only can our methods account for changing transport modes, but can actually lead to their identification or verification.

For the transport of a metabolite where the form of Equation 4.17 is clearly superior, the boundary parameter $t_{i,b}$ indicates an approximate time for when the qualitative change in the process occured. Similar to the use of the piecewise equation, the value of the parameter was not set *a priori* but was optimized using data. In this respect, our methods actually allow for the computational approximation of the critical points. The question therefore is whether the values we obtained make biological sense. To a large extent, we believe that the answer is yes. With respect to ornithine and proline, it is remarkable that their $t_{i,b}$ values, which represent the points at which they switch from production to consumption, are close to each other, and that both are near the time when arginine is depleted. This makes sense because arginine is consumed rapidly by the cells, and both ornithine and proline are downstream of its catabolic route. Indeed, the rapid depletion of arginine simultaneously occurring with the rapid accumulation of ornithine in the medium is a clear indication of a very active arginine fermentation pathway (see Section 4.8.2). Clearly, the $t_{i,b}$ parameters we obtained, at least in this case, are not merely mathematically convenient, but in point of fact correspond to, or reproduce, actual biological processes. Thus, as evidenced by this example, not only do our methods identify the presence of distinct modes in the transport processes but can actually lead to the elucidation of the reasons behind.

Valine uptake accelerates at approximately $t = 77$ h, near the time when isoleucine is depleted. Similarly, the uptake rate of methionine accelerates at about $t = 80$ h, near the time when leucine is depleted. While it is hard to conclude with certainty, given the resolution and quality of the current data, that the acceleration of valine uptake actually precedes that of methionine, it is very likely that the increased consumption rates for both are the result of cells compensating for the depletion of leucine and isoleucine. Note that the two branched-chain amino acids that were depleted seem to be preferred metabolites of *Halobacterium salinarum* under aerobic conditions. Accordingly, cells would have had to switch their metabolism to utilize more of the substitute sources in order to sustain their growth. Alternative explanations are provided in Section 4.5.

The other supplied metabolites with utilization patterns that we found to exhibit distinct modes are alanine, serine, threonine and glycine. The first switches from gradual production to rapid consumption, and the latter three demonstrate uptake rates that significantly accelerate. In each of these cases, the critical point seems to be near $t = 92$ h, which is also near the time when proline, methionine and valine are depleted. Although we believe that the increased consumption rates are also compensatory measures for the depleted metabolites, it is again difficult to conclude this with certainty given the current data. Moreover, if true, it is also difficult to determine whether which nutrients serve to compensate for what.

4.5   Validation of the Critical Points

The consumption and/or production of a metabolite with a transport pattern that qualitatively changes during growth, i.e., exhibits distinct modes, was modeled using Equation 4.17. The $t_{i,b}$ parameter of this equation carries significant qualitative interpretation because it indicates a point near where the qualitative change occurs. Its value is therefore crucial for identifying possible biological reasons for the presence of the distinct modes. For this reason, we performed further steps to investigate how well-defined the values we obtained are given the available data. Specifically, for each appropriate metabolite X_i, we fixed $t_{i,b}$ to specific values, and optimized for the best residual error that can be achieved leaving all the other parameters free. As with before, we used systematically defined initial guesses (for the other parameters). The results are summarized in Figure 4.7. In the figure, we plot for each metabolite the best residual error value as a function of $t_{i,b}$ (red broken curves). The values inidicated are normalized to the overall minimum, and therefore good values for $t_{i,b}$ are points at which the curve is near 1.0 (low).

Through manual inspection we found that most of the $t_{i,b}$ parameter values fall at points that are near the time of depletion of at least one other supplied metabolite. For example, the $t_{i,b}$ values initially obtained for ornithine and proline, where both metabolites switch from production to consumption, are near the point at which arginine is depleted. Similarly, the $t_{i,b}$ value of valine is near the depletion point of isoleucine. This can readily be seen by comparing Figures 2 and 4.7. In such cases, we manually synchronized the events by adjusting the $t_{i,b}$'s. The final values we used in our model are indicated by green, inverted triangles in Figure 4.7.

It is clear that a global $t_{i,b}$ parameter value is not possible because of metabolites
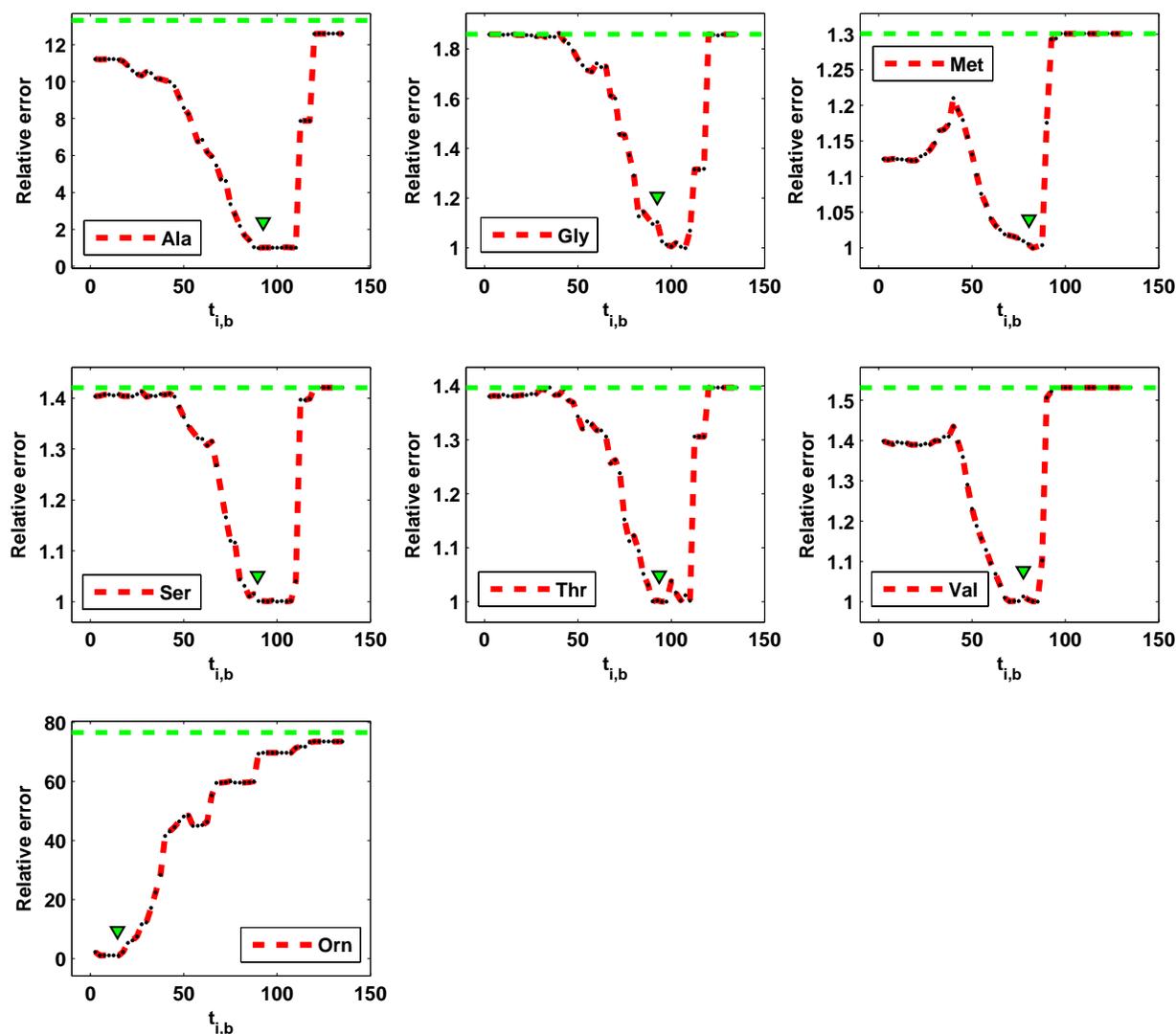
Figure 4.7: Parameter ($t_{i,b}$) exploration for aerobically grown cells. The uptake patterns of some supplied metabolites qualitatively change during growth. For such nutrients, the points at which the changes occur are represented by the $t_{i,b}$ parameter of Equation 2. The above graphs show for each metabolite the best residual error we were able to obtain for specific values of $t_{i,b}$ normalized to the overall minimum (red broken curve). Accordingly, ideal values for $t_{i,b}$ should be points where the curve is near 1.0 (lowest point). The actual parameter values used in the final model are indicated by green, inverted triangles. For comparison, we also indicate the best residual error value we were able to obtain using the non-piecewise equation form (green horizontal line).

like ornithine that plainly switch mode before the rest. Nevertheless, arguments relating to model simplicity may be made, given the current data, for using a common $t_{i,b}$ value for all metabolites that change modes later in the growth period, particularly alanine, glycine, methionine, serine, threonine and valine. Such synchronization could, for example, be the result of a global (metabolic) shift that occurred at that time, which induced the observed mode changes in the metabolites. While such an interpretation remains a possibility, it is inferior to the current interpretation that most of the mode changes are simply compensatory measures for the depletion of other nutrients in a number of ways. The reasons include: (1) From Figure 4.7, it can be seen that there is no single value for which the consumption/production models of the relevant nutrients are all at their best, i.e., each is at a point where the residual error is lowest. (2) Also from Figure 4.7, it is difficult to reconcile the idea that the qualitative changes occur nearly simultaneously with the observation that the curves of the relevant metabolites are quite diverse in their shapes; for example, some curves reach their lowest point (1.0) a lot earlier than others, and also begin rising earlier than others. If the events were indeed synchronized, then one would expect the curves to be more similar than they are. And (3) non-synchronized $t_{i,b}$ values allow for a more proportional total material uptake throughout growth. In contrast, forcing a single common value, which should be taken from a very narrow window around $t = 87.5$ h (Figure 4.7), would mean that the total material uptake of cells from the time when major metabolites (eg., leucine and isoleucine) are depleted up until the critical point will be relatively very low compared to the rest of the growth period. Moreover, given that all of the relevant mode changes are described by either an acceleration in uptake rate or a switch from production to consumption, then the total material uptake starting from the critical point on will be exceedingly high. Indeed, for methionine and valine, an acceleration that takes place as late as $t \approx 87.5$ would imply an incredibly high uptake rate because both should already be depleted by about $t = 92$. We should note that an abrupt increase in total material uptake is not something that is to be expected of a culture that is slowing down with respect to growth (late log phase). Finally, nutrient substitution has been described in bacteria and archaea. For example, *E. coli* has been shown to start consumption of acetate once the glucose supply is depleted (Varma and Palsson, 1994; Mahadevan *et al.*, 2002).

4.6   Carbon Fates

Nutrients consumed by cells can have various fates; they can be: (1) incorporated into the biomass with at most very minimal changes, such as in the case of amino acids incorporated as protein residues or free metabolites; (2) converted to other biomass components, typically after partial degradation; (3) oxidatively degraded to CO_2 for the production of energy through respiration; and (4) completely or partially degraded, for example down to the level of small metabolites, and then secreted. Given that the rates at which most of the supplied amino acids were consumed far exceeded the requirements for biomass incorporation as described in the fate (1) mentioned above, then at least one of the other possible fates must also happen to the relevant amino acids. The results for individual metabolites are summarized in Figure 4.8, and the comparison between total (global) carbon uptake and total carbon incorporation (fates 1 and 2) is provided in Figure 4.9.

4.6.1   Degradation of Essential Amino Acids

Surprisingly, leucine, isoleucine, valine and methionine are among the nutrients that are mostly degraded (see Figure 4.8) by *Halobacterium salinarum*, whether for energy or as carbon skeleton donors. This was unexpected because the four are, along with lysine, essential amino acids for the archaeon. That is, the microorganism cannot synthesize these metabolites, and as such cannot grow if they are not exogenously supplied. For this reason, one would typically expect that it is better for cells to use essential nutrients sparingly, for example in accordance only with the requirements for biomass production, especially when there are other non-essential sources of energy and carbon available. However, as discussed earlier, leucine and isoleucine are even among the preferred metabolites of *Halobacterium salinarum* (Section 4.4). The heavy rates at which the halophile catabolizes the essential amino acids can cause it to prematurely terminate its own growth.

By $t = 140$ h, which is the boundary of our simulations, only about 7%, 8% and 4% of the amounts of leucine, isoleucine, and methionine, respectively, that have been consumed have been directly incorporated into the biomass (fate 1). This means that in each case, over 90% of the material that was taken up was used for other purposes, which very likely involved at least partial degradation. While the exhibited behavior is initially counterintuitive, we believe that it is actually an adaptation of the microorganism to its environments, which are characterized not only by extreme salinity but also by eratic nutritional availability. This is also supported by additional findings (see the rest of the
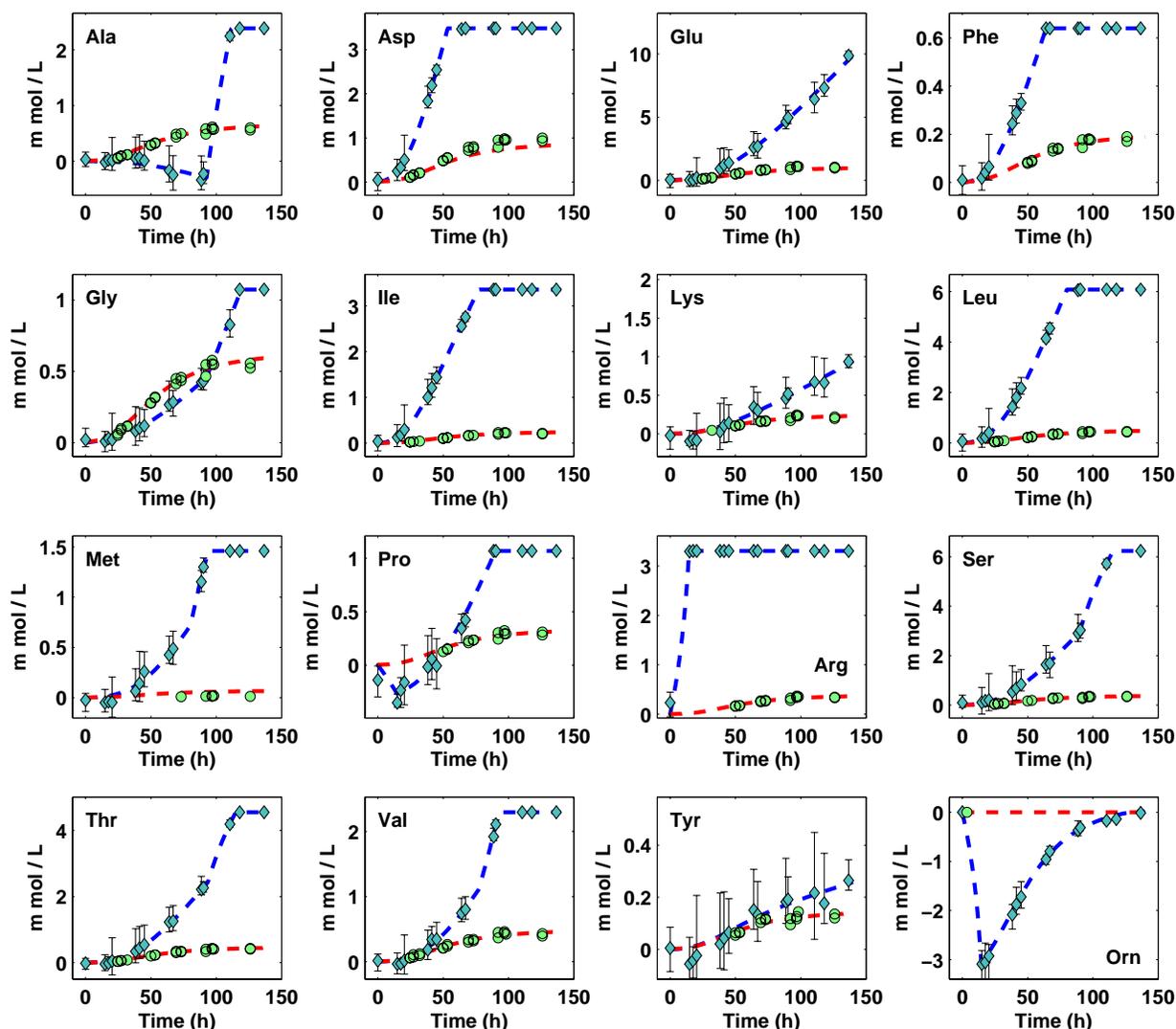
Figure 4.8: Summary of nutrient uptake and incorporation rates under aerobic conditions. The blue curves indicate the net amount of each amino acid that has been consumed (positive) or produced (negative) as a function of time. The red curves, on the other hand, show the total amount of each that has been incoporated into the biomass, whether integrated into proteins or as free metabolites. Clearly, most of the supplied amino acids are taken up from the medium at rates that far exceed the speeds at which they are directly incorporated. This implies that the pertinent amino acids are, in addition, significantly catabolized for energy, and/or are used to synthesize the other biomass constituents that are not supplied.
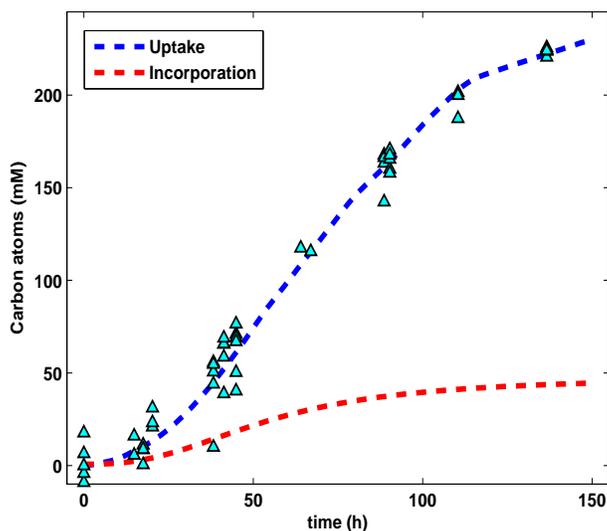
Figure 4.9: Total carbon consumption and biomass incorporation of aerobically grown cells. The red curve shows the total amount of carbon that has been incorporated into the biomass (computed using the biomass reaction definition and the time-dependent population size). On the other hand, the blue curve indicates the total amount of carbon that has disappeared from the medium, calculated using the supplied nutrients and ornithine (filled triangles are experimental data). At $t = 140$, only about 19% of the total carbon that has been consumed can be accounted for by the biomass.

chapter and the next), and will be discussed in more detail in Section 5.3.

4.6.2 Investigating By-product Secretion

The four possible nutrient fates mentioned above can be grouped into two general categories; biomass incoporation (fates 1 and 2) and metabolic by-product secretion (fates 3 and 4). Only fate 1 is directly measurable using the experimental methods used in this study. Nevertheless, with the approximation of the biomass composition in hand (Section 4.3), which was obtained using measurements, calculations and assumptions, analysis of the other fates becomes possible. In particular, we correlated the nutrient utilization data (Figure 4.6) with the growth function and the observed growth rates using flux balance analysis.

We repeated the determination of the amino acid content of the biomass for the specific cultures represented in Figure 4.6. The results are summarized in Figure 4.10. While the relationships between the amino acid concentrations and the optical density are not strictly linear, average errors from linear fits are below 12.5% in all cases. Most of the
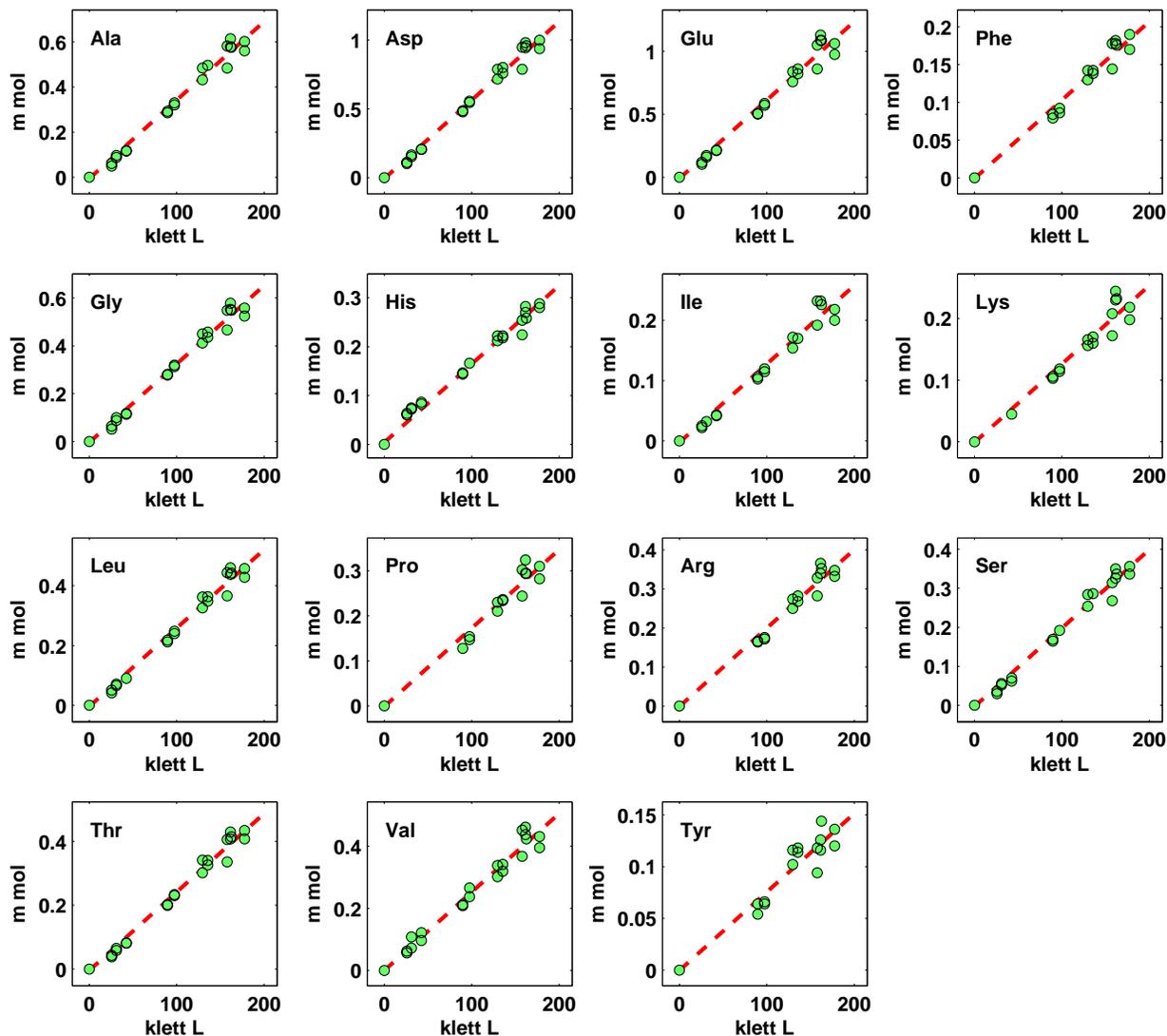
Figure 4.10: Amino acid composition of biomass during aerobic growth. We repeated biomass amino acid content determination for the specific cultures represented in Figure 4.6. The amount of each was quantified at different optical densities. The measurements include those that are integrated into proteins (residues) and those that are free inside the cells. The red line in each graph indicates the best linear fit for the particular molecule. Average deviation from the line in all cases is below 12.5%. Due to experimental limitations, we were not able to obtain values for cysteine and tryptophan, and the values for aspartate and glutamate are already inclusive of asparagine and glutamine, respectively.

deviations are situated in the lower optical density range. With respect to the previous set of measurements (Section 4.3.1), the most significant difference is that the current results revealed a total amino acid content that is about 30% higher at approximately 503 μg/OD·ml. Possible factors that could have contributed to this discrepancy include differences in the conditions used, including media composition, and the that fact we more rigorously adapted the current cultures. The discrepancy is well within the range of inaccuracy that may be introduced by morphological changes in the cells to the relationship between biomass and optical density. Accordingly, we also adjusted the biomass requirement (composition) of the other constituents for the rest of the analysis described in this chapter.

CO_2 is an expected metabolic end-product of cells under aerobic conditions. Indeed, if it is assumed that all carbon atoms that are taken up and do not appear in the biomass are completely degraded through respiration, then CO_2 should account for the bulk of the by-product pool. We calculated the (theoretical) amount of oxygen that would have been needed in our cultures under such a scenario at various points during growth, using flux balance analysis. However, measurements of actual oxygen consumption showed that cells used only about 20% of the oxygen that they would have otherwise needed to completely oxidize all the material. Therefore, respiratory-linked degradation was not the fate of most of the carbon atoms that were consumed but did not get incorporated into the biomass. Moreover, significant quantities of (partially oxidized) by-products must have been formed and accumulated in the medium, including very likely those that result from forms of overflow metabolism. The results also show that the described situation already seems to be the case even before the oxygen supply became limiting (about $t = 30$ h). The computations are summarized in Figure 4.11. Note that the optimality assumptions of flux balance analysis have little to do with the predicted high by-product secretion rates. Rather, the optimality principles only affect the ratios between the by-products themselves; for example, what percent is secreted as CO_2 and what percent is secreted as partially degraded metabolites. The latter type is likely because the repiratory rates exhibited by the cells are clearly not sufficient to handle all the materials they consumed. Further flux analyses show that the most efficient by-products with respect to energy production are common intermediates such as acetate and succinate.

We observed a maximal respiratory rate of approximately 10 nmol O_2/klett·ml·hr during growth. At about $t = 30$ h, the amount of dissolved oxygen in the medium was reduced to 0%. Given that prior to this point we could already observe large amounts of nutrients
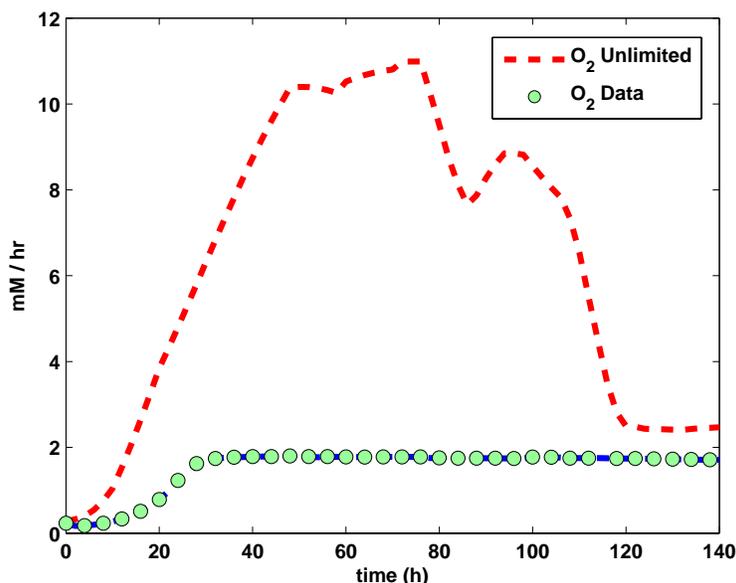
Figure 4.11: Theoretical (energy-optimal) and actual oxygen consumption rates. The red broken curve indicates the amount of oxygen that would have been needed if all nutrients that were taken up but did not get incorporated into the biomass were used for respiration. The blue broken curve shows the actual oxygen utilization rate. The large discrepancy between the two curves suggest that overflow metabolism is prevalent. The oxygen supply starts to become limiting at about $t = 30$. The smaller peaks and troughs of the red curve are due to the discontinuities of the piecewise models used for the uptake of several metabolites.

that were consumed but neither incorporated into the biomass nor subjected to oxidative phosphorylation, then the respiratory process itself was likely the bottleneck prior to it. After about $t = 30$ h however, cells could only respire in at most the rate at which oxygen dissolved into the medium. Note that a 0% oxygen saturation level does not mean that oxygen is no longer available to the cells because flasks were kept open. Since growth continued well past the point when the oxygen supply started to become limiting, the rate of respiration dropped steadily from then on.

4.7   Bioenergetics

The bioenergetic fexibility of *Halobacterium salinarum* (Section 3.1.2) makes it an ideal model organism for the integrated analysis of different energy producing mechanisms which are typically investigated separately. In particular, under the aerobic conditions that we used, cells could derive energy (ATP) through respiration and substrate level phosphory-lation (including arginine fermentation). We calculated the maximum (theoretical) energy

that the system can produce at various points during growth, using flux balance analysis subject to the nutrient utilization (Figure 4.6) and oxgen consumption data, and evaluated the result with respect to the contribution of respiration. The values are indicated in Figure 4.12 by the red and blue curves, respectively. Specifically, we calculated the energy produced by respiration (blue curve) using the observed oxygen consumption rates and an ATP to O_2 ratio of approximately 1:1 (Hartmann *et al.*, 1977). On the other hand, the system maximum energy (red curve) was calculated as

$$E_{max}(t) = v_{ATP}(t) + \mu \dot{\gamma}(t) \tag{4.27}$$

where $\dot{\gamma}(t)$ is the current growth rate, and μ is the growth-related energy that is implicitly taken into account by the metabolic network in synthesizing the compounds included in the biomass (Teusink *et al.*, 2006). In order to find flux distributions that maximize energy production, we introduced an ATP hydrolysis reaction ($ATP + H_2O = ADP + P$) into the network, and performed flux balance analysis with it defined as the objective function. $v_{ATP}(t)$ in Equation (4.27) refers to the flux through this reaction.
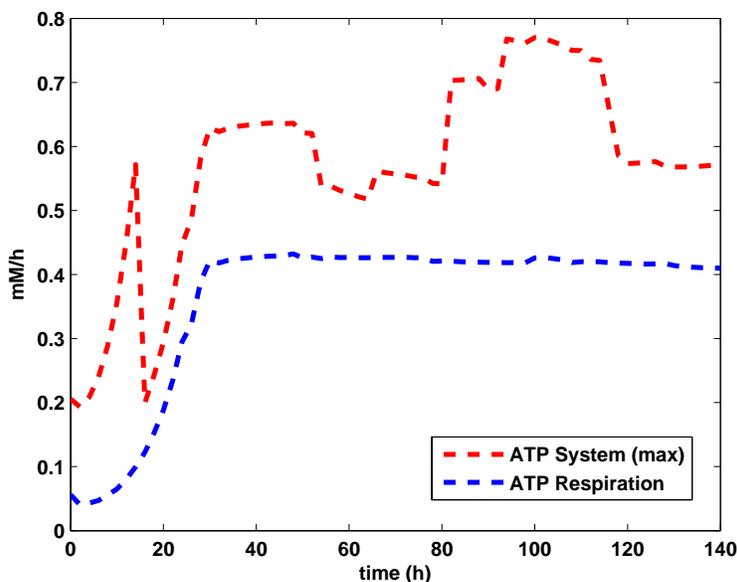
Figure 4.12: System energy in ATP equivalent. The red curve shows the maximum amount of energy, computed using FBA, that the system can produce at various points during growth. The blue curve indicates the amount of energy that is produced through respiration. Interestingly, prior to the depletion of arginine ($t \approx 15$), fermentation of the amino acid generates most of the energy in the system, even more than respiration.

From Figure 4.12, it can be observed that the theoretical maximum amount of energy

that the system can produce (red curve in Figure 9) is significantly higher than what is generated by respiration (blue curve). Although it may be the case that the actual energy production of the system is lower than what the theoretical values indicate, it is clear that at the early stages of growth (before approximately $t = 15$), very large amounts of energy are generated that cannot be attributed to respiration. Inspection of the model revealed that this is due to activity of the arginine fermentation pathway during that period. This finding is quite significant because arginine fermentation is normally assumed to play a secondary bioenergetic role, i.e., an alternative to the primary modes of respiration and photosynthesis. However, in contrast to the typical supposition, the fermentation process occurs simultaneously with respiration, and is moreover highly active. We should note that this conclusion, because of the constraints imposed by the observed rapid depletion and accumulation of arginine and ornithine, respectively, is not subject to the optimality assumptions used during FBA.

Our findings also indicate that arginine fermentation occurs at such high velocities that its energy output becomes comparable to that of the respiration, even though it is much less efficient in terms of ATP production per unit substrate. Indeed, even if we increased the efficiency of the ATP to O_2 ratio that we used, which was obtained from oxygen pulse experiments (Hartmann *et al.*, 1977) that were not corrected for possible competing processes, by a factor of 4 from 1:1 to 4:1, calculations would still indicate that arginine fermentation out-produces respiration in our cultures. This is consistent with the fact that *Halobacterium salinarum* can be grown with neither oxygen nor light just as readily, by supplying large amounts of the amino acid.

After arginine was depleted at about $t = 15$ h (Figure 4.12), respiration accounted for most of the energy produced in the system. The additional energy that can be produced through the further non-respiratory-related degradation of other (non-arginine) nutrients is significantly smaller. This being the case, and in connection with the observations made in Section 4.6, it is remarkable that non-respiratory-related processes account for approximately 80% of the nutrients that were consumed but not incorporated into the biomass. Again, the pertinent metabolites include essential ones. Unfortunately, the nature and purpose of the catabolic processes remain unlcear.

4.8   Fluxome Prediction

The determination of internal metabolic fluxes is very important because they describe at the molecular level how an organism lives and grows. Unfortunately, in sharp contrast to other 'omics' technologies, fluxome analysis has remained a laborious endeavor (Sauer, 2004). This is especially true for organisms such as *Halobacterium salinarum*, where carbon labeling, the experimental basis for flux analysis, is complicated by the necessary presence of multiple substrates in the medium. Accordingly, genome-scale measurements of fluxes for the archaeon are still very difficult and impractical. In this study, we used our constraints-based model to predict values for the fluxes, or at least analyze their allowable ranges. Specifically, we used flux balance analysis to determine particular distributions, and performed flux variability analysis to calculate for each reaction its maxmimum and minimum flux values independent of any optimality assumption.

4.8.1   Variability Analysis

The fluxome illustrated in Figure 4.13 was calculated using the observed consumption and production rates during the log phase. In it, compounds that are taken up from the medium are represented by yellow ellipses, and compounds which are accumulated in the medium by red ellipses. Fluxes through reactions are represented by arrows, and thickness is used to indicate strength. Considering that it is uncertain to what extent energy production, or any of the optimality principles used in FBA, is actually the objective of cells *in vivo*, we used flux variability analysis (Mahadevan and Schilling, 2003) to complement the results by computing the minimum and maximum possible fluxes through each reaction after removing all optimality assumptions. Fluxes through reactions for which the values are either both positive or both negative are drawn in black. Intuitively, these are reactions that are constrained enough by the network stucture, pseudo-stability and the observed exchange fluxes, to have non-zero activity regardless of any optimality assumption. That is, the network structure and the measured data are already enough to guarantee the existence of the fluxes and their indicated directions. We refer to these fluxes as qualitatively constrained.

During the log phase ($t = 30$ h), we found 378 (67%) of the 567 internal reactions of the metabolic network to have qualitatively invariable fluxes. Of these, 246 carry non-zero fluxes, and 132 are blocked under the conditions. The fact that measurements of a relatively few input and output fluxes already qualitatively determine the fluxes through most of the
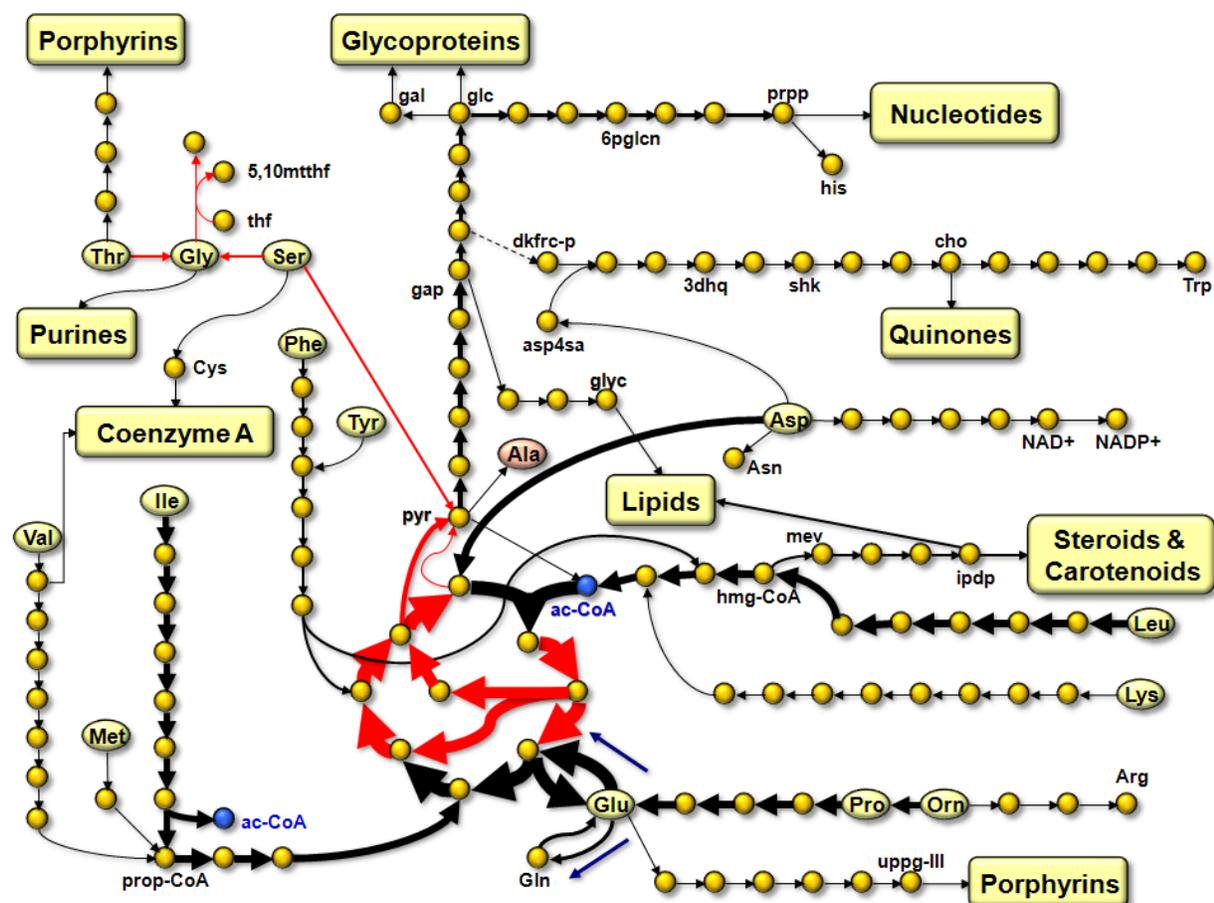
Figure 4.13: Predicted fluxome during log phase ($t=30$) that maximizes energy production. Yellow ellipses represent compounds that are taken up from the medium, while light-red ellipses represent those that are accumulated in the medium. Larger fluxes are drawn with thicker arrows. Black and red arrows correspond to qualitatively invariable and qualitatively variable (see main text) fluxes, respectively. Side reactants, for the most part, are not depicted in the figure. Yellow rounded boxes represent biochemical pathways.

Metabolite abbreviations: 3dhq - 3-dehydroquinate, 5,10mtthf - 5,10-methylenetetrahydrofolate, 6pglcn - 6-phosphogluconate, ac-CoA - acetyl-CoA, Ala - alanine, Asn - asparagine, Asp - aspartate, asp4sa - aspartate 4-semialdehyde, Arg - arginine, Cys - cysteine, dkfrcp - 6-deoxy-5-ketofructose 1-phosphate, gal - galactose, gap - glyceraldehyde 3-phosphate, cho - chorismate, glc - glucose, Gly - glycine, glyc - glycerol, Glu - glutamate, Gln - glutamine, His - histidine, hmg-CoA - 3-hydroxy-3-methyl-glutaryl-CoA, ipdp - isopenteny diphosphate, Ile - isolecuine, Leu - leucine, Lys - lysine, Met - methionine, mev - mevalonate, Orn - ornithine, Phe - phenylalanine, Pro - proline, prop-CoA - propanoyl-CoA, prpp - 5-phosphoribosyl diphosphate, pyr - pyruvate, Ser - serine, shk - shikimate, thf - tetrahydrofolate, Thr - threonine, Trp - tryptophan, Tyr - tyrosine, uppg-III - uroporphyrinogen III, Val - valine.

reactions is related to the bow tie structure of the metabolic network (Gonzalez *et al.*, 2008; Csete and Doyle, 2004). Catabolic pathways, which together form one fan of the bow tie, fan-in into the knot of central metabolism to supply the common intermediates. These pathways are highly linear and convergent in structure. Accordingly, nutrients which are taken up and not directly incorporated into the biomass will have to go through the associated catabolic pathway with no variability until the level of the common intermediates is reached. Similarly, biosynthetic pathways, which form the other fan of the bow tie, fan-out and branch from the knot of central metabolism, and are also highly linear in structure. Thus, most biosynthetic routes of cellular constituents will also be invariable once past the level of the metabolic core.

4.8.2 Nutrient Utilization

When an organism is grown on defined substrates, it is often of interest to know which pathways are active, and for what processes and to what extents the supplied substrates are utilized. This information can be used to guide labeling studies as well as to interpret results from experiments using similar conditions, such as transcript and protein expression data. As a simple example, given that the precise mechanisms for lipid formation in *Halobacterium salinarum* is still unclear, it will be of interest to maximize the flow of central metabolites into lipid formation to facilitate labeling experiments. From the fluxome (Figure 4.13), it is clear that the high catabolic fluxes from leucine and and lysine will have to be reduced, such as by using mutants deficient in genes associated with the pertinent pathways. In this section, we describe how the supplied nutrients are utilized under the aerobic conditions, and relate them with the current knowledge in the literature.

About 90% of the leucine taken up by cells is converted to hydroxymethylglutaryl-CoA (hmg-CoA) during the log phase. In turn, most of the hmg-CoA enter the TCA cycle through acetoacetate as acetyl-CoA, and some are converted to mevalonate (EC 1.1.1.34) on the way to the synthesis of archaeal isoprenoid lipids. Note that unlike higher organisms that synthesize their isoprenoid chains from acetate, halobacterial species at least in part synthesize their lipids from amino acids (Ekiel *et al.*, 1986). While no gene could be assigned for the enzyme hmg-CoA lyase (EC 4.1.3.4) in *Halobacterium salinarum*, simulations indicated that the biochemical function must be present as it is needed to channel the substantial quantities of leucine molecules that are not directly integrated into the biomass. One candidate is the gene annotated as hmg-CoA synthase (EC 2.3.3.10), which classically catalyzes a reaction that proceeds in the other direction. Alternatively,

it is also possible that the organism uses a mechanism that is currently uncharacterized, or that it is using a non-homologous gene.

Approximately 93% of isoleucine is degraded and enters the TCA cycle at two entry points; C2 units enter as acetyl-CoA, and the final C3 unit, propionate, is converted to succinyl-CoA. Since the consumption rates of leucine and isoleucine exhibited among the greatest reductions during phototrophic growth (Chapter 5) compared to aerobic growth, it is likely that the two are the main respiratory substrates of *Halobacterium salinarum* under the conditions used.

In contrast to the other two branched-chain amino acids, valine exhibited relatively very little uptake activity during the log phase. However, consumption of the amino acid greatly accelerated when isoleucine was depleted. Considering this and the fact that isoleucine and valine share the catabolic route through propionate, it would seem that the latter is used by the archaeon as an alternative to the former. Valine is in addition used to synthesize coenzyme-A (minor flux).

While it is difficult to interpret the individual fluxes into central metabolism from glycine, serine and threonine due to interconversions among them, simulations showed that the combined flux is about 53% of their total uptake. The scheme depicted in Figure 4.13 represents an optimal configuration where threonine and serine, at least in part, are converted to glycine, which is then cleaved to NH_3 and CO_2. Prior to the depletion of threonine and serine, the uptake rate of glycine was below the requirement for direct biomass incorporation of the amino acid. This does not even consider the fact that glycine also serves as the methyl donor, via tetrahydrofolate, in a number of biosynthetic reactions. However, after serine or threonine was depleted, or both since the resolution of the data does not allow us to differentiate between the two events, glycine uptake accelerated considerably. This is consistent with the idea that threonine and serine were partially converted to the amino acid in order to shoulder some of the requirements for it.

Aspartate exhibited the highest consumption rate next to arginine. Most of it likely enters the TCA cycle as oxaloacetate, although, in addition, it is also: (1) used to synthesize asparagine (EC 6.3.5.4); (2) oxidated to iminoaspartate (EC 1.4.3.16) on the way to the synthesis of NAD+ (minor flux); (3) used to synthesize the probable aromatic amino acid precursor aspartate-4-semialdehyde (Section 3.7.3); and (4) is used in the metabolism of nucleotides.

None of the key degradative enzymes for phenylalanine and tyrosine could be identified in *Halobacterium salinarum* during the initial stages of reconstruction. However, similar

to the case for hmg-CoA lyase (EC 4.1.3.4), simulations indicated that the archaeon likely has the capability to degrade both amino acids. The uptake rate of each exceeded the corresponding incorporation rate (Section 4.4), more notably for phenylalanine. Accordingly, it is likely that the archaeon either uses non-homolgous enzymes or employs as of yet uncharacterized pathways.

Analogous to the situations with phenylalanine and tyrosine, degradative enzymes could also not be identified for lysine even though uptake of the amino acid significantly exceeded the requirement for biomass incorporation. Therefore, a catabolic pathway for the amino acid is also likely. Indeed, it was observed in a previous study that labeled carbon atoms from lysine turn up in isoprenoid molecules (Ekiel *et al.*, 1986). Based on the observed labeling pattern, it is likely that the pathway goes through acetoacetate (or acetoacetyl-CoA).

We mentioned earlier that the $t_{i,b}$ parameters for ornithine and proline are remarkably close to each other, and that both occur near the time when arginine is depleted (Section 4.4). We also pointed out that the situation revealed by our methods made sense since both ornithine and proline are downstream of arginine's catabolic route. Here, we discuss the relationships between the three metabolites in more detail. Growth simulation at a point before arginine depletion shows the situation depicted in Figure 4.14. From the figure, it can be seen that arginine, which exhibited the highest uptake rate prior to its depletion, is mostly deaminated to citrulline via the action of arginine deiminase (EC 3.5.3.6). In turn, citrulline is converted to ornithine and carbamoyl-phosphate by ornithine carbamoyltransferase (EC 2.1.3.3). On one hand, ornithine is mainly transported outside cells through an arginine-ornithine antiporter ($\approx 95\%$). This means that most of the arginine is transported into the cells for "free". Small amounts of ornithine continue through the catabolic route that, via proline, eventually leads to α-ketoglutarate. On the other hand, carbamoyl-phosphate is mostly degraded to NH_3 and CO_2 in a reaction that produces ATP from ADP (EC 2.7.2.2). To summarize the situation, arginine is rapidly converted to ornithine through a process that results in the net production of ATP. The carbon skeleton of ornithine is not further degraded, but is is rather secreted. As was discussed earlier, arginine fermentation accounts for most of the energy in the cells at the early stages of growth (Section 4.7).

Ornithine was not one of the metabolites in the medium that were originally measured at the early iterations of our analysis of aerobic cultures. However, since we believed that arginine is essential for *Halobacterium salinarum* and we observed that cells continued to
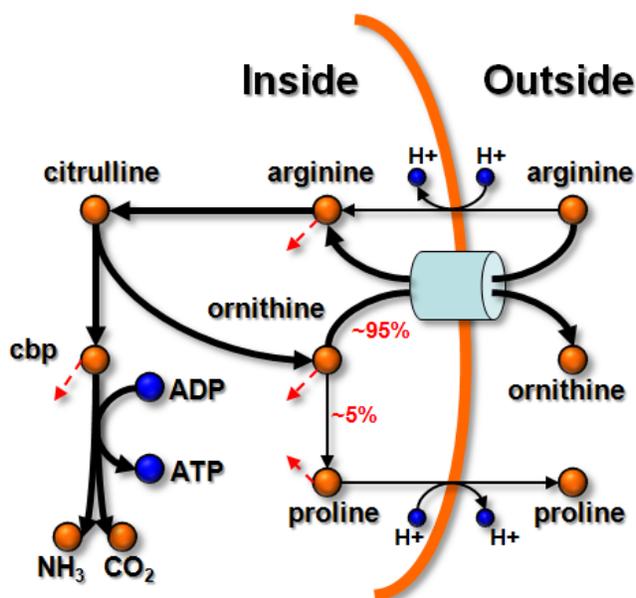
Figure 4.14: Arginine, proline and ornithine metabolism. Prior to its depletion, arginine exhibits the highest uptake rate among the supplied nutrients. Most of it is deaminated to citrulline, which is then converted to ornithine and carbamoyl-phosphate (cbp). Most of the ornithine ($\approx 95\%$) is transported outside through an arginine-ornithine antiporter. On the other hand, carbamoyl phosphate is primarily degraded to NH_3 and CO_2 in a reaction that produces ATP from ADP. Simulations show that at the beginning of growth this process produces most of the energy in the cells. The red broken arrows indicate connections to other parts of the network.

grow well past the point when the amino acid was depleted, we concluded that the archaeon was simply storing the carbon skeleton of arginine in a reusable form. Despite the fact that arginine fermentation is often assumed to be repressed by the primary bioenergetic modes of respiration and photosynthesis, our first intuition then was still that the reusable form is ornithine. This is the reason why we subsequently included it in the set of measured metabolites. As has been described above, our initial intuition proved to be correct as the succeeding results showed that ornithine is rapidly accumulated, and then gradually consumed once the arginine supply is exhausted. In addition, our proposed scheme also implied that ornithine should be able to substitute for arginine in the medium. Accordingly, we also confirmed this by growing *Halobacterium salinarum* in a medium were arginine is replaced by ornithine. However, unexpected results where obtained when, in the interest of completeness, we also tried to grow the archeaon with neither of the metabolites supplied, and got positive results. Details on this aspect are discussed in Section 3.7.1.

Members of *Halobacterium* (*salinarum, NRC-1, cutirubrum, halobium*) have long been reported not to grow on sugars. Indeed, active transport of glucose has been experimentally excluded (Severina *et al.*, 1991). Despite these findings however, the exogenous addition of glucose and galactose have been described to stimulate growth to some extent (Gochnauer and Kushner, 1969; Sonawat *et al.*, 1990). We believe that this is in part due to the significant biosynthetic burden that is alleviated from the cells. Note that the surface layer of the Halobacteria contain saccharide elements (Section 4.3). Under the specific set of conditions being analyzed, simulations show that the archaeon must be using a reversal

of the glycolytic pathway (gluconeogenesis) to fulfil its sugar needs. Pyruvate that enters gluconeogenesis during the log phase is produced either through the decarboxylation of malate or oxaloacetate (EC 1.1.1.40), or from serine (EC 4.3.1.19).

Chapter 5

Phototrophic Growth of *H. salinarum*

This chapter is a continuation of the analysis of *Halobacterium salinarum*'s growth. The main difference from the previous chapter is that the focus is on phototrophic conditions. This allows for a comparison between the two bioenergetic modes of respiration and photosynthesis. A discussion of the main findings for both conditions is provided in Section 5.3.

5.1 General Comparison with Aerobic Case

Phototrophic cultures were grown under generally similar conditions to the aerobic preparations, with the obvious exceptions of the absence of oxygen and the presence of illumination (white light). The flasks were first flushed with nitrogen in order to remove oxygen, and were then closed using air-tight septa. So as to maintain the anaerobic conditions of the cultures while taking samples, syringes with long needles inserted through the septa were used. We were able to observe final optical densities well above 150 klett (1.5 OD), comparable to the population levels achieved by their counterparts (see Figure 5.1). The most apparent difference is with respect to having slower growth rates. While aerobic preparations exhibited log phase doubling times of only roughly 11 hours, phototrophic cultures demonstrated times close to 50 hours. The reason for this disparity is unclear. Representative aerobic and phototrophic growth curves are depicted in Figure 5.1.

We repeated the procedure for the determination of biomass amino acid content (Section 4.3.1) on the phototrophic cultures (Figure 5.2). Measurements revealed that they in
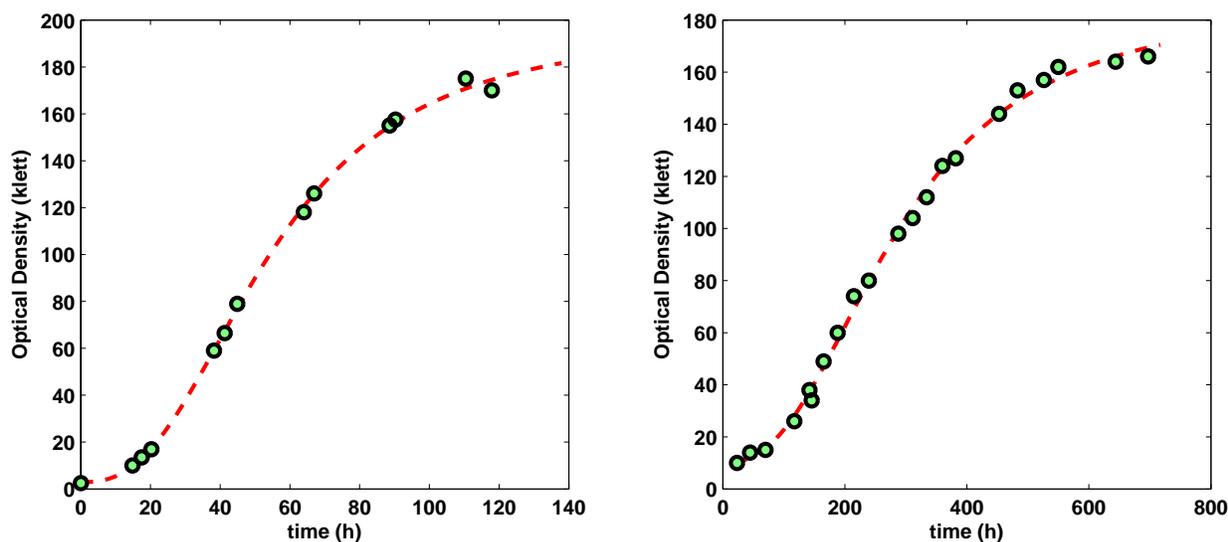
Figure 5.1: Representative growth curves. (Left:) Cells grown aerobically in the dark. A doubling time of approximately 11 hours was observed. **(Right:)** Cells grown anaerobically in light. The culture exhibited a much slower growth rate with a doubling time of about 50 hours. Please take note of the very different time scales between the two graphs.

general have approximately 35% less of each amino acid per klett·ml compared to their aerobic counterparts (Figure 5.3, top graph). However, this does not necessarily mean that the content per cell is really lower. The discrepancies can very well be due to morphological differences between cells under the two conditions, which could have affected the relationship between optical density and cell count. In contrast, the relative contribution of each amino acid to the total is remarkably constant between the two conditions (Figure 5.3, bottom graph).

5.2   Metabolite Consumption and Production

We modeled the consumption and production of metabolites during phototrophy using the same procedure employed in the aerobic case (Section 4.2.1). Again, the $t_{i,b}$ parameters, when applicable, reflect the points at which the qualitative changes in the transport processes occur. The experimental data and the corresponding model simulations are illustrated in Figure 5.4.

Analogous to the aerobic case, arginine was already depleted near the start of growth with most of it converted to ornithine and secreted by cells into the medium. This is interesting because arginine fermentation genes (*arcABC*) have been found to be repressed
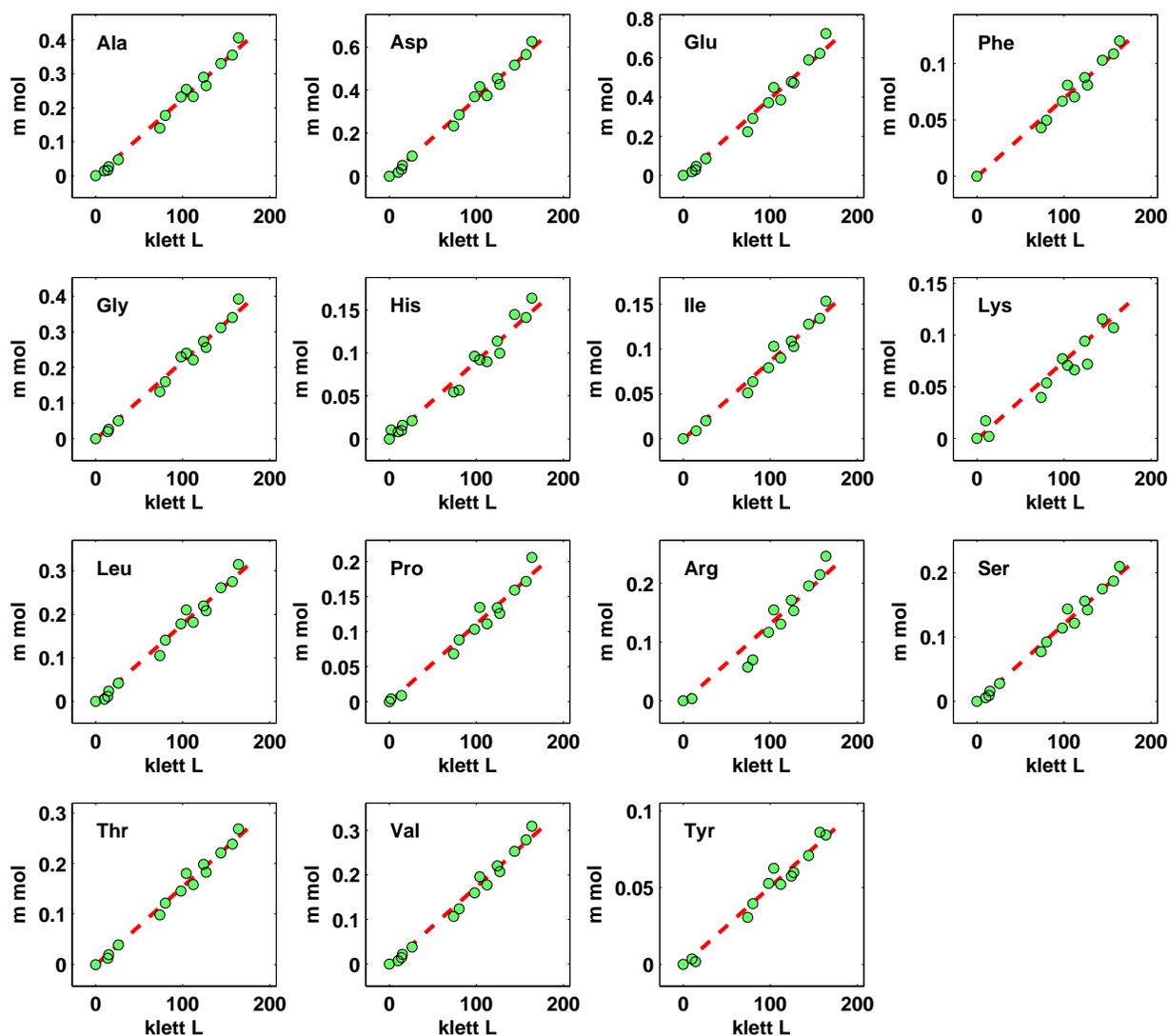
Figure 5.2: Amino acid composition of biomass during phototrophic growth. The amount of each amino acid was measured at different optical densities. The values indicated are inclusive of those that are integrated into proteins (residues) and those that are free inside cells. The red lines indicate the best linear fits. Due to experimental limitations, we were not able to obtain values for cysteine and tryptophan, and the values for aspartate and glutamate are already inclusive of asparagine and glutamine, respectively.
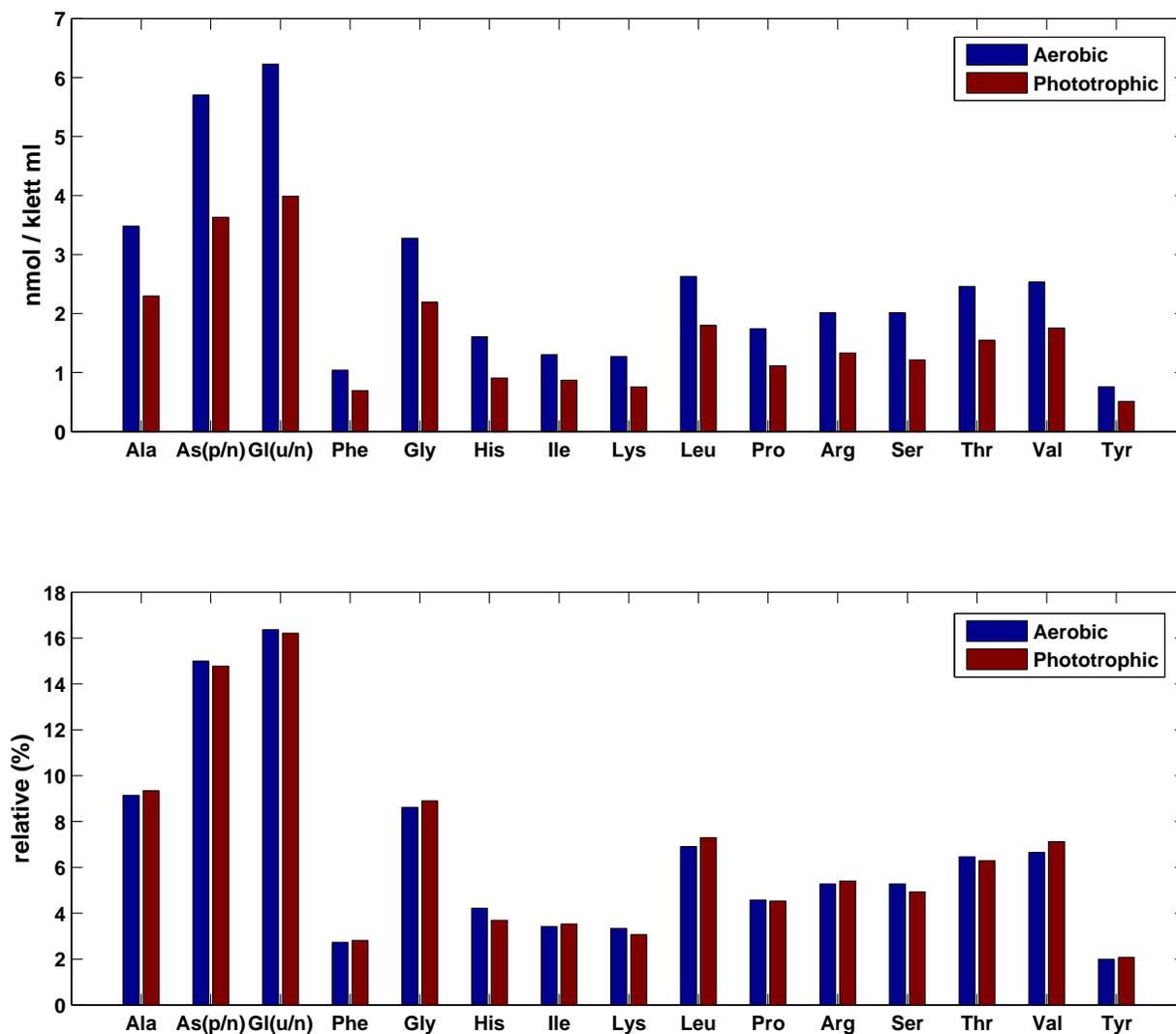
Figure 5.3: Comparison of the biomass amino acid composition between aerobic and phototrophic cultures. The top and bottom graphs show absolute and relative values, respectively. Phototrophic cultures in general have approximately 35% less of each amino acid per klett·ml. However, this does not necessarily mean that the content per cell is really lower. The discrepancies can very well be due to morphological differences between the cells under the two conditions, which could have affected the relationship between optical density and cell count. The relative contribution of each amino acid to the total is remarkably constant between the two conditions (bottom graph). However, it is worth noting that hydrophobic amino acids seem to be slightly overrepresented under phototrophic conditions, and hydrophilic amino acids slightly underrpresented. This is indicative of a higher expression ratio between membrane and cytosolic proteins.

during phototrophy, presumably to repress the secondary energy source (Pan *et al.*, 2002). However, the rapid depletion of the amino acid and the equally rapid accumulation of ornithine is a clear indication that the fermentation process was in fact active even under the phototrophic condition. Considering that we used inoculants taken from cultures already growing under phototrophy, it is unlikely that the observations are only due to the time (lag) that is required for regulation to take effect.

Next to arginine, aspartate exhibited the highest uptake rate, and is the second supplied amino acid that is depleted at about $t = 200$ h. Up to that point, very little consumption of glutamate could be observed, if at all. However, after the exogenous supply of aspartate was exhausted, consumption of glutamate either started or accelerated appreciably. Therefore, it would seem that glutamate is used as a substitute for aspartate under the conditions. Note that both amino acids are just one step away from the TCA cycle. Methionine consumption also accelerates at about the same point.

Some 200 hours later ($t \approx 400$ h), phenylalanine is depleted, and threonine and serine are nearly exhausted. In addition, alanine switches from production to consumption, and the uptake of glycine accelerates. As pointed out earlier (Section 4.8.2), the three amino acids glycine, serine and threonine are connected to each other through their degradative pathways. Presumably, when glycine is no longer produced through the degradation of excess amounts of either serine or threonine, then cells will have to transport more of it. This is consistent with the observation under both aerobic and anaerobic conditions that prior to the depletion of serine or threonine or both, glycine consumption is significantly lower than the rate at which it is incorporated into the biomass, not even considering its role as the methyl donor in various biosynthetic pathways.

The depletion of phenylalanine coincided with a number of discernable qualitative changes in the transport pattern of other metabolites, unlike in the aerobic case. In addition to the ones stated above, consumption of the amino acid serine seems to have stopped at about the same time, even though growth, albeit already in the late log phase, could still be observed. The reason for this is unclear. Serine was already nearly exhausted at about 0.2 mM by then. Nevertheless, we should point out that it is also at this point that alanine switches from production to consumption, and that serine and alanine are metabolically related to each other by virtue of degradation to pyruvate. Phenylalanine also, at least in part, enters central metabolism through pyruvate, so the seeming synchronization of the qualitative change in the transport pattern of this amino acid with those of serine and alanine is possibly related to this fact. Moreover, we should also note that in the aerobic
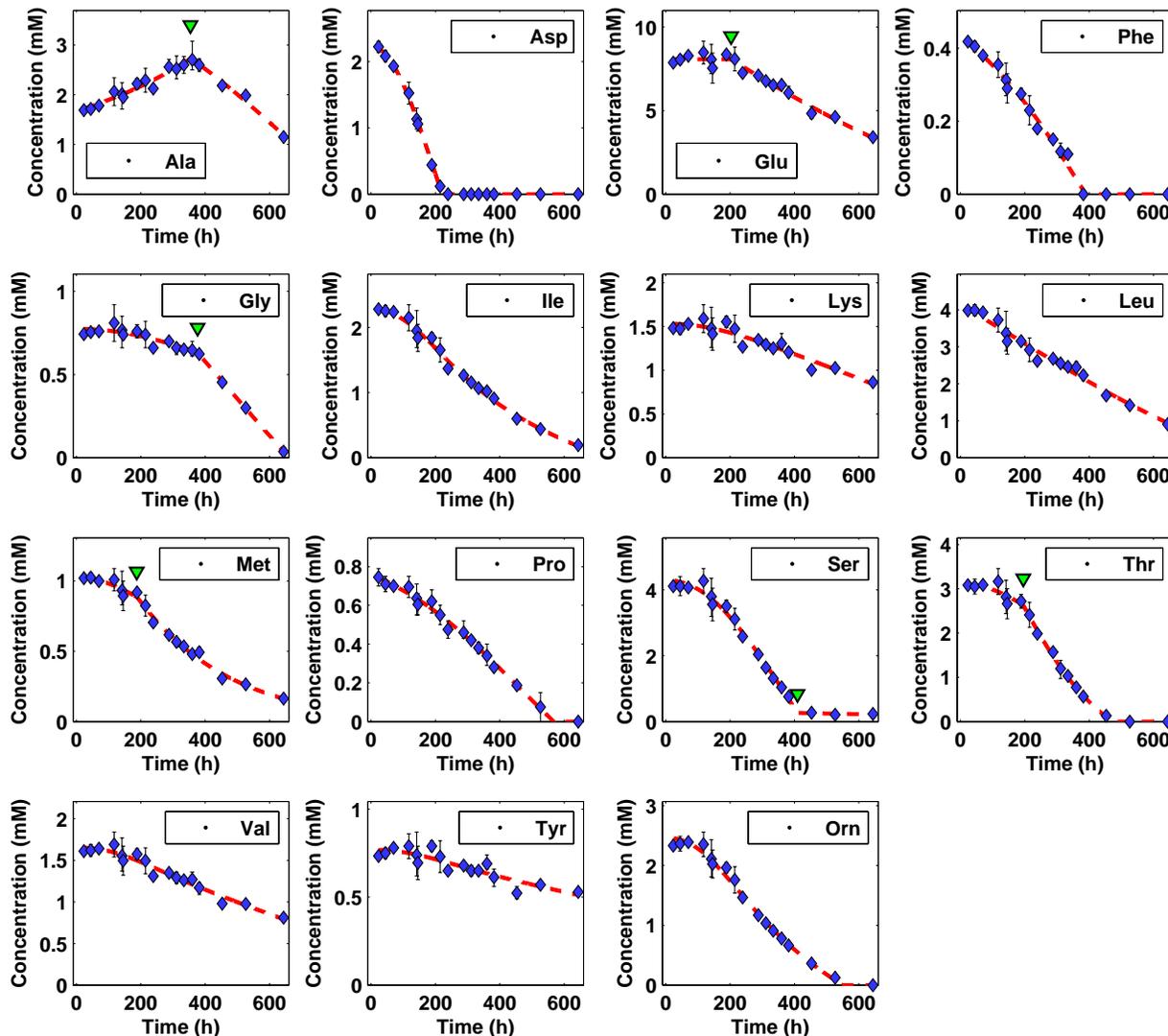
Figure 5.4: Nutrient consumption and production data from cells grown anaerobically in light. Experimental data are indicated using diamonds (average) with error bars provided. Model simulations are illustrated using red broken curves. The transport patterns of several metabolites qualitatively change during growth. For such metabolites, the piecewise Equation (3) was used to model utilization, and the corresponding $t_{i,b}$ parameter value, which represents a point near where the qualitative change occurs, is indicated with an inverted green triangle. Note that serine uptake stops at around $t = 400$ h, and does not go down to zero. This figure corresponds to Figure 4.6 for the aerobic case.

case, not only were we unable to observe an arrest in serine uptake, but the switch of alanine from production to consumption actually conincided with the acceleration of serine consumption. Therefore, an anaerobic condition is likely also a factor for the observations, ie., it is possible that phenylalanine played a more important role in the phototrophic case. In this respect, we should note that the early steps in the biosynthesis of the aromatic amino acids (Section 3.7.3) are shared with the pathway for the production of quinones, which play critical roles in respiration.

Somewhat unexpectedly, the rates at which most of the supplied amino acids were taken up from the medium far exceeded the rates at which they were directly incorporated into the biomass (fate 1), analogous to the aerobic case. The results are summarized in Figure 5.5. Again, the implication is that the consumed quantities of the pertinent amino acids are considerably catabolized by the cells. Certainly, significant portions of these are used as building blocks (carbon skeleton donors) for the production of other biomass constituents (fate 2), such as nucleotides and lipid molecules. However, even the substrate requirements for the synthesis of these unsupplied biomass components are not enough to explain the differences. This can readily be seen in Figure 5.6, which shows the total carbon uptake against the total carbon incorporation. Accordingly, considerable quantities of other metabolites (by-products) must also be produced and accumulated in the medium. This is somewhat more unexpected under anaerobic conditions because the necessary oxidative breakdown of nutrients by respiration does not occur. Under phototrophic conditions, one would typically expect material consumption to be only near the quantity sufficient for biomass production because energy can already be derived from light rather than from the supplied metabolites. The energy yield of the non-respiratory degradation of these materials should be very small compared to the energy produced through photosynthesis, especially considering that arginine is the only fermentative substrate of *H. salinarum*. Nevertheless, in retrospect, this behavior is consistent with the emerging picture that the organism takes an approach toward growth that is focused on the here and now, even at the cost of longer-term concerns. Indeed, the essential amino acids leucine, lysine, isoleucine, methionine and valine are again among the amino acids that are taken up in amounts far more than required for biomass formation.

When phototrophic cultures reached 150 klett (1.5 OD), approximately 30% of the supplied carbon that has been consumed can be attributed to the biomass. In comparison, 19% was incorporated when aerobic cultures reached comparable levels of population (also in the late log phase). This means that 70% and 81% are not incorporated under the
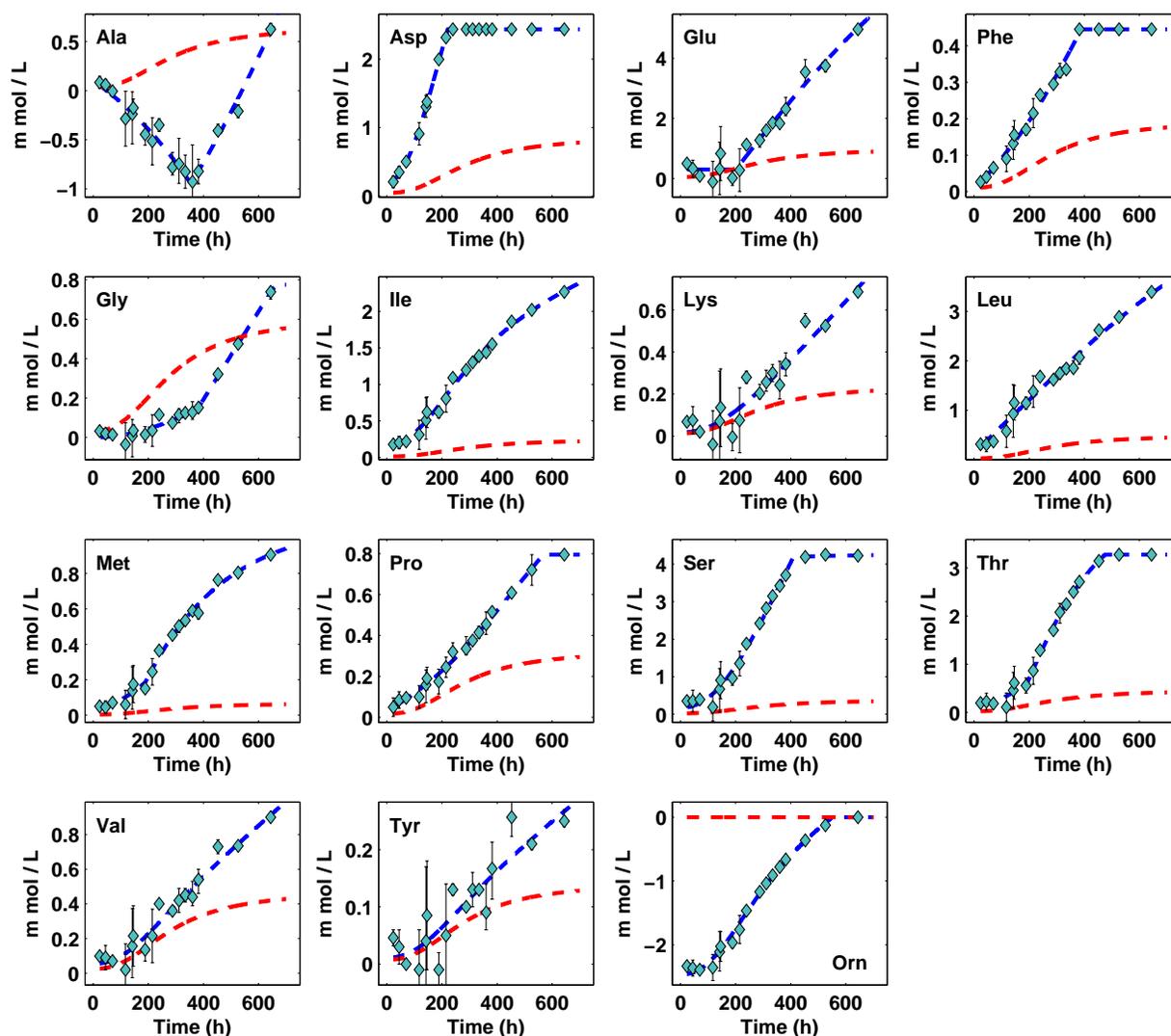
Figure 5.5: Summary of nutrient uptake and incorporation rates under phototrophic (anaerobic) conditions. The blue curves indicate the total amount of each amino acid that has been consumed/produced. The red curves show the total amount of that particular amino acid that has been incoporated into the biomass, wether integrated into proteins or as free metabolites, as calculated from the population size and the growth (biomass) function. Similar to the aerobic case, and somewhat unexpectedly, most of the supplied amino acids exhibited uptake rates that far exceeded the rates at which they were incorporated. This implies that the amino acids, in addition to biomass incorporation, are also catabolized significantly.
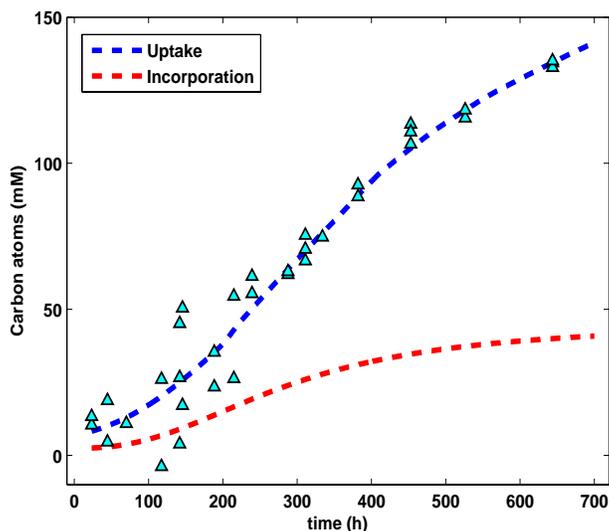
Figure 5.6: Total carbon consumption and biomass incorporation of phototrophically grown cells. The red curve shows the total amount of carbon that has been incorporated into the biomass (computed using the biomass reaction definition and the time-dependent population size). On the other hand, the blue curve indicates the total amount of carbon which has disappeared from the medium, calculated using the supplied nutrients and ornithine (filled triangles are experimental data). At $t = 650$, only $\approx 30\%$ of the total carbon that has been consumed can be found in the biomass.

phototrophic and aerobic conditions, respectively. Further adjusting the aerobic value of 81% to account for the respiratory-related oxidative degradation of nutrients (assuming a 1:1 CO_2 to O_2 ratio) reduces it to approximately 66%. Considering that the the growth rates observed under phototrophy were nearly five-fold slower, it is remarkable that the ratios are as close to each other as they are. Indeed, aerobic cultures already reached 150 klett after only slightly more than 80 hours, while, in stark contrast, phototrophic cultures took about 450 hours to reach similar levels. It would therefore seem that a large part of the nutrients that were taken up but neither incorporated into the biomass nor used as respiratory substrates has more to do with growth (biomass production) than with any form of maintenance. Accordingly, it is likely that some as of yet uncharacterized, non-maintenance growth processes are at least partially responsible for the low carbon incorporation rates.

5.3   Environmental Adaptations

Several unexpected findings regarding the growth of *Halobacterium salinarum* were made during the course of this study. First, essential amino acids are degraded under both aerobic and phototrophic (anaerobic) conditions, even though energy should already be abundant. That this is the case during respiration is already unexpected because this can lead to cells stunting their own growth. However, that this is also true under phototrophic conditions is surprising since energy can already be derived from light. Second, arginine fermentation occurs simultaneously with respiration and photosynthesis. This is in contrast to the typical assumption that the fermentation process plays a secondary, alternative role to respiration and photosynthesis. Third, considerable amounts of metabolites are produced and accumulate in the medium under both conditions, as a result of nutrients that are consumed but not incorporated into the biomass nor used as respiratory substrates. And fourth, in connection with the previous point, the total carbon incorporation rates under both conditions are very low. In particular, the incorporation rate during phototrophy was approximately 30%, when one would typically expect nutrient consumption to be just in quantities sufficient for biomass production since energy is already derived from light.

All of the findings mentioned above are consistent in depicting a seemingly "greedy" behavior by *Halobacterium salinarum*. We believe that these are actually adaptations of the halophile to its natural environments (Gonzalez *et al.*, 2008), where nutrient availability is not only irregular but can also be absent for extended periods of time. In the salt lakes and solar salterns where *Halobacterium salinarum* may be found, life is characterized by blooms that may not occur for years after a previous episode. In the Dead Sea for example, no growth of the unicellular green alga *Dunaliella parva*, which is responsible for all of the primary productivity, is possible when the salt concentration of the water column is invariably high. Blooms of *Dunaliella* in the Dead Sea occur only after significant dilution of the upper levels by the influx of freshwater. In turn, this allows the archaeal community to bloom at the expense of the organic material produced by the alga (Oren and Shilo, 1982). The fact that the conditions which allow these blooms may not be realized for years after a previous episode (Oren and Gurevich, 1995a; Oren, 1983b) may have predisposed *Halobacterium salinarum* to grow as much as possible when the conditions become favorable. In turn, this increases the probability of its survival until the next growth-conducive period. Consistent with this, the capacity of the halobacteria to survive in adverse conditions for extended periods is well established (Grant and Norton, 1998; Grant *et al.*, 1998). In this respect, we should note that a considerable number of the reported viable ancient cells

that have been recovered are at least moderately halophilic (halotolerant) (Vreeland *et al.*, 2000; Denner *et al.*, 1994; Norton *et al.*, 1993).

Chapter 6

Analysis of *N. pharaonis* Growth

6.1 The Polyextremophile

Natronomonas pharaonis can be isolated from soda lakes where it has to cope with two extreme conditions; high salt concentrations and an alkaline pH of 11. Two strains have been described so far; strain Gabara from lake Gabara in Egypt (DSM 2160) (Soliman and Truper, 1982) and strain SP1 from lake Magadi in Kenya (DSM 3395) (Tindall *et al.*, 1984). Similar to *Halobacterium salinarum*, *Natronomonas pharaonis* belongs to the order Halobacteriales, and accordingly also to the family Halobacteriaceae.
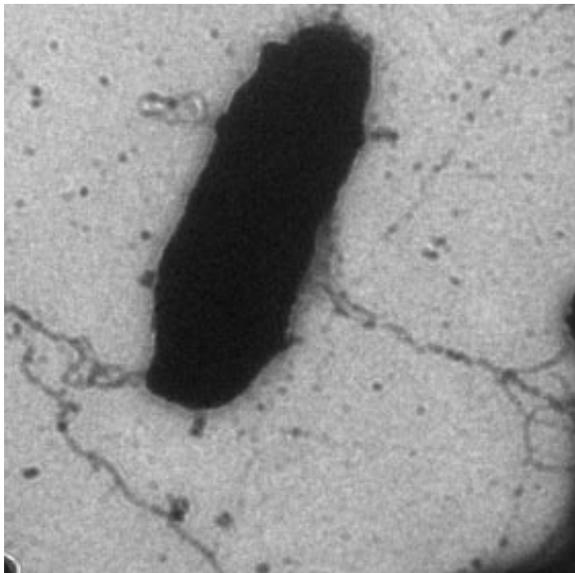
Figure 6.1: Electron microscopic image *of Natronomonas pharaonis*. The archaeon is a polyextremophile that is adapted to both high salinity and alkaline pH. The image shows two long flagellar bundles that emerge from the rod-shaped cell body. The photo was taken from the Max-Planck Institute for Biochemistry website[1].

---

[1]http://www.biochem.mpg.de/en/rd/oesterhelt/web_page_list/Org_Napha/

The genome of *Natronomonas pharaonis* type strain Gabara (DSM 2160) is organized into a 2.6-Mb chromosome, a typical haloarchaeal 131-kb plasmid (PL131) and a unique multicopy 23-kb plasmid (PL23) (Falb *et al.*, 2005). The chromosome is GC rich (63.4%GC) and contains an integrated copy of PL23. 2843 protein-coding genes were predicted, 65% of which could be assigned to a COG (cluster of orthologous group, Tatusov *et al.*, 1997, 2001). 45% and 12% of the predicted proteome could be assigned specific and general functions, respectively.

Consistent with other haloarchaea, cytoplasmic *Natronomonas pharaonis* proteins contain a high proportion of acidic amino acids (Falb *et al.*, 2005). This leads to low isoelectric points (average pI 4.6), which can increase protein stability in the presence of extremely high ionic concentrations (Section 2.3.1). In addition, the archaeon was reported to use 2-sulfotrehalose as an osmolyte, by producing it in amounts that increase with the concentration of external NaCl (Desmarais *et al.*, 1997). Its genome also codes for a typical compatible solute transporter (NP3588A).

6.2 Metabolic Reconstruction

We reconstructed the metabolic network of *Natronomonas pharaonis* using the same procedure used for *Halobacterium salinarum* (Section 3.3). The resulting network, which generally has the same coverage as the first, consists of 592 genes and 675 reactions. 56% of the ORFs have reliable proteomic identification (Konstantinidis *et al.*, 2007; Aivaliotis *et al.*, 2007). A breakdown of the reactions into high-level functional categories is provided in Figure 6.2. Compared to *Halobacterium salinarum*, very little literature was available for *Natronomonas pharaonis*. As such, most reactions were included in the network with bioinformatic evidence only.

6.2.1 Central Metabolism

Similar to *Halobacterium salinarum*, a gene for the key Embden-Meyerhof pathway enzyme 6-phosphofructokinase (EC 2.7.1.11) could not be assigned, and neither for the new archaeal types of the enzyme that use ADP as co-substrate instead of ATP. Accordingly, glucose degradation via glycolysis is also unlikely in *Natronomonas pharaonis*. In addition, the polyextremophile also seems to have the last two enzymes of the methylglyoxal bypass, glyoxylase I (EC 4.4.1.5) and glyoxylase II (EC 3.1.2.6), but not methylglyoxal synthase (EC 4.2.3.3) (see Section 3.5.3).
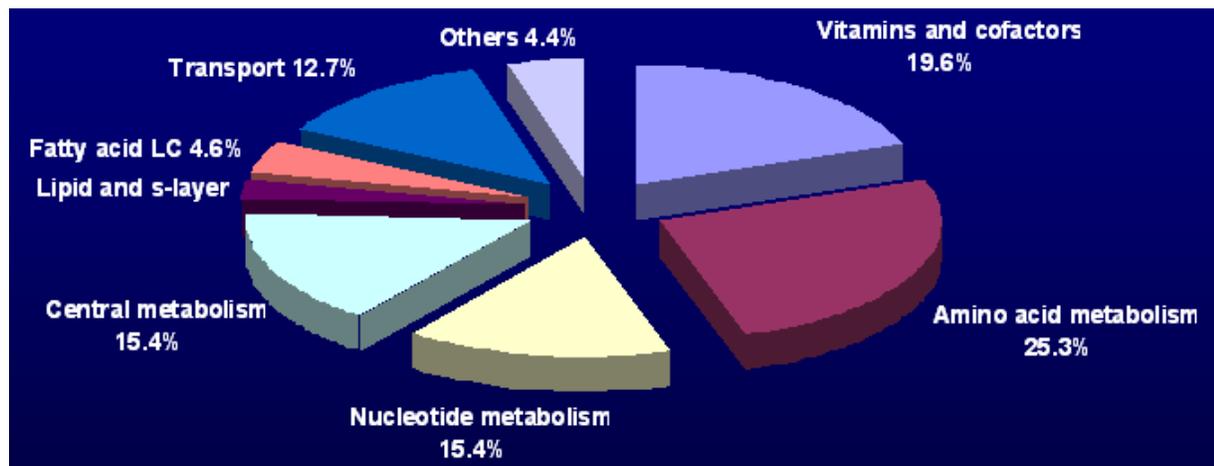
Figure 6.2: The distribution of reactions by category. The reconstructed metabolic network for *Natronomonas pharaonis* includes 676 reactions, which were subdivided into eight high-level functional categories. The "Lipid and s-layer" and "Vitamins and cofactors" categories are likely significantly underrepresented because several pathways belonging to these are yet to be characterized.

Natronomonas pharaonis has the complete set of enzymes associated with the tricarboxylic acid (TCA) cycle (Figure 6.3). However, while the cycle allows for the catabolism of acetate for the production of energy, neither it nor its reversal is enough for the process of carbohydrate biosynthesis from the compound (Section 3.5.3). Rather, this capability is afforded for by the glyoxylate shunt, which bypasses the reactions of the TCA cycle where carbon is lost in the form of CO_2. Since *Natronomonas pharaonis* is able to grow with acetate as the sole carbon source, the occurrence of the glyoxylate cycle is necessary. Indeed, genes for both isocitrate lyase (EC 4.1.3.1; NP4432A) and malate synthase (EC 2.3.3.9; NP4430A) could be assigned.

6.2.2 Nutritional Self-sufficiency and Bioenergetics

The enzyme inventory of *Natronomonas pharaonis* indicates a great level of nutritional self-sufficiency. Unlike that of *Halobacterium salinarum*, it shows the capability to synthesize all 20 proteinogenic amino acids. In addition, genes for the production of critical vitamins and cofactors, such as quinones, thiamine, coenzyme A, flavins, tetrahydrofolate, hemes, and molydopterin, are also present. Consistent with this, we were able to grow the archaeon in defined medium with only acetate, inorganic phosphate, sulfate, amonia, and other inorganic ions provided. We should mention that prior to this study, the simplest medium used to grow *Natronomonas pharaonis* contained leucine because it was previously thought
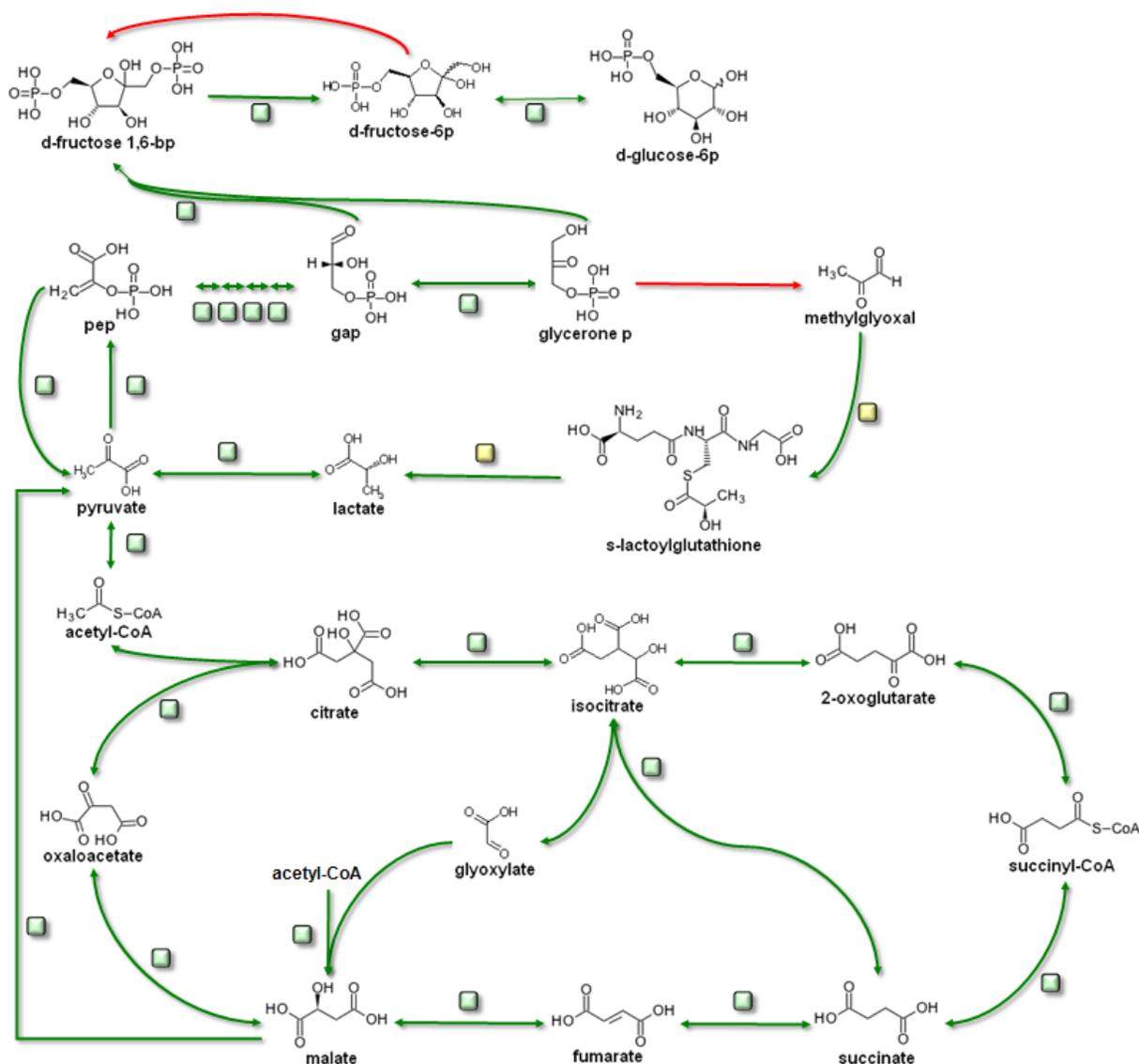
Figure 6.3: EM pathway, TCA cycle and some associated reactions in *Natronomonas pharaonis*. Green and red arrows represent reactions that occur and do not occur in the halophile, respectively. The level of bioinformatic evidence associated with each occuring reaction is indicated by the color of a nearby box: green - strong, red - none, and yellow - intermediate. Similar to *Halobacterium salinarum*, glucose degradation via the EM pathway is not possible, and a complete TCA cycle seems to be present. However, wheras only experimental evidence is present for the glyoxylate cycle enzymes in the other halophile, literature evidence is not available for *Natronomonas pharaonis* but the relevant genes could be assigned. Note that the glyoxylate cycle is necessary in the case of the polyextremophile because it can grow on 2-carbon substrates.

that the amino acid is essential for growth of the microorganism, due to disruption in the 5'-region of the isopropylmalate synthase gene (NP2206A) (Falb *et al.*, 2005). However, it is now clear that leucine is not required.

The respiratory chains of the archaea show high degrees of plasticity with respect to composition, often differing greatly from the classical five complex systems found in mitochondria and *Escherichia coli* (Falb *et al.*, 2005). In particular, the type, number, and composition of terminal oxidases vary widely, and NADH is not oxidized by proton pumping type I NADH dehydrogenase complexes because the relevant acceptor modules are not present. In the specific case of *Natronomonas pharaonis*, it is particularly interesting that while analogs of Complexes I, II, IV and V were found, not a single Complex III subunit could be identified. Since the archaeon only has the Cl$^-$ pump halorhodopsin and not the proton pump bacteriorhodopsin, it is therefore not capable of sustained photosynthetic growth, and accordingly is likely to rely primarily on respiration for the production of energy. Indeed, a working respiratory system has been shown for *Natronomonas pharaonis* (Falb *et al.*, 2005), despite the apparent absence of one of the respiratory complexes.

6.3   Methods

6.3.1   Culture Conditions and Sample Preparation

Natronomonas pharaonis strain Gabara (DSM 2160) cells were grown under aerobic conditions in chemically-defined media that were prepared according to Table 6.1. Preparatory cultures were grown in 100 ml flasks containing 35 ml of the medium to a cell density of approximately 50 Klett (0.5 OD), from which 1 ml inoculants were taken to start succeeding ones. Cultures were prepared in flasks that had side arms to measure turbidity (cell density) via a Klett photometer. All cultures were carried out in duplicates. Cell suspensions were shaken at 105 rpm at 40°C in the dark. At specific points, samples were taken from the cultures so that 6 samples were collected over the growth period, and these were stored at 4°C. To separate the cells from the medium, the samples were centrifuged for five minutes at 15,000 rpm, using a SS34 rotor. Pellets were resuspended in 500 μl basal salt (medium without acetate) and spun down as before. Amino acid analysis was performed on the pellets using an Amino Acid Analyzer (Biotronik LC3000).

Table 6.1: Composition of defined media used for *Natronomonas pharaonis*[a]

Description	Value	Description	Value
NaCl	3.4 M	$FeSO_4$ x 7 H_2O	5 μM
Na_2CO_3	175 mM	$CuSO_4$ x 5 H_2O	4 μM
KCl	27 mM	$MnCl_2$	4 μM
Na_2HPO_4 x H_2O	2mM	$CaCl_2$ x 2 H_2O	3 μM
NaH_2PO_4 x 7 H_2O	2mM	$ZnSO_4$ x 7 H_2O	3μM
$MgSO_4$ x 7 H_2O	1mM	Acetate	20 mM
NH_4Cl	12mM		
	5 μM		

[a]pH was adjusted to 7.2

6.3.2   Acetate Assay

The concentration of acetic acid in the medium was measured using the K-ACETRM assay kit from Megazyme (Wicklow, Ireland). It is principled on the following sequence of reactions: acetate kinase in the presence of ATP converts acetic acid into acetyl-phosphate

$$\text{Acetic acid} + \text{ATP} \rightarrow \text{acetyl-phosphate} + \text{ADP}$$

This reaction is driven to completion by the rapid conversion of acetyl-phosphate into acetyl-CoA and inorganic phosphate via the action of phosphotransacetylase in the presence of CoA

$$\text{acetyl-phosphate} + \text{CoA} \rightarrow \text{acetyl-CoA} + \text{Pi}$$

The ATP used by the first reaction is regenerated from ADP via the action of pyruvate kinase

$$\text{ADP} + \text{PEP} \rightarrow \text{pyruvate} + \text{ATP}$$
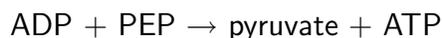
Via the action of D-lactate dehydrogenase, pyruvate is reduced to D-lactate by NADH with the production of NAD^+

$$\text{pyruvate} + \text{NADH} \rightarrow \text{NAD}^+ + \text{D-lactate}$$
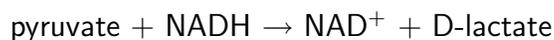
Since the amount of NAD^+ formed by the above reactions is stoichiometric with the amount of acetic acid, then the amount of NADH consumption, which can be measured through the decrease in absorbance at 340 nm, can be used to determine the concentration of acetate.

6.4 Biomass

The procedure used to determine the biomass amino acid content of *Halobacterium sali-narum* (Section 4.3) was repeated for *Natronomonas pharaonis* (Figure 6.4). Despite the differences between the two halophiles, their amino acid contents are remarkably similar, especially in relative terms (Figure 6.5). From the ones that we were able to measure, only histidine seems to be an exception, which is underrepresented by more than 50% in *Natronomonas pharaonis.*

With the exclusion of amino acids, all other biomass components of *Natronomonas pharaonis* were determined computationally. Specifcially, we calculated the ratio between its amino acid content and that of *Halobacterium salinarum,* and then simply multiplied the other biomass components of the latter by this factor. These include nucleic acids, lipids, vitamins and cofactors.

6.5 Aerobic Growth

6.5.1 Growth as a Function of Acetate and Oxygen Consumption

Given the much simpler nutritional requirements of *Natronomonas pharaonis* compared to *Halobacterium salinarum,* it presents an opportunity to analyze in more detail the relationships between network inputs and growth. For example, when the alkaliphile is grown aerobically on just acetate, it has to derive all of its energy and carbon skeleton solely from the compound. Figure 6.6 shows the maximum (theoretical) biomass production for specific combinations of oxygen and acetate consumption, calculated using flux balance analysis.

Distinct growth regimes can readily be discerned from Figure 6.6. Clearly, both acetate and oxygen need to be present in order for biomass production to be possible. That growth is unachievable without a carbon source (acetate), no matter how much oxygen is available, is obvious. However, that simulations predict that growth cannot occurr if only the carbon source is present indicates that there is currently no way to produce energy from acetate without oxygen (or any terminal electron acceptor), given the current reactions that are believed to occur within the halophile.

Figure 6.7 shows a cross section of Figure 6.6 for a particular oxygen consumption rate (190 nmol / ml hr). As expected, maximum growth increases as acetate consumption is increased from zero. However, after the rate has reached a particular point, the
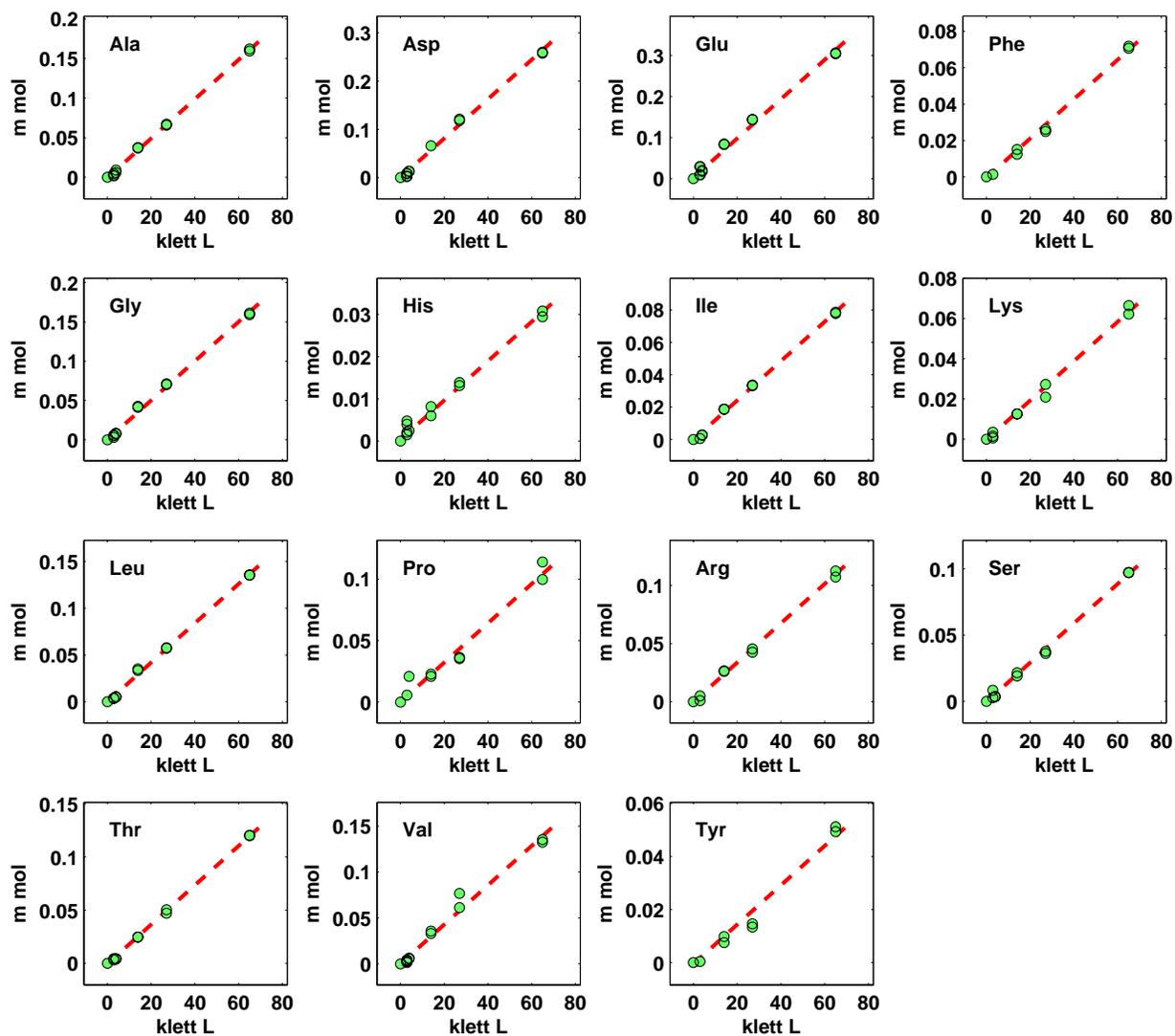
Figure 6.4: Amino acid composition of the *Natronomonas pharaonis* biomass. The values indicated are inclusive of protein residues and free metabolites. The red broken lines show the best linear fits. Due to experimental limitations, we were not able to obtain values for cysteine and tryptophan, and the values for aspartate and glutamate are already inclusive of asparagine and glutamine, respectively.
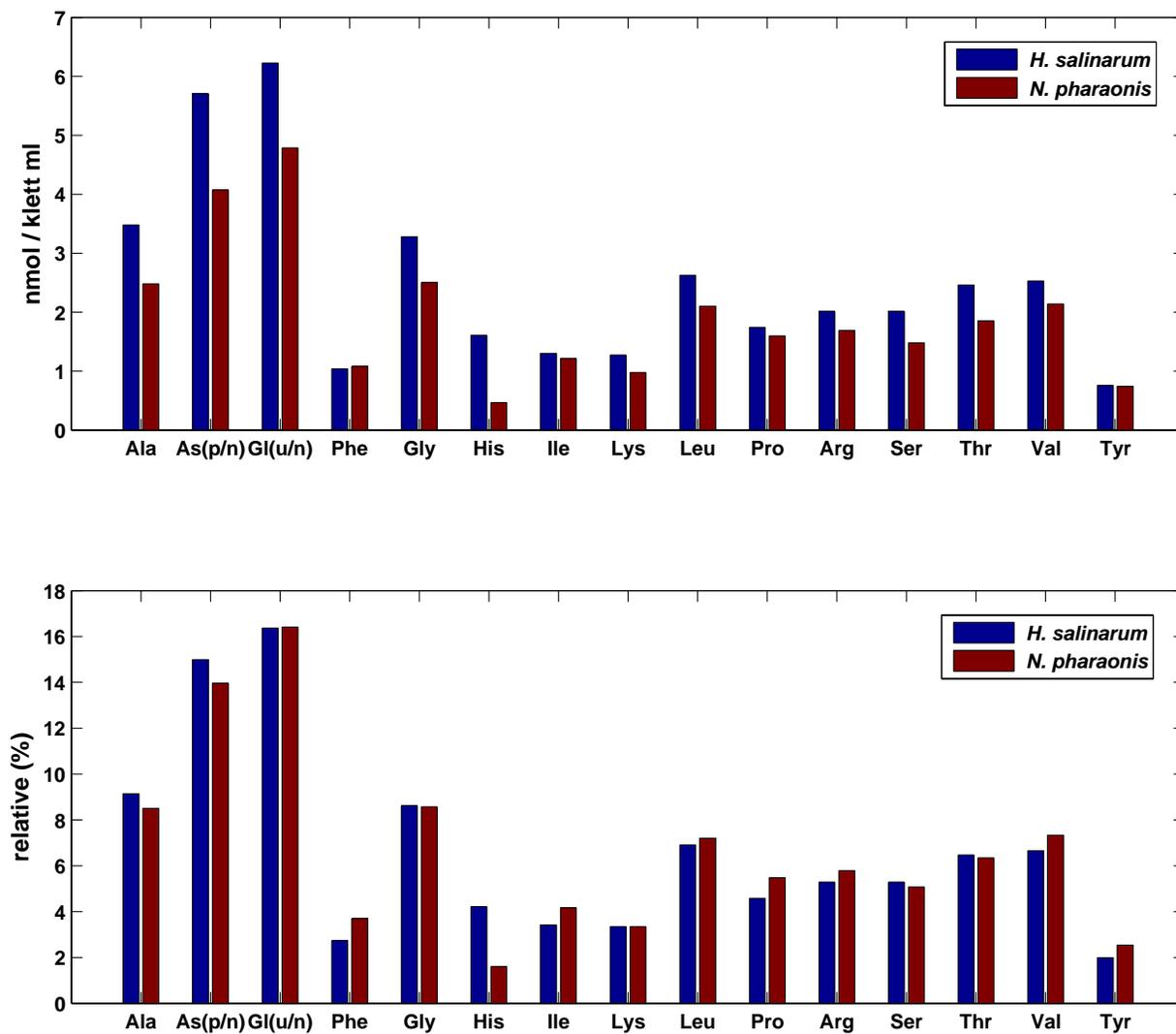
Figure 6.5: Biomass amino acid composition comparison between *Natronomonas pharaonis* and *Halobacterium salinarum*. Despite the differences between the two halophiles, their amino acid contents are remarkably similar, especially in relative terms (bottom). From the ones that we were able to measure, the only exception seems to be histidine, which is underrepresented by more than 50% in *Natronomonas pharaonis*.
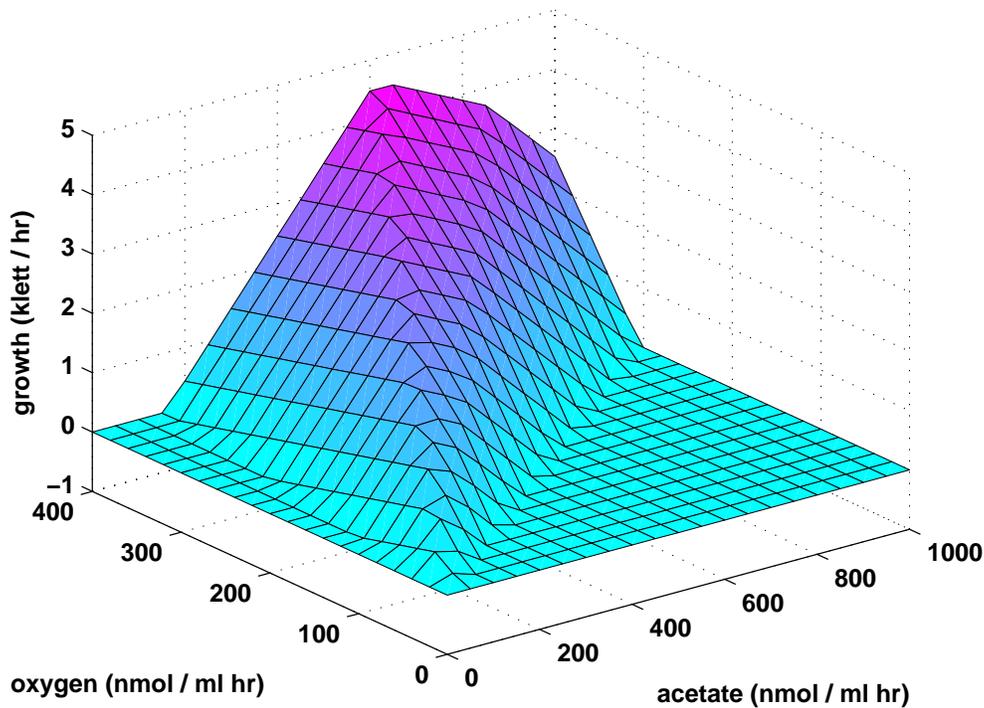
Figure 6.6: Maximum growth as a function of acetate and oxygen consumption. The maximum biomass production was calculated using flux balance analysis for specific combinations of oxygen and acetate consumption. Distinct growth regimes brought about by the interactions between these two vaiables can readily be discerned.
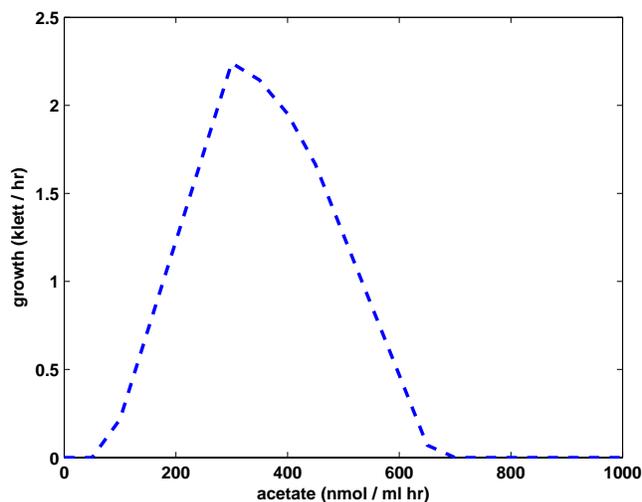
Figure 6.7: Maximum growth as a function of acetate consumption for a specific oxygen utilization rate. The image is a cross section of Figure 6.6 (Oxygen utilization = 190 nmol / ml hr). Maximum (theoretical) growth increases as acetate consumption is increased from zero. However, after the rate has reached a particular point, the effect becomes negative, and growth starts to diminish.

effect becomes negative, and growth starts to diminish. The reason is that when oxygen is limiting, the network does not have the capability to degrade excess amounts of acetate to CO_2. This in turn necessitates that the excess materials be secreted as more complex compounds. In our simulations, the first non CO_2 by-product produced was lactate, a 3-carbon coumpound, and was then followed by the 4-carbon compound succinate after acetate consumption was further increased. The energy used to produce these larger molecules competes with the energy that could have been used for biomass production, and accordingly diminishes growth. Clearly, such a behavior, if it occurs in vivo, is suboptimal.

6.5.2 An Actual Culture

The left graph in Figure 6.8 shows the growth of an actual culture of *Natronomonas pharaonis* grown aerobically on acetate. We observed a log phase doubling time of about 6.5 hours and a maximum optical density of about 64 klett. The concentration of acetate in the medium as a function of time is shown in the right graph of the same figure. Simulation of acetate consumption, which is represented by the red broken curve, was performed with the same procedure used for the consumption of metabolites by *Halobacterium salinarum* (Section 4.2.1).
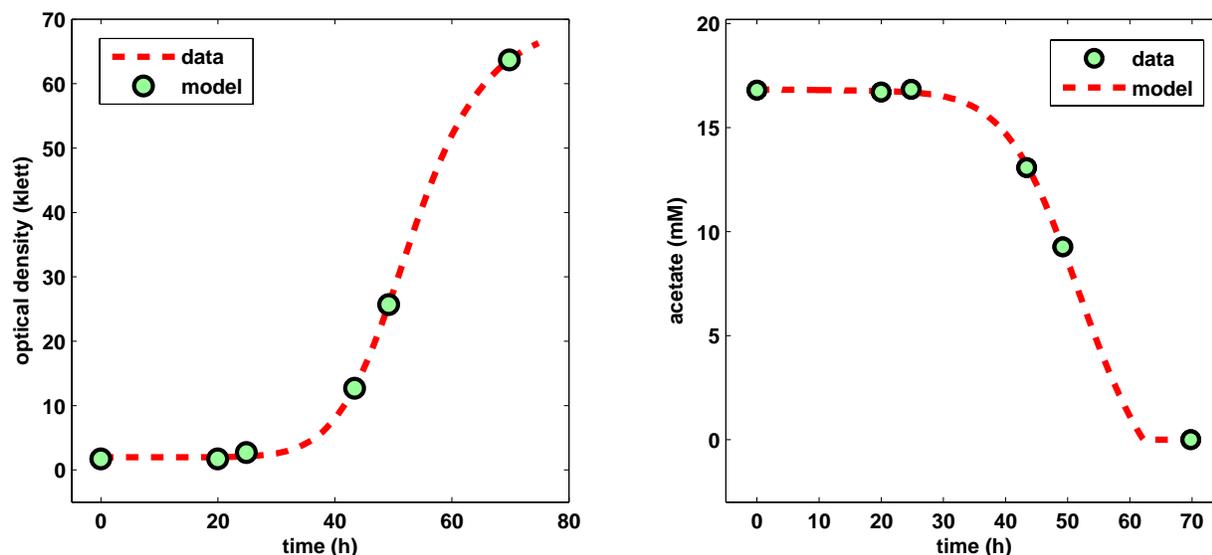
Figure 6.8: *Natronomonas pharaonis* **grown aerobically on acetate.** The left and right graphs show the population size and the concentration of acetate in the medium, respectively, as a function of time. We observed a log phase doubling time of about 6.5 hours.
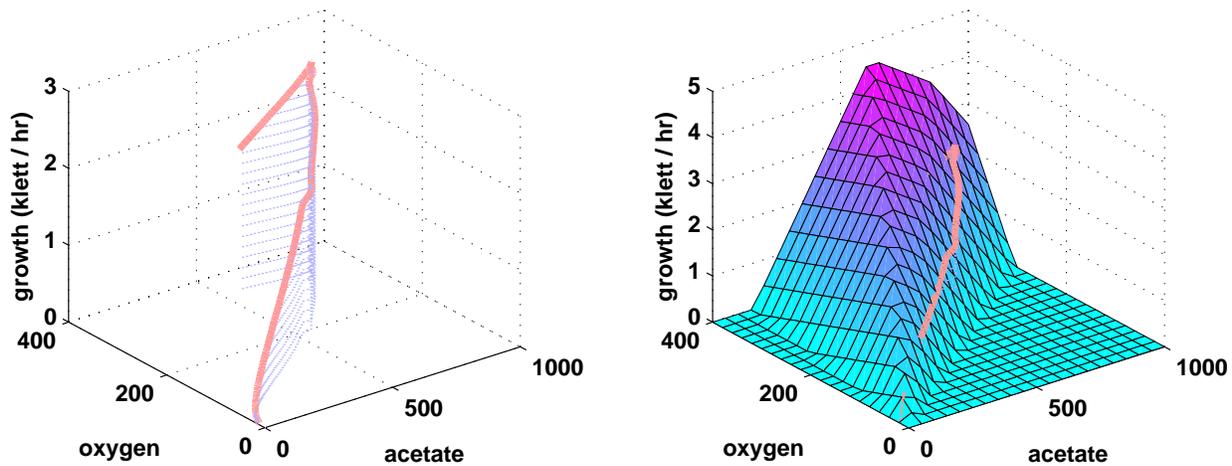
Figure 6.9: Aerobic growth summary. (Left:) The red curve summarizes the data for the aerobic culture grown on acetate. The x and y axes correspond to acetate and oxygen consumption rates, respectively, and the z-axis shows the growth rates for the corresponding points. Blue lines are drawn to provide a clearer impression of height. (Right:) The curve on the left graph (experimental data) superimposed on the calculations for the maximum (theoretical) possible growth for specific combinations of acetate and oxygen consumption rates. The proximity of the curve to the surface indicates that our model parameters, including those of the growth (biomass) function, are already within reasonable ranges, but are still in need of improvement.

The red curve in the left graph of Figure 6.9 summarizes the data for the aerobic culture grown on acetate. The x and y axes correspond to acetate and oxygen consumption rates, respectively, and the z-axis shows the growth rates for the corresponding points. The right graph in the same figure shows this summary curve superimposed on the calculations for the maximum (theoretical) possible growth for specific combinations of acetate and oxygen consumption rates (from Figure 6.6). Since the surface corresponds to maximum growth, then all parts of the curve, which represents actual data, should be on or below it. Accordingly, the proximity of the curve to the surface indicates that the parameters of our model, including those of the growth (biomass) function, are already within reasonable ranges, but are still in need of improvement.

There are several possible reasons why the curve in Figure 6.9 is above the surface in some areas. One is that oxygen measurments may be too low (by at least about 20%); note that moving the curve a little bit in the positive direction of the y-axis will put the curve below the surface. Another possibility is that the acetate measurements are too high. However, the experimental method used to assay the compound is much more accurate than the method used to determine oxygen consumption, so this is more unlikely. And a third possibility is that the growth function parameters are too high, i.e., we overestimated

the biomass content. Note that components of the growth function, save for the amino acids, were simply taken in proportion to the values from *Halobacterium salinarum* for lack of available data.

6.5.3   Summary of the Napha Model

While the computational model for *Natronomonas pharaonis* is still not at the same level as the model for *Halobacterium salinarum*, in part because the body of knowledge that exists for the latter organism is considerably larger, the reconstructed network for the haloalkaliphile has already proven valuable by providing a platform for integrating available information. For example, in the previous section we showed that some of the settings for biologically relevant parameters cannot explain actual growth data. For these, we suggested possible explanations that can then be validated experimentally, which can then be subsequently used to adjust the parameters.

Chapter 7

Conclusions and Outlook

Metabolism refers to the set of biochemical processes that occur within living organisms that allow them to maintain life. The availability of complete genome sequences for numerous organisms permit genome-scale reconstructions of metabolic networks because they can be used to supplement literature data, which is often very limited. In this work we described the reconstruction of the metabolic networks of two extremely halophilic archaeal species and the development and use of computational models derived from them.

7.1 Contributions of this Thesis

This thesis consists of interdisciplinary work by the author; it contains both computational modeling and experimental work. Most of the experimental results used for the modeling and simulations were produced by the author in Prof. Oesterhelt's group at the Max Planck Institute of Biochemistry (MPI).

One of the most important contributions of this work is the reconstructed network for *Halobacterium salinarum.* This network is composed of 695 reactions, which are linked to either genetic or literature evidence, or both. Extensive manual curation was performed on the network to achieve high reliability. Although one of the primary uses of such reconstructed networks is to serve as foundation for the development of computational models, the network in itself is already valuable, prior to any formal modeling effort, because it represents a summary of the current knowledge regarding the metabolism of *Halobacterium salinarum.* We demonstrated this value by systematically searching the network for "knowledge gaps", and then forming biological hypotheses from them. Two separate research efforts – one related to aromatic amino acid metabolism and the other

to pentose synthesis – were initiated, the work done by colleagues at the MPI, in part due to the results of the reconstruction effort. In addition, the reconstruction effort itself lead to improvements to the existing primary annotation for *Halobacterium salinarum.*

Important components of the two core constraints-based models used in this study are the growth (pseudo-)reactions. Each of these two reactions was derived from a characterization of the respective biomass composition. In the case of *Halobacterium salinarum*, the protein fraction was determined experimentally using measurements of individual amino acids at different optical densities, and the rest were derived from external data. The protein fraction of the *Natronomonas pharaonis* biomass was also determined experimentally, which allowed for a comparison between the two species.

We performed growth simulations by employing constraints-based models with the use of appropriate growth data, including nutrient consumption and growth rates. These data were integrated into the model by translating them into additional constraints. All of the relevant data were generated in this study. In the case of *Halobacterium salinarum*, the data includes time-series of 16 nutrients, under both aerobic and phototrophic conditions, and continuous measurements of oxygen usage for the aerobic case. This allowed us to compare growth under the two bioenergetic modes of respiration and photosynthesis. In the case of *Natronomonas pharaonis*, the data includes time series of acetate and oxygen consumption.

The models we used in our simulations are hybrids that have both kinetic (dynamic) and constraints-based components. Although similar models have been created in the past, one common limitation of the previously used methodologies is that they cannot account for transport patterns that change qualitatively during growth. This proved to be a particularly restrictive limitation for this study; multiple nutrients were included in the growth media, and most of these metabolites displayed qualitative changes in their transport patterns. To overcome this, we developed a new system for constructing the dynamic components of the hybrid models. This method not only could account for qualitatively changing transport patterns, but could actually automatically detect their presence. Moreover, our method could also uncover relationships between the metabolites themselves, by detecting possible interactions (e.g., competitive inhibition). This latter feature was demonstrated by automatically detecting the relationship between arginine and ornithine, which is rooted in the arginine fermentation pathway. More possible relationships between other pairs of the supplied nutrients were also discussed.

While the computational models were derived from the reconstructed metabolic net-

works, in turn, the computational models through simulations lead to improvements to the reconstruction. Moreoever, by virtue of the improvements, the computational models also contributed information to the relevant physiology. For example, during the initial reconstruction, neither bioinformatic nor experimental evidence could be found for degradative pathways for phenylalanine, tyrosine and lysine. However, simulations indicated that these capabilities are likely present.

Our integrative analysis of bioenergetics, nutrient utilization and growth of *Halobacterium salinarum* allowed us a glimpse of life under the harsh conditions that very few other species on Earth have been able to adapt to. Our analysis led to several unexpected, biologically relevant findings, including: (1) that the organism degrades essential amino acids under both aerobic and phototrophic conditions. Such behavior can lead to cells terminating their growth earlier than necessary. This was particularly surprising because cells should already have abundant energy and nonessential nutrients were available; (2) that carbon incorporation rates were surprisingly low even under phototrophic conditions. During phototrophy, one would typically expect nutrient consumption to be just in quantities that are sufficient for biomass production since energy could already be derived from light rather than from the supplied nutrients. Due to the low incorporation rates, the model predicts that considerable amounts of metabolites are likely produced and accumulated in the medium, some of which may be derived from the essential amino acids; and (3) that arginine fermentation occurs simultaneously with, and is in fact comparable in terms of energy output to, respiration and photosynthesis. This is in contrast to the common assumption that the fermentation process only plays a secondary, alternative role to the two main bioenergetic modes.

We hypothesized that the initially surprising behavior exhibited by *Halobacterium salinarum* actually consists of adaptations to its natural environments, where nutrients are not only irregularly available but may altogether be absent for extended periods that may span several years. Such a setting probably predisposed the cells to grow as much as possible when the conditions become favorable. In turn, this increases the probability of its survival until the next growth-conducive period.

7.2 Future Directions

Systems biology is an inter-disciplinary field of study that focuses on the systematic analysis of complex interactions in biological systems. Accordingly, the integration of as much

relevant information or components into the models is often desirable. While the networks analyzed in this study are already quite large, they still only account for small fractions of the complexities of the biological systems to which they belong. For example, although the reactions in the two metabolic networks are linked to their respective genomes by virtue of the fact that the reactions are linked to enzymes which are encoded by genes, details of the interactions at the genomic level are not integrated in the models. The models would probably benefit from the inclusion of gene regulatory networks, which could further restrict the solution spaces via the imposition of additional constraints (Covert and Palsson, 2003). While the construction of detailed gene regulatory networks at a genome-scale would probably not be feasible for some time, except for organisms that are as widely studied as *E. coli*, specific aspects of interest could be focused on for inclusion into the network. For example, the genes related to switching from photosynthetic to aerobic growth, or vice versa, could be determined with targeted experiments. Even if the regulatory relationships between these genes (e.g., causality) prove difficult to establish, the regulation data themselves (eg., regulation factors) could be compared to the results of the basic metabolic model, and then subsequently used to add more constraints (albeit static ones).

In addition to *Halobacterium salinarum* and *Natronomonas pharaonis*, complete genome sequences are also available for four other haloarchaeal species, and more are currently being sequenced. Metabolic reconstructions for these other species would be useful for performing a comparative study.

Analysis with constraints-based models typically do not result in unique solutions. For example, with flux variability analysis we could only describe fluxes that are compatible with our experimental data. While we found that fluxes for a large fraction of the reactions in the network are already constrained enough to be qualitatively invariable, more detailed information on these fluxes are desirable. Not to mention the reactions that are not even qualitatively constrained. To address this issue, quantitative measurements of the fluxes themselves would have to be experimentally obtained, for example using flux analysis with labeled substrates. While the task will be very laborious, due to the available methods or techniques, the data that could be obtained is potentially very useful. For example, it would reveal details on the fates of the individual carbon sources. In addition, performing flux analysis under different conditions (e.g., phototrophic v.s. aerobic) could yield insights on the metabolic strategy of the microorganism.

One of the predictions of our model is that large amounts of metabolites, possibly by-products, are formed and accumulate in the medium, under both aerobic and phototrophic

conditions. This is particularly remarkable because our analysis suggests that the secreted metabolites account for over 50% of the carbon consumed by the cells, some of which may be derived from the degradation of essential nutrients. In the future, we intend to experimentally confirm this prediction, and elucidate the identities of the metabolites themselves. Subsequently, we would also like to find out the reasons why these metabolites are secreted in the first place.

Besides the integration of other components (e.g., the genetic regulatory networks mentioned above) into the models, one of the future directions we would like to take is to introduce dynamics to at least some more parts of the network; in the current hybrid model for *Halobacterium salinarum*, only parts associated with nutrient transport have dynamic elements. Although creating a fully kinetic metabolic model with the same scale is currently infeasible due to the fact that only a very small percentage of reactions have available kinetic information, we would nevertheless like to integrate the current model with kinetic models of respiration and photosynthesis, which are currently being developed by colleagues at the MPI.

Appendix A

Reconstructed Networks

Table A.1: *Halobacterium salinarum* reaction list. In cases where there is ambiguity in the genetic evidence, all of the candidate genes are listed.

Id	Definition	EC No.	Genes	Ref
R01737	ATP + D-Gluconic acid => ADP + 6-Phospho-D-gluconate	2.7.1.12		
R01737	ATP + D-Gluconic acid => ADP + 6-Phospho-D-gluconate	2.7.1.12		
R00396	L-Alanine + NAD+ + H2O <=> Pyruvate + NH3 + NADH + H+	1.4.1.1		PMID:486498 PMID:7371854
R03869	(S)-Methylmalonate semialdehyde + NAD+ + H2O => Methylmalonate + NADH	1.2.1.3	OE2133R OE2190R OE2367F OE4529F	
R01731	L-Aspartate + Prephenate <=> Oxaloacetate + L-Arogenate	2.6.1.57		
R01734	Dimethylbenzimidazole <=> Riboflavin	-		
R01736	(R)-S-Lactoylglutathione + H2O <=> Glutathione + (R)-Lactate	3.1.2.6	OE7071F OE6191F OE7077F OE6197F	
R00782	Hydrogen sulfide + Pyruvate + NH3 => L-Cysteine + H2O	4.4.1.1 4.4.1.8	OE2173F OE2681F	
R00489	L-Aspartate => beta-Alanine + CO2	4.1.1.11 4.1.1.15	OE1498R	
R01728	Prephenate + NAD+ => 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADH + H+	1.3.1.12 1.3.1.52 1.3.1.43	OE2770F	
R02487	Glutaryl-CoA + FAD => FADH2 + Crotonoyl-CoA + CO2	1.3.99.7	OE4500R	
R02485	Deoxycytidine + H2O => Deoxyuridine + NH3	3.5.4.5	OE2313R	
R02742	gamma-L-Glutamyl-L-cysteine + NADP+ <=> Bis-gamma-glutamylcystine + NADPH	1.8.1.13		PMID:3136140 PMID:2910862
R00885	GTP + D-Mannose 1-phosphate <=> Pyrophosphate + GDPmannose	2.7.7.13	OE4022R	
R02748	Deoxyinosine + Orthophosphate <=> Hypoxanthine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.1		PMID:9457844
R02340	Indoleglycerol phosphate <=> Indole + D-Glyceraldehyde 3-phosphate	4.2.1.20	OE1470F OE1471F	
R00481	L-Aspartate + Oxygen => Iminoaspartate + H2O2	1.4.3.16	OE3646F	
R02749	2-Deoxy-D-ribose 1-phosphate <=> 2-Deoxy-D-ribose 5-phosphate	5.4.2.7	OE2318R	
R00789	Nitrite + NADPH + H+ => NADP+ + NH3 + H2O	1.7.1.4		PMID:629538
R00480	ATP + L-Aspartate => ADP + 4-Phospho-L-aspartate	2.7.2.4	OE4333R	
R01724	Nicotinate + 5-Phospho-alpha-D-ribose 1-diphosphate => Nicotinate D-ribonucleotide + Pyrophosphate	2.4.2.11	OE4565F	

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00485	L-Asparagine + H2O => L-Aspartate + NH3	3.5.1.1	OE3633R	
R05048	2,5-Diaminopyrimidine nucleoside triphosphate => 2,5-Diamino-6-(5'-triphosphoryl-3',4'-trihydroxy-2'-oxopentyl)-	3.5.4.16		PMID:17032654
R00479	Isocitrate <=> Succinate + Glyoxylate	4.1.3.1		PMID:5780095
R05046	Formamidopyrimidine nucleoside triphosphate + H2O => 2,5-Diaminopyrimidine nucleoside triphosphate + Formate	3.5.4.16		PMID:17032654
R00376	dGTP => Pyrophosphate + dGMP	2.7.7.7	OE4293R OE1777F OE4390F OE6036F OE7114F	
R00377	dCTP => Pyrophosphate + dCMP	2.7.7.7	OE4293R OE1777F OE4390F OE6036F OE6234F OE7114F	
R00891	L-Serine + Sulfide => L-Cysteine + H2O	4.2.1.22	OE1916F OE2860R	
R00378	dTTP => Pyrophosphate + dTMP	2.7.7.7	OE4293R OE1777F OE4390F OE6036F OE6234F OE7114F	
R05815	Cobyrinate + L-Glutamine + ATP + H2O => Cob(II)yrinate a,c diamide + L-Glutamate + ADP + Orthophosphate	6.3.1.- 6.3.5.9	OE3243F	
R01090	L-Leucine + 2-Oxoglutarate <=> 4-Methyl-2-oxopentanoate + L-Glutamate	2.6.1.42 2.6.1.6	OE3959R	
R05814	Cobalt-precorrin 8 => Cobyrinate	5.4.1.2	OE3237F	
R02472	(R)-Pantoate + NADP+ <=> 2-Dehydropantoate + NADPH	1.1.1.169	OE2078R	
R00897	O-Acetyl-L-serine + Hydrogen sulfide => L-Cysteine + Acetate	2.5.1.47	OE1916F OE2860R	
R05811	Cobalt-precorrin 5 + S-Adenosyl-L-methionine + H2O => Cobalt-precorrin 6 + S-Adenosyl-L-homocysteine + Acetaldehyde	2.1.1.-	OE3994F	
R00375	dATP => Pyrophosphate + dAMP	2.7.7.7	OE4293R OE1777F OE4390F OE6036F OE6234F OE7114F	
R00896	Mercaptopyruvate + Glutamate <=> L-Cysteine + 2-Oxoglutarate	2.6.1.1	OE1755F OE1944R OE2619F	
R02473	ATP + (R)-Pantoate + beta-Alanine => AMP + Pyrophosphate + Pantothenate	6.3.2.1		
R05810	Cobalt-precorrin 4 + S-Adenosyl-L-methionine => Cobalt-precorrin 5 + S-Adenosyl-L-homocysteine	2.1.1.133	OE3212F	
R05813	Cobalt-dihydro-precorrin 6 + S-Adenosyl-L-methionine => Cobalt-precorrin 8 + S-Adenosyl-L-homocysteine + CO2	2.1.1.132	OE3207F OE3238F	
R05812	Cobalt-precorrin 6 + NADPH + H+ => Cobalt-dihydro-precorrin 6 + NADP+	1.3.1.54		
R00751	L-Threonine <=> Glycine + Acetaldehyde	4.1.2.5	OE4436R	
R01752	D-Glyceraldehyde + NAD+ + H2O <=> D-Glycerate + NADH + H+	1.2.1.3	OE2133R OE2190R OE2367F OE4529F	
R02372	dUTP + Cytidine => dUDP + CMP	2.7.1.48	OE2749F	
R03990	Tetradecanoyl-CoA + FAD => trans-Tetradec-2-enoyl-CoA + FADH2	1.3.3.6 1.3.99.3 1.3.99.13	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R	
R03991	Tetradecanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxopalmitoyl-CoA	2.3.1.16	OE3884F	
R05617	1,4-Dihydroxy-2-naphthoate + all-trans-Octaprenyl diphosphate => 2-Demethylmenaquinone + Pyrophosphate + CO2	2.5.1.-	OE2559R OE2650F OE3503F OE3505R OE4010F	
R01899	Isocitrate + NADP+ <=> Oxalosuccinate + NADPH + H+	1.1.1.42	OE3634F	PMID:4393394 PMID:5058691 PMID:5780095 PMID:5802601 PMID:9827332
R05612	all-trans-Hexaprenyl diphosphate + Isopentenyl diphosphate => all-trans-Heptaprenyl diphosphate + Pyrophosphate	2.5.1.30	OE4010F OE2650F	PMID:2857171

Id	Definition	EC No.	Genes	Ref
R05613	all-trans-Pentaprenyl diphosphate + Isopentenyl diphosphate <=> all-trans-Hexaprenyl diphosphate + Pyrophosphate	2.5.1.33	OE4010F OE2650F	PMID:2857171
R00369	L-Alanine + Glyoxylate => Pyruvate + Glycine	2.6.1.44	OE4391F	PMID:3700337 PMID:8157586
R05611	all-trans-Heptaprenyl diphosphate + Isopentenyl diphosphate <=> all-trans-Octaprenyl diphosphate + Pyrophosphate	2.5.1.11 2.5.1.-	OE2559R OE2650F OE3503F OE3505R OE4010F	PMID:2857171
R00362	Citrate <=> Acetate + Oxaloacetate	4.1.3.6	OE1942F	
R02765	(R)-2-Methyl-3-oxopropanoyl-CoA <=> (S)-2-Methyl-3-oxopropanoyl-CoA	5.1.99.1	OE1718R	
R03194	S-Adenosyl-L-methionine + Uroporphyrinogen III => S-Adenosyl-L-homocysteine + Precorrin 2	2.1.1.107	OE4277F	
R00762	D-Fructose 1,6-bisphosphate + H2O => D-Fructose 6-phosphate + Orthophosphate	3.1.3.11	OE2020F	PMID:2256927 PMID:3255682 PMID:6642538 PMID:9827332
R02864	Iron + Sirohydrochlorin => Siroheme + H+	4.99.1.4	OE3498R	
R05809	Cobalt-precorrin 3 + S-Adenosyl-L-methionine => Cobalt-precorrin 4 + S-Adenosyl-L-homocysteine	2.1.1.131	OE3214F OE3216F	
R05808	Cobalt-precorrin 2 + S-Adenosyl-L-methionine => Cobalt-precorrin 3 + S-Adenosyl-L-homocysteine	2.1.1.130	OE3209F	
R00768	L-Glutamine + D-Fructose 6-phosphate => L-Glutamate + D-Glucosamine 6-phosphate	2.6.1.16	OE1013R	
R05807	Sirohydrochlorin + Cobalt => Cobalt-precorrin 2	4.99.1.3	OE3221F	
R02872	Presqualene diphosphate + NADPH + H+ => Pyrophosphate + Squalene + NADP+	2.5.1.21	OE2014F	PMID:2857171
R00771	D-Glucose 6-phosphate <=> D-Fructose 6-phosphate	5.3.1.9	OE3792F	PMID:2256927 PMID:3255682 PMID:6642538 PMID:9827332
R00354	(3S)-Citryl-CoA <=> Acetyl-CoA + Oxaloacetate	4.1.3.34	OE1942F	
R00355	L-Aspartate + 2-Oxoglutarate <=> Oxaloacetate + L-Glutamate	2.6.1.1	OE1755F OE1944R OE2619F	
R00357	L-Aspartate + H2O + Oxygen => Oxaloacetate + NH3 + H2O2	1.4.3.16	OE3646F	
R02771	dTDPglucose <=> dTDP-4-dehydro-6-deoxy-alpha-D-glucose + H2O	4.2.1.46	OE1115F	
R07280	5-Amino-6-(5'-phosphoribitylamino)uracil + H2O => 4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine + Orthophosphate	3.1.3.-	OE2299F	
R07281	D-Ribulose 5-phosphate => 3,4-Dihydroxy-2-butanone 4-phosphate + Formate	5.4.99.-	OE3963R	
R01073	Anthranilate + 5-Phospho-alpha-D-ribose 1-diphosphate => N-(5-Phospho-D-ribosyl)anthranilate + Pyrophosphate	2.4.2.18	OE3334R	
R01072	L-Glutamine + 5-Phospho-alpha-D-ribose 1-diphosphate + H2O => 5-Phosphoribosylamine + Pyrophosphate + L-Glutamate	2.4.2.14	OE3139R	
R01776	Acetyl-CoA + L-Homoserine <=> CoA + O-Acetyl-L-homoserine	2.3.1.31	OE4397F	
R01775	L-Homoserine + NADP+ <=> L-Aspartate 4-semialdehyde + NADPH + H+	1.1.1.3	OE4722R	
R01074	ATP + Biotin <=> Pyrophosphate + Biotinyl-5'-AMP	6.3.4.15	OE3186F OE1987F	
R02914	Urocanate + H2O <=> 4-Imidazolone-5-propanoate	4.2.1.49	OE2734F	
R01880	dGTP + Uridine => dGDP + UMP	2.7.1.48	OE2749F	
R02918	ATP + L-Tyrosine + tRNA(Tyr) <=> AMP + Pyrophosphate + L-Tyrosyl-tRNA(Tyr)	6.1.1.1	OE4139R	
R01071	ATP + 5-Phospho-alpha-D-ribose 1-diphosphate <=> Phosphoribosyl-ATP + Pyrophosphate	2.4.2.17	OE4152R	PMID:3700337
R01676	Guanine + H2O => Xanthine + NH3	3.5.4.3	OE2730R	PMID:9457844

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R01070	D-Fructose 1,6-bisphosphate <=> Glycerone phosphate + D-Glyceraldehyde 3-phosphate	4.1.2.13	OE2019F	
R01876	Uridine + Orthophosphate <=> Uracil + alpha-D-Ribose 1-phosphate	2.4.2.3	OE2311R OE3603R	
R06977	L-Glutamate + L-Aspartate 4-semialdehyde <=> 2-Oxoglutarate + L-2,4-Diaminobutanoate	2.6.1.76	OE5094F	
R01878	Cytidine + H2O => Uridine + NH3	3.5.4.5	OE2313R	
R03084	3-Dehydroquinate <=> 3-Dehydroshikimate + H2O	4.2.1.10 4.2.1.11	OE1477R OE2640F	PMID:3700337
R01975	Acetoacetyl-CoA + NADH <=> (S)-3-Hydroxybutanoyl-CoA + NAD+	1.1.1.35	OE2015R OE2871F	
R01976	Acetoacetyl-CoA + NADPH <=> (S)-3-Hydroxybutanoyl-CoA + NADP+	1.1.1.157	OE2015R OE2871F	
R01977	Acetoacetyl-CoA + NADPH <=> (R)-3-Hydroxybutanoyl-CoA + NADP+	1.1.1.36	OE4695F	
R01978	Acetyl-CoA + H2O + Acetoacetyl-CoA => (S)-3-Hydroxy-3-methylglutaryl-CoA + CoA	2.3.3.10	OE3296F	PMID:3700337
R01086	N-(L-Arginino)succinate <=> Fumarate + L-Arginine	4.3.2.1	OE4419R	
R00086	ATP + H2O => ADP + Orthophosphate	3.6.1.15	OE3863R	
R01083	N6-(1,2-Dicarboxyethyl)-AMP <=> Fumarate + AMP	4.3.2.2	OE1623F	
R00084	Porphobilinogen + H2O => Hydroxymethylbilane + NH3	2.5.1.61	OE4276F	
R01870	Orotidine 5'-phosphate + Pyrophosphate <=> Orotate + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.10	OE1672F OE3953R	
R02237	ATP + Dihydropteroate + L-Glutamate => ADP + Orthophosphate + Dihydrofolate	6.3.2.17 6.3.2.12 2.1.1.45	OE1615R	
R01082	(S)-Malate <=> Fumarate + H2O	4.2.1.2	OE2935R	PMID:5780095 PMID:9827332
R01665	ATP + dCMP => ADP + dCDP	2.7.4.14	OE3429F	
R01664	dCMP + H2O => Deoxycytidine + Orthophosphate	3.1.3.5	OE3017R	
R02235	Dihydrofolate + NAD+ <=> Folate + NADH + H+	1.5.1.3		PMID:17416665 PMID:15554970
R02236	Dihydrofolate + NADP+ <=> Folate + NADPH + H+	1.5.1.3	OE2921R	PMID:15554970 PMID:17416665
R01867	(S)-Dihydroorotate + Oxygen => Orotate + H2O2	1.3.3.1	OE4508R	
R02245	ATP + (R)-Mevalonate => ADP + (R)-5-Phosphomevalonate	2.7.1.36	OE2645F	PMID:3700337
R01968	dGMP + H2O => Deoxyguanosine + Orthophosphate	3.1.3.5	OE3017R	
R01969	Deoxyguanosine + Orthophosphate <=> Guanine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.1		PMID:9457844
R01863	Inosine + Orthophosphate <=> Hypoxanthine + alpha-D-Ribose 1-phosphate	2.4.2.1		PMID:9457844
R01771	ATP + L-Homoserine => ADP + O-Phospho-L-homoserine	2.7.1.39	OE3531R	PMID:3700337
R02100	dUTP + H2O => dUMP + Pyrophosphate	3.6.1.23	OE4610R	
R02331	ATP + dUDP <=> ADP + dUTP	2.7.4.6	OE2667F	
R01057	alpha-D-Ribose 1-phosphate <=> D-Ribose 5-phosphate	5.4.2.7	OE2318R	PMID:15205420
R02332	dUTP + Uridine => dUDP + UMP	2.7.1.48	OE2749F	
R01056	D-Ribose 5-phosphate <=> D-Ribulose 5-phosphate	5.3.1.6	OE4185F	
R01051	ATP + D-Ribose => ADP + D-Ribose 5-phosphate	2.7.1.15	OE3606R	
R02719	Xanthosine 5'-phosphate + H2O => Xanthosine + Orthophosphate	3.1.3.5	OE3017R	
R01655	5,10-Methenyltetrahydrofolate + H2O <=> 10-Formyltetrahydrofolate + H+	3.5.4.9 6.3.4.3	OE3038F	
R01658	Dimethylallyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + Geranyl diphosphate	2.5.1.10 2.5.1.29 2.5.1.1	OE2650F OE4010F	PMID:3700337
R01857	ATP + dGDP <=> ADP + dGTP	2.7.4.6	OE2667F	

Id	Definition	EC No.	Genes	Ref
R02328	dTTP + D-Glucose 1-phosphate <=> Pyrophosphate + dTDPglucose	2.7.7.24	OE1578F OE1016R OE1078F OE1114F OE2530F OE1014R	
R01858	dGDP + Phosphoenolpyruvate => dGTP + Pyruvate	2.7.1.40	OE1495R	
R02327	dCTP + Uridine => dCDP + UMP	2.7.1.48	OE2749F	
R01859	ATP + Propanoyl-CoA + HCO3- => ADP + Orthophosphate + (S)-2-Methyl-3-oxopropanoyl-CoA	6.4.1.3	OE1939F OE3175F	
R02329	dTTP + alpha-D-Galactose 1-phosphate <=> Pyrophosphate + dTDPgalactose	2.7.7.32	OE1578F OE1016R OE1078F OE1114F OE2530F OE1014R	
R01993	(S)-Dihydroorotate + H2O <=> N-Carbamoyl-L-aspartate	3.5.2.3	OE4558R	
R00999	O-Succinyl-L-homoserine + H2O => 2-Oxobutanoate + Succinate + NH3	2.5.1.48	OE2173F OE2681F	
R02519	Homogentisate + Oxygen => 4-Maleylacetoacetate	1.13.11.5		
R02326	ATP + dCDP <=> ADP + dCTP	2.7.4.6	OE2667F	
R01049	ATP + D-Ribose 5-phosphate <=> AMP + 5-Phospho-alpha-D-ribose 1-diphosphate	2.7.6.1	OE4085R	
R00996	L-Threonine => 2-Oxobutanoate + NH3	4.3.1.19	OE3931R	
R02325	dCTP + H2O => dUTP + NH3	3.5.4.13	OE1384F	
R01068	D-Fructose 1,6-bisphosphate <=> Glycerone phosphate + D-Glyceraldehyde 3-phosphate	4.1.2.13	OE2019F	PMID:2256927 PMID:3255682 PMID:6642538 PMID:9827332
R10005	Prephytoene diphosphate + H+ + NADPH => Pyrophosphate + Phytoene + NADP+	2.5.1.32	OE3093R OE3376F	PMID:2857171
R01066	2-Deoxy-D-ribose 5-phosphate <=> D-Glyceraldehyde 3-phosphate + Acetaldehyde	4.1.2.4	OE3616F	
R10004	all-trans-Octaprenyl diphosphate + Isopentenyl diphosphate => all-trans-Nonaprenyl diphosphate + Pyrophosphate	2.5.1.11		PMID:2857171
R10003	Geranylgeranyl diphosphate + Isopentenyl diphosphate => all-trans-Pentaprenyl diphosphate + Pyrophosphate	2.5.1.-	OE2650F OE4010F	PMID:2857171
R01061	D-Glyceraldehyde 3-phosphate + Orthophosphate + NAD+ <=> 3-Phospho-D-glyceroyl phosphate + NADH + H+	1.2.1.12 1.2.1.59	OE1154F	PMID:2256927 PMID:3255682 PMID:6642538 PMID:9827332
R02530	Glutathione + Methylglyoxal => (R)-S-Lactoylglutathione	4.4.1.5	OE2610F	FEMS Microbiology Letters, 1995, 125:83-88
R02531	Methylglyoxal + NADH + H+ <=> Lactaldehyde + NAD+	1.1.1.21	OE1698R OE2451R OE2486F OE3312R	
R01648	4-Aminobutanoate + 2-Oxoglutarate <=> Succinate semialdehyde + L-Glutamate	2.6.1.19	OE3168R	
R02722	L-Serine + Indoleglycerol phosphate <=> L-Tryptophan + Glyceraldehyde 3-phosphate + H2O	4.2.1.20	OE1470F OE1471F	
R02413	Shikimate + NADP+ <=> 3-Dehydroshikimate + NADPH + H+	1.1.1.25	OE1565F	PMID:3700337
R02412	ATP + Shikimate => ADP + Shikimate 3-phosphate	2.7.1.71	OE2785R	PMID:3700337
R05064	(S)-3-Hydroxyisobutyryl-CoA + H2O <=> CoA + (S)-3-Hydroxyisobutyrate	3.1.2.4		
R02521	3-(4-Hydroxyphenyl)pyruvate + Oxygen => Homogentisate + CO2	1.13.11.27		
R05066	(S)-3-Hydroxyisobutyrate + NAD+ <=> (S)-Methylmalonate semialdehyde + NADH + H+	1.1.1.35	OE2015R	
R01058	D-Glyceraldehyde 3-phosphate + NADP+ + H2O => 3-Phospho-D-glycerate + NADPH	1.2.1.9	OE2133R OE2190R OE2367F OE4529F	PMID:2256927 PMID:9827332
R02313	N6-(L-1,3-Dicarboxypropyl)-L-lysine + NAD+ + H2O <=> L-Glutamate + L-2-Aminoadipate 6-semialdehyde + NADH + H+	1.5.1.9		PMID:3700337
				Continued on next page

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R04509	ATP + 4-Amino-2-methyl-5-phosphomethylpyrimidine => ADP + 2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate	2.7.4.7	OE4654F	
R00691	L-Arogenate => L-Phenylalanine + H2O + CO2	4.2.1.51	OE4117F	PMID:3700337
R02408	L-Cystine + H2O => Pyruvate + NH3 + Thiocysteine	4.4.1.1 4.4.1.8	OE2173F OE2681F	
R00694	L-Phenylalanine + 2-Oxoglutarate <=> Phenylpyruvate + L-Glutamate	2.6.1.9 2.6.1.1 2.6.1.5 2.6.1.57	OE2507R OE1755F OE1944R OE2619F	
R00405	ATP + Succinate + CoA <=> ADP + Orthophosphate + Succinyl-CoA	6.2.1.5	OE3195F OE3196F	PMID:5780095 PMID:9827332
R00303	D-Glucose 6-phosphate + H2O => D-Glucose + Orthophosphate	3.1.3.9		PMID:3255682 PMID:9827332
R00408	Succinate + FAD <=> FADH2 + Fumarate	1.3.99.1	OE2865R OE2866R OE2867R OE2868R	PMID:5780095
R03045	3-Hydroxypropionyl-CoA <=> Propenoyl-CoA + H2O	4.2.1.17	OE3846R	
R03038	ATP + L-Alanine + tRNA(Ala) <=> AMP + Pyrophosphate + L-Alanyl-tRNA	6.1.1.7	OE4198F	
R00586	L-Serine + Acetyl-CoA => O-Acetyl-L-serine + CoA	2.3.1.30	OE3122F	
R03243	3-(Imidazol-4-yl)-2-oxopropyl phosphate + L-Glutamate <=> L-Histidinol phosphate + 2-Oxoglutarate	2.6.1.9	OE2507R	PMID:3700337
R00585	Hydroxypyruvate + L-Alanine <=> L-Serine + Pyruvate	2.6.1.51	OE4391F	
R03145	Ferricytochrome b1 + Pyruvate + H2O => Ferrocytochrome b1 + Acetate + CO2	1.2.2.2		PMID:8157586
R02685	(3S)-3-Hydroxyacyl-CoA <=> trans-2,3-Dehydroacyl-CoA + H2O	4.2.1.17	OE3846R	
R01168	L-Histidine => Urocanate + NH3	4.3.1.3	OE2739F	
R06447	trans,trans,cis-Geranylgeranyl diphosphate + Isopentenyl diphosphate => di-trans,poly-cis-Undecaprenyl diphosphate + Pyrophosphate	2.5.1.31 2.5.1.-	OE3503F OE3505R OE2559R OE2650F OE4010F	PMID:2857171
R00582	O-Phospho-L-serine + H2O => L-Serine + Orthophosphate	3.1.3.3	OE4405R	
R03035	ATP + Pantetheine 4'-phosphate => Pyrophosphate + Dephospho-CoA	2.7.7.3	OE2992R	
R03530	ATP + dIDP <=> ADP + dITP	2.7.4.6	OE2667F	
R00578	ATP + L-Aspartate + L-Glutamine + H2O => AMP + Pyrophosphate + L-Asparagine + L-Glutamate	6.3.5.4	OE2278F	
R03027	(R)-3-Hydroxybutanoyl-CoA <=> Crotonoyl-CoA + H2O	4.2.1.55	OE1641R	
R00575	ATP + L-Glutamine + HCO3- + H2O => ADP + Orthophosphate + L-Glutamate + Carbamoyl phosphate	6.3.5.5	OE3554F OE3556R	
R01175	Butanoyl-CoA + FAD => FADH2 + Crotonoyl-CoA	1.3.3.6 1.3.99.2 1.3.99.3	OE2013R OE2710F OE1555F OE2138F OE3123R OE4500R	
R01177	Acetyl-CoA + Butanoyl-CoA <=> CoA + 3-Oxohexanoyl-CoA	2.3.1.9 2.3.1.16	OE3884F	
R03665	ATP + L-Valine + tRNA(Val) <=> AMP + Pyrophosphate + L-Valyl-tRNA(Val)	6.1.1.9	OE4572R	
R00674	L-Serine + Indole => L-Tryptophan + H2O	4.2.1.20	OE1470F OE1471F	
R00673	L-Tryptophan + H2O => Indole + Pyruvate + NH3	4.1.99.1	OE4331R	
R01171	Butanoyl-CoA + NAD+ <=> Crotonoyl-CoA + NADH + H+	1.3.99.2	OE2013R OE1555F OE2138F OE2710F OE3123R OE4500R	
R03132	O-Acetyl-L-serine + Thiosulfate => S-Sulfo-L-cysteine + Acetate	2.5.1.47 2.5.1.48	OE1916F OE2860R OE2173F OE2681F	

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00425	GTP + H2O => Formate + 2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine + Pyrophosphate	3.5.4.25		
R00420	UDP-N-acetyl-D-glucosamine <=> UDP-N-acetyl-D-mannosamine	5.1.3.14	OE4703R	
R03067	2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine + 4-Aminobenzoate => Pyrophosphate + Dihydropteroate	2.5.1.15	OE1615R	
R00428	GTP + H2O => Formamidopyrimidine nucleoside triphosphate	3.5.4.16		PMID:17032654
R00570	ATP + CDP <=> ADP + CTP	2.7.4.6	OE2667F	
R00571	ATP + UTP + NH3 => ADP + Orthophosphate + CTP	6.3.4.2	OE3572R	
R02557	Deoxyadenosine + Orthophosphate <=> Adenine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.1		PMID:9457844
R00573	ATP + UTP + L-Glutamine + H2O => ADP + Orthophosphate + CTP + L-Glutamate	6.3.4.2	OE3572R	
R01698	Dihydrolipoamide + NAD+ <=> Lipoamide + NADH	1.8.1.4	OE4116F	PMID:6266826
R03905	Glutaminyl-tRNA + L-Glutamate + Orthophosphate + ADP <=> L-Glutamyl-tRNA(Gln) + L-Glutamine + ATP	6.3.5.7	OE3591F	
R00566	L-Arginine => Agmatine + CO2	4.1.1.19	OE3803R	
R00568	CTP + H2O => UTP + NH3	3.5.4.13	OE1384F	
R03260	O-Succinyl-L-homoserine + L-Cysteine => L-Cystathionine + Succinate	2.5.1.48	OE2173F OE2681F	
R00430	GDP + Phosphoenolpyruvate => GTP + Pyruvate	2.7.1.40	OE1495R	
R02662	S-(2-Methylpropanoyl)-dihydrolipoamide + CoA => 2-Methylpropanoyl-CoA + Dihydrolipoamide	2.3.1.-		
R02661	2-Methylpropanoyl-CoA + FAD <=> 2-Methylprop-2-enoyl-CoA + FADH2	1.3.99.3	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R	
R02660	2-Methylpropanoyl-CoA + Oxygen => 2-Methylprop-2-enoyl-CoA + H2O	1.3.99.2	OE2013R OE1555F OE2138F OE2710F OE3123R OE4500R	
R03696	L-threo-3-Methylaspartate <=> Mesaconate + NH3	4.3.1.2	OE4207F	
R03509	N-(5-Phospho-D-ribosyl)anthranilate <=> 1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate	5.3.1.24	OE3333R	
R03508	1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate => Indoleglycerol phosphate + CO2 + H2O	4.1.1.48	OE1469F	
R00414	UDP-N-acetyl-D-glucosamine + H2O <=> N-Acetyl-D-mannosamine + UDP	5.1.3.14	OE4703R	
R04798	zeta-Carotene <=> Neurosporene + Hydrogen	1.14.99.30	OE1426F OE3381R OE3468R	
R04799	zeta-Carotene + NADP+ <=> Lycopene + NADPH	1.14.99.- -	OE3381R OE3468R	
R00410	Succinyl-CoA + Acetoacetate <=> Succinate + Acetoacetyl-CoA	2.8.3.5	OE5189F OE4211F	
R03503	ATP + 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine => AMP + 2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine	2.7.6.3		
R06613	5,10-Methylenetetrahydrofolate + dUMP + FADH2 => Tetrahydrofolate + dTMP + FAD	2.1.1.148	OE2898R	PMID:12423760
R02568	D-Fructose 1-phosphate <=> Glycerone phosphate + D-Glyceraldehyde	4.1.2.13	OE2019F	
R03504	2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- <=> Glycolaldehyde + 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	4.1.2.25		
R03815	Dihydrolipoylprotein + NADP+ <=> Lipoylprotein + NADPH	1.8.1.4	OE4116F	

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00552	L-Arginine + H2O => L-Citrulline + NH3	3.5.3.6	OE5208R	PMID:6933439 PMID:7868583 PMID:8759859
R02571	Glutaryl-CoA + Dihydrolipoamide <=> CoA + S-Glutaryldihydrolipoamide	2.3.1.61		PMID:3700337
R01290	L-Serine + L-Homocysteine => L-Cystathionine + H2O	4.2.1.22	OE1916F OE2860R	
R01196	Oxidized ferredoxin + Pyruvate + CoA => Reduced ferredoxin + Acetyl-CoA + CO2	1.2.7.1	OE2622R OE2623R	PMID:6266826 PMID:6266827 PMID:8157586 PMID:9827332
R06558	Adenosyl cobinamide + GTP => Adenosyl cobinamide phosphate + GDP	2.7.1.156		
R01197	Reduced ferredoxin + Succinyl-CoA + CO2 <=> Oxidized ferredoxin + 2-Oxoglutarate + CoA	1.2.7.3	OE1710R OE1711R	PMID:5780095 PMID:6266826 PMID:6266827 PMID:9827332
R04620	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H2O => 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- + Orthophosphate	3.1.3.1	OE5192R	
R04621	Dihydroneopterin phosphate + H2O => 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- + Orthophosphate	3.6.1.-		
R01199	Oxidized ferredoxin + 2-Oxobutanoate + CoA <=> Reduced ferredoxin + Propanoyl-CoA + CO2	1.2.7.1 1.2.7.2	OE2622R OE2623R	PMID:6816594 FEBS Letters 1980, 118:271-273. FEBS Letters 1977, 83:197-201
R02577	Lactaldehyde + NADPH + H+ <=> Propane-1,2-diol + NADP+	1.1.1.21	OE1698R OE2451R OE2486F OE3312R	
R06411	cis-2-Methyl-5-isopropylhexa-2,5-dienoyl-CoA + H2O => 3-Hydroxy-2,6-dimethyl-5-methylene-heptanoyl-CoA	4.2.1.17	OE3846R	
R06412	trans-2-Methyl-5-isopropylhexa-2,5-dienoyl-CoA + H2O => 3-Hydroxy-2,6-dimethyl-5-methylene-heptanoyl-CoA	4.2.1.17	OE3846R	
R00344	ATP + Pyruvate + HCO3- + H+ => ADP + Orthophosphate + Oxaloacetate	6.4.1.1	OE3177F	PMID:8157586 PMID:9827332
R00443	UTP => Pyrophosphate + UMP	2.7.7.6	OE4740R OE4739R OE4741R OE4742R	
R00442	CTP => Pyrophosphate + CMP	2.7.7.6	OE4740R OE4739R OE4741R OE4742R	
R00549	ATP + Riboflavin => ADP + FMN	2.7.1.26		
R00351	Acetyl-CoA + H2O + Oxaloacetate => Citrate + CoA	2.3.3.1 2.3.3.3	OE3934R	PMID:5780095 PMID:9827332
R03181	4-Maleylacetoacetate <=> 4-Fumarylacetoacetate	5.2.1.2		
R00548	FMN + H2O => Riboflavin + Orthophosphate	3.1.3.2	OE2782F	
R03823	Lycopene + NADPH => beta-Carotene + NADP+	-	OE3983R	PMID:12003928
R03778	Octanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxodecanoyl-CoA	2.3.1.16	OE3884F	
R03777	Octanoyl-CoA + FAD => trans-Oct-2-enoyl-CoA + FADH2	1.3.3.6 1.3.99.3 1.3.99.13	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R	
R03269	(R)-4'-Phosphopantothenoyl-L-cysteine => Pantetheine 4'-phosphate + CO2	4.1.1.36	OE1856R	
R00439	GDP => Orthophosphate + GMP	2.7.7.8		The Enzymes, 3rd. Ed. (Boyer, P.D., ed.) 15B, 517-553
R00332	ATP + GMP <=> ADP + GDP	2.7.4.8		

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R04639	2,5-Diamino-6-(5'-triphosphoryl-3',4'-trihydroxy-2'-oxopentyl)- $<=>$ 2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H_2O	3.5.4.16		PMID:17032654
R03425	Glycine + Lipoylprotein $=>$ S-Aminomethyldihydrolipoylprotein + CO_2	1.4.4.2	OE3274R OE3275R	
R04638	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H_2O $=>$ Dihydroneopterin phosphate + Pyrophosphate	3.6.1.-		
R00438	UDP $=>$ Orthophosphate + UMP	2.7.7.8		The Enzymes, 3rd. Ed. (Boyer, P.D., ed.) 15B, 517-553
R00437	ADP $=>$ Orthophosphate + AMP	2.7.7.8		The Enzymes, 3rd. Ed. (Boyer, P.D., ed.) 15B, 517-553
R00435	ATP $=>$ Pyrophosphate + AMP	2.7.7.6	OE4740R OE4739R OE4741R OE4742R OE4218F	
R04640	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)- $<=>$ N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5''-phospho-D-	5.3.1.16		PMID:3700337
R00440	CDP $=>$ Orthophosphate + CMP	2.7.7.8		The Enzymes, 3rd. Ed. (Boyer, P.D., ed.) 15B, 517-553
R00441	GTP $=>$ Pyrophosphate + GMP	2.7.7.6	OE4740R OE4739R OE4741R OE4742R	
R03174	S-(2-Methylbutanoyl)-dihydrolipoamide + CoA $=>$ 2-Methylbutanoyl-CoA + Dihydrolipoamide	2.3.1.-		
R03172	2-Methylbutanoyl-CoA + FAD $<=>$ 2-Methylbut-2-enoyl-CoA + FADH2	1.3.99.3	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R OE4323F	
R00342	(S)-Malate + NAD+ $<=>$ Oxaloacetate + NADH + H+	1.1.1.37		PMID:901751 PMID:901752 PMID:7297556 PMID:5780095 PMID:9827332
R03170	2-Methylbutanoyl-CoA + Oxygen $=>$ 2-Methylbut-2-enoyl-CoA + H_2O	1.3.99.2	OE2013R OE1555F OE2138F OE2710F OE3123R OE4500R OE2569R	
R00533	Sulfite + Oxygen + H_2O $=>$ Sulfate + H2O2	1.8.3.1		
R03026	Crotonoyl-CoA + H_2O $<=>$ (S)-3-Hydroxybutanoyl-CoA	4.2.1.17	OE3846R	
R03661	ATP + L-Proline + tRNA(Pro) $<=>$ AMP + Pyrophosphate + L-Prolyl-tRNA(Pro)	6.1.1.15	OE1595F	
R03662	ATP + L-Serine + tRNA(Ser) $<=>$ AMP + Pyrophosphate + L-Seryl-tRNA(Ser)	6.1.1.11	OE3893F	
R03663	ATP + L-Threonine + tRNA(Thr) $<=>$ AMP + Pyrophosphate + L-Threonyl-tRNA(Thr)	6.1.1.3	OE3580R	
R03664	ATP + L-Tryptophan + tRNA(Trp) $<=>$ AMP + Pyrophosphate + L-Tryptophanyl-tRNA(Trp)	6.1.1.2	OE4101R OE4132R	
R00671	L-Ornithine $<=>$ L-Proline + NH3	4.3.1.12	OE4121R OE2945F	
R03457	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate $<=>$ 3-(Imidazol-4-yl)-2-oxopropyl phosphate + H_2O	4.2.1.19	OE4220F	PMID:3700337
R03458	5-Amino-6-(5'-phosphoribosylamino)uracil + NADPH $=>$ 5-Amino-6-(5'-phosphoribitylamino)uracil + NADP+	1.1.1.193	OE2802F	
R03660	ATP + L-Phenylalanine + tRNA(Phe) $<=>$ AMP + Pyrophosphate + L-Phenylalanyl-tRNA(Phe)	6.1.1.20	OE4505F OE4507F	PMID:8061624
R03459	2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine + H_2O $=>$ 5-Amino-6-(5'-phosphoribosylamino)uracil + NH3	3.5.4.26		

Id	Definition	EC No.	Genes	Ref
R03659	ATP + L-Methionine + tRNA(Met) <=> AMP + Pyrophosphate + L-Methionyl-tRNA	6.1.1.10	OE1496R	
R00472	Acetyl-CoA + H2O + Glyoxylate => (S)-Malate + CoA	2.3.3.9		PMID:5780095
R03658	ATP + L-Lysine + tRNA(Lys) <=> AMP + Pyrophosphate + L-Lysyl-tRNA	6.1.1.6	OE3826F	
R00529	ATP + Sulfate => Pyrophosphate + Adenylylsulfate	2.7.7.4	OE1684F	
R03165	Hydroxymethylbilane <=> Uroporphyrinogen III + H2O	4.2.1.75	OE4281F	
R06529	Adenosyl cobyrinate hexaamide + D-1-Aminopropan-2-ol O-phosphate + ATP => Adenosyl cobinamide phosphate + ADP + Orthophosphate	6.3.1.10	OE3253F OE3259F	PMID:14645280
R03460	Phosphoenolpyruvate + Shikimate 3-phosphate <=> Orthophosphate + 5-O-(1-Carboxyvinyl)-3-phosphoshikimate	2.5.1.19	OE2762R	PMID:3700337
R03655	ATP + L-Histidine + tRNA(His) => AMP + Pyrophosphate + L-Histidyl-tRNA(His)	6.1.1.21	OE3812R	
R03654	ATP + Glycine + tRNA(Gly) <=> AMP + Pyrophosphate + Glycyl-tRNA(Gly)	6.1.1.14	OE4307F	
R03657	ATP + L-Leucine + tRNA(Leu) <=> AMP + Pyrophosphate + L-Leucyl-tRNA	6.1.1.4	OE4118R	
R03947	Precorrin 2 + NAD+ => Sirohydrochlorin + NADH + H+	2.1.1.107 1.3.1.76	OE4277F OE3498R	
R03656	ATP + L-Isoleucine + tRNA(Ile) <=> AMP + Pyrophosphate + L-Isoleucyl-tRNA(Ile)	6.1.1.5	OE4068F	
R03758	L-2-Amino-3-oxobutanoate => Aminoacetone + CO2	-		
R00330	ATP + GDP <=> ADP + GTP	2.7.4.6	OE2667F	
R03018	ATP + Pantothenate => ADP + D-4'-Phosphopantothenate	2.7.1.33		
R03759	Aminoacetone + NADH + H+ <=> (R)-1-Aminopropan-2-ol + NAD+	1.1.1.75		
R03013	L-Histidinol phosphate + H2O => L-Histidinol + Orthophosphate	3.1.3.15	OE2276F	
R03012	L-Histidinol + NAD+ => L-Histidinal + NADH + H+	1.1.1.23	OE3071F	PMID:3700337
R00316	ATP + Acetate => Pyrophosphate + Acetyl adenylate	6.2.1.1	OE2450F OE1726F	
R03651	ATP + L-Glutamate + tRNA(Gln) <=> AMP + Pyrophosphate + L-Glutamyl-tRNA(Gln)	6.1.1.17	OE2652F	
R00658	2-Phospho-D-glycerate <=> Phosphoenolpyruvate + H2O	4.2.1.11	OE2640F	PMID:2256927 PMID:3255682 PMID:6642538 PMID:9827332
R00516	UTP + Cytidine => UDP + CMP	2.7.1.48	OE2749F	
R03647	ATP + L-Aspartate + tRNA(Asn) <=> AMP + Pyrophosphate + L-Aspartyl-tRNA(Asn)	6.1.1.12	OE1687F	
R00517	GTP + Cytidine => GDP + CMP	2.7.1.48	OE2749F	
R03646	ATP + L-Arginine + tRNA(Arg) <=> AMP + Pyrophosphate + L-Arginyl-tRNA(Arg)	6.1.1.19	OE5203F	
R00654	L-Methionine + H2O => Methanethiol + NH3 + 2-Oxobutanoate	4.4.1.11	OE2173F OE2681F OE4398F	PMID:8206953
R00651	O-Acetyl-L-homoserine + Methanethiol => L-Methionine + Acetate	2.5.1.49	OE4398F	PMID:8206953
R00511	CMP + H2O => Cytidine + Orthophosphate	3.1.3.5	OE3017R	
R03858	Lauroyl-CoA + Acetyl-CoA <=> CoA + 3-Oxotetradecanoyl-CoA	2.3.1.16	OE3884F	
R03857	Lauroyl-CoA + FAD => 2-trans-Dodecenoyl-CoA + FADH2	1.3.3.6 1.3.99.3 1.3.99.13	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R	
R04520	Geranylgeranyl diphosphate + sn-3-O-(Geranylgeranyl)glycerol 1-phosphate <=> Pyrophosphate + 2,3-Bis-O-(geranylgeranyl)glycerol 1-phosphate	2.5.1.42		PMID:16788058

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00512	ATP + CMP <=> ADP + CDP	2.7.4.14	OE3429F	
R03005	ATP + Nicotinate D-ribonucleotide => Pyrophosphate + Deamino-NAD+	2.7.7.1 2.7.7.18	OE1462R	
R00513	ATP + Cytidine => ADP + CMP	2.7.1.48	OE2749F	
R01354	ATP + Propanoate => Pyrophosphate + Propinol adenylate	6.2.1.1	OE2450F OE1726F	
R05218	Cob(II)yrinate a,c diamide + H+ + NADH <=> Cob(I)yrinate a,c diamide + NAD+	1.16.8.1		
R00220	L-Serine => Pyruvate + NH3	4.3.1.17 4.3.1.19	OE3931R	
R04463	ATP + 5'-Phosphoribosyl-N-formylglycinamide + L-Glutamine + H2O => ADP + Orthophosphate + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine + L-Glutamate	6.3.5.3	OE2274R OE3731R	
R90107	L-Proline (E) + Sodium (E) [pmf] <=> L-Proline + Sodium		OE2322F	PMID:889797
R90106	Daunorubicin + ATP + H2O => Daunorubicin (E) + Orthophosphate + ADP + H+		OE1704R	
R04560	10-Formyltetrahydrofolate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide <=> Tetrahydrofolate + 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	2.1.2.3	OE1620R	
R90105	ATP + H2O + Copper (E) => ADP + Orthophosphate + H+ + Copper		OE2042F	
R90104	Arsenite (E) + ATP + H2O => Arsenite + ADP + Orthophosphate + H+		OE1547R OE4649F OE5440F	
R90103	Vitamin B12 (E) + ATP + H2O => Vitamin B12 + ADP + Orthophosphate + H+		OE1480F	
R90102	3-Phospho-D-Glycerate (E) + ATP + H2O => 3-Phospho-D-glycerate + ADP + Orthophosphate + H+		OE1420F	
R90101	Sodium + H+ (pmf associated) <=> H+ + Sodium (E) [pmf]		OE3960F OE5204R OE3961R	PMID:35540
R90100	Nitrate (E) + Photon => Nitrate		OE1299R	PMID:9233791
R04859	O-Acetyl-L-serine + Thiosulfate + Thioredoxin + H+ => L-Cysteine + Sulfite + Oxidized thioredoxin + Acetate	2.5.1.47 2.5.1.49	OE1916F OE2860R OE4398F	
R04559	1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole <=> Fumarate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	4.3.2.2	OE1623F	
R04558	N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5''-phospho-D- + L-Glutamine <=> 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + L-Glutamate + D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate	2.4.2.-	OE2268R OE3913F	
R04170	(S)-3-Hydroxydodecanoyl-CoA <=> 2-trans-Dodecenoyl-CoA + H2O	4.2.1.17	OE3846R	
R03346	Nicotinate D-ribonucleotide + H2O => Nicotinate D-ribonucleoside + Orthophosphate	3.1.3.5	OE3017R	
R90108	Cl- (E) + H+ (pmf associated) => Cl- + H+		OE3200R	
R04173	O-Phospho-L-serine + 2-Oxoglutarate <=> 3-Phosphonooxypyruvate + L-Glutamate	2.6.1.52	OE4391F	
R03348	Pyridine-2,3-dicarboxylate + 5-Phospho-alpha-D-ribose 1-diphosphate => Nicotinate D-ribonucleotide + Pyrophosphate + CO2	2.4.2.19	OE3648F	
R90109	Pantothenate (E) + Sodium (E) [pmf] => Pantothenate + Sodium		OE3511F	
R05221	Adenosyl cobinamide + ATP => Adenosyl cobinamide phosphate + ADP	2.7.1.156		
R05222	Adenosyl cobinamide phosphate + GTP => Adenosine-GDP-cobinamide + Pyrophosphate	2.7.7.62	OE3227F	
R00216	(S)-Malate + NADP+ => Pyruvate + CO2 + NADPH	1.1.1.38 1.1.1.39	OE3308F	PMID:8157586 PMID:9827332

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00217	Oxaloacetate => Pyruvate + CO2	1.1.1.40	OE3308F	
R05220	Cob(I)yrinate a,c diamide + ATP => Adenosyl cobyrinate a,c diamide + Pyrophosphate + Orthophosphate	2.5.1.17	OE3245F	
R05225	Adenosyl cobyrinate a,c diamide + L-Glutamine + ATP + H2O => Adenosyl cobyrinate hexaamide + L-Glutamate + Orthophosphate + ADP	6.3.5.10	OE3246F	PMID:14645280
R05226	Adenosyl cobyrinate hexaamide + (R)-1-Aminopropan-2-ol <=> Adenosyl cobinamide	6.3.1.10	OE3253F OE3259F	PMID:14645280
R04457	4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine + 3, 4-Dihydroxy-2-butanone 4-phosphate => 6,7-Dimethyl-8-(1-D-ribityl)lumazine + Orthophosphate + H2O	2.5.1.9	OE1946R OE4683F	
R05223	Adenosine-GDP-cobinamide + alpha-Ribazole => Cobamide coenzyme + GMP	2.7.8.26	OE3255F	
R00908	beta-Alanine + 2-Oxoglutarate <=> 3-Oxopropanoate + L-Glutamate	2.6.1.19	OE3168R OE5094F	
R00214	(S)-Malate + NAD+ => Pyruvate + CO2 + NADH	1.1.1.38 1.1.1.39	OE3308F	
R01364	4-Fumarylacetoacetate + H2O => Acetoacetate + Fumarate	3.7.1.2		
R01360	(S)-3-Hydroxy-3-methylglutaryl-CoA => Acetyl-CoA + Acetoacetate	4.1.3.4	OE3296F	
R00502	UTP + alpha-D-Galactose 1-phosphate <=> Pyrophosphate + UDP-D-galactose	2.7.7.10	OE1578F OE1016R OE1078F OE1114F OE2530F OE1014R	
R01220	5,10-Methylenetetrahydrofolate + NADP+ <=> 5,10-Methenyltetrahydrofolate + NADPH	1.5.1.5	OE3038F	
R01221	5,10-Methylenetetrahydrofolate + NH3 + CO2 + NADH + H+ <=> Glycine + Tetrahydrofolate + NAD+	2.1.2.10	OE3278R	
R00509	ATP + Adenylylsulfate => ADP + 3'-Phosphoadenylyl sulfate	2.7.1.25		
R01214	L-Valine + 2-Oxoglutarate <=> 3-Methyl-2-oxobutanoate + L-Glutamate	2.6.1.42 2.6.1.6	OE3959R	
R04740	(S)-3-Hydroxytetradecanoyl-CoA <=> trans-Tetradec-2-enoyl-CoA + H2O	4.2.1.17	OE3846R	
R00615	Thiamin diphosphate + H2O => Thiamin monophosphate + Orthophosphate	3.6.1.15	OE3863R	
R04742	Decanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxododecanoyl-CoA	2.3.1.16	OE3884F	
R00617	ATP + Thiamin monophosphate => ADP + Thiamin diphosphate	2.7.4.16	OE3818F	
R04741	(S)-3-Hydroxydodecanoyl-CoA + NAD+ <=> 3-Oxododecanoyl-CoA + NADH	1.1.1.35	OE2015R	
R03471	ATP + 4-Amino-5-hydroxymethyl-2-methylpyrimidine => ADP + 4-Amino-2-methyl-5-phosphomethylpyrimidine	2.7.1.49	OE4654F	
R03472	Aminoimidazole ribotide => 4-Amino-5-hydroxymethyl-2-methylpyrimidine		OE2057F	
R01230	ATP + Xanthosine 5'-phosphate + NH3 => AMP + Pyrophosphate + GMP	6.3.4.1 6.3.5.2	OE1363F OE3571R	
R00610	Sarcosine + H2O + Oxygen => Glycine + Formaldehyde + H2O2	1.5.3.1	OE1432F	
R01231	ATP + Xanthosine 5'-phosphate + L-Glutamine + H2O => AMP + Pyrophosphate + GMP + L-Glutamate	6.3.5.2	OE1363F OE3571R	
R01229	Guanine + 5-Phospho-alpha-D-ribose 1-diphosphate => GMP + Pyrophosphate	2.4.2.7	OE1504F	PMID:9457844
R01228	ATP + Guanosine => ADP + GMP	2.7.1.73	OE3606R	PMID:9457844
R01227	GMP + H2O => Guanosine + Orthophosphate	3.1.3.5	OE3017R	
R01226	5,10-Methylenetetrahydrofolate + 3-Methyl-2-oxobutanoate + H2O => Tetrahydrofolate + 2-Dehydropantoate	2.1.2.11	OE3119R	

Continued on next page

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R01325	Citrate <=> cis-Aconitate + H2O	4.2.1.3 4.2.1.4	OE4613F	PMID:5780095 PMID:9827332
R04737	(S)-3-Hydroxyhexadecanoyl-CoA + NAD+ <=> 3-Oxopalmitoyl-CoA + NADH	1.1.1.35	OE2015R OE2871F OE3846R	
R01324	Citrate <=> Isocitrate	4.2.1.3	OE4613F	PMID:5780095 PMID:9827332
R04738	(S)-3-Hydroxyhexadecanoyl-CoA <=> trans-Hexadec-2-enoyl-CoA + H2O	4.2.1.17	OE3846R OE1641R	
R04739	(S)-3-Hydroxytetradecanoyl-CoA + NAD+ <=> 3-Oxotetradecanoyl-CoA + NADH	1.1.1.35	OE2015R	
R04591	ATP + 1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate + L-Aspartate <=> ADP + Orthophosphate + 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole	6.3.2.6	OE3724F	
R04751	Hexanoyl-CoA + FAD => trans-Hex-2-enoyl-CoA + FADH2	1.3.3.6 1.3.99.2 1.3.99.3 1.3.99.13	OE2013R OE2710F OE1555F OE2138F OE3123R OE4500R	
R04594	N1-(5-Phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole + H2O => alpha-Ribazole + Orthophosphate	3.1.3.73		
R00622	2-Oxoglutarate + Thiamin diphosphate => Succinate semialdehyde-thiamin diphosphate anion + CO2	2.5.1.64	OE2563R	
R00104	ATP + NAD+ => ADP + NADP+	2.7.1.23	OE3671F	
R04749	(S)-Hydroxyhexanoyl-CoA <=> trans-Hex-2-enoyl-CoA + H2O	4.2.1.17	OE3846R	
R00203	Methylglyoxal + NAD+ + H2O => Pyruvate + NADH + H+	1.2.1.22	OE4529F OE2133R OE2190R OE2367F	
R04747	Hexanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxooctanoyl-CoA	2.3.1.16	OE3884F	
R04391	ATP + N-((R)-Pantothenoyl)-L-cysteine => ADP + (R)-4'-Phosphopantothenoyl-L-cysteine	2.7.1.33		
R00200	ADP + Phosphoenolpyruvate => ATP + Pyruvate	2.7.1.40	OE1495R	PMID:3255682
R04748	(S)-Hydroxyhexanoyl-CoA + NAD+ <=> 3-Oxohexanoyl-CoA + NADH	1.1.1.35	OE2015R	
R04745	(S)-Hydroxyoctanoyl-CoA + NAD+ <=> 3-Oxooctanoyl-CoA + NADH	1.1.1.35	OE2015R	
R00206	ATP + Pyruvate + Orthophosphate => AMP + Phosphoenolpyruvate + Pyrophosphate	2.7.9.1	OE1500R	PMID:2256927 PMID:9827332
R03102	L-2-Aminoadipate 6-semialdehyde + NAD+ + H2O => L-2-Aminoadipate + NADH + H+	1.2.1.31		PMID:3700337
R04746	(S)-Hydroxyoctanoyl-CoA <=> trans-Oct-2-enoyl-CoA + H2O	4.2.1.17	OE3846R	
R04743	(S)-Hydroxydecanoyl-CoA + NAD+ <=> 3-Oxodecanoyl-CoA + NADH	1.1.1.35	OE2015R	
R04744	(S)-Hydroxydecanoyl-CoA <=> trans-Dec-2-enoyl-CoA + H2O	4.2.1.17	OE3846R	
R01446	(S)-Lactaldehyde + NAD+ + H2O => (S)-Lactate + NADH	1.2.1.22	OE4529F OE2133R OE2190R OE2367F	
R90148	GAP (E) + H+ (pmf associated) <=> D-Glyceraldehyde 3-phosphate + H+			
R90149	Fumarate (E) + H+ (pmf associated) <=> Fumarate + H+			
R90146	ATP + H2O + Molybdenum (E) => ADP + Orthophosphate + H+ + Molybdenum			
R90000	NADH + Menaquinone + H+ => NAD+ + Menaquinol + H+ (pmf associated)	1.6.5.3	OE1953F OE1954F OE1956F OE1957F OE1958F OE1959F OE1960F OE1963F OE1964F OE1965F OE1967F OE4005F OE4007F	PMID:9932647
R90147	(S)-Malate (E) + H+ (pmf associated) <=> (S)-Malate + H+			
R90144	Isopentenyl phosphate + ATP => Isopentenyl diphosphate + ADP		OE2647F	PMID:1621811 PMID:3700337

Continued on next page

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R01253	(S)-1-Pyrroline-5-carboxylate + NADPH <=> L-Proline + H2O + NADP+	1.5.99.8	OE3955F	PMID:9827332
R90145	Pyruvate (E) + H+ (pmf associated) <=> Pyruvate + H+			
R90142	Molybdenum cofactor + GTP => MoCo dinucleotide (Guanine) + Pyrophosphate		OE2217R OE2219R	
R01398	Carbamoyl phosphate + L-Ornithine <=> Orthophosphate + L-Citrulline	2.1.3.3	OE5205R	PMID:7868583 PMID:8759859
R90003	ADP + Orthophosphate + H+ (pmf associated) <=> ATP + H2O + H+	3.6.3.14	OE3984R OE3985R OE3986R OE3987R OE3988R OE3989R OE3991R OE3992R OE3978R	PMID:21098
R01397	Carbamoyl phosphate + L-Aspartate => Orthophosphate + N-Carbamoyl-L-aspartate	2.1.3.2	OE5201F OE5202F	
R90004	Reduced ferredoxin + Menaquinone + H+ => Oxidized ferredoxin + Menaquinol + H+ (pmf associated)	1.6.5.3	OE1953F OE1954F OE1956F OE1957F OE1958F OE1959F OE1960F OE1963F OE1964F OE1965F OE1967F OE4005F OE4007F	
R90143	(R)-5-Phosphomevalonate + ATP => Isopentenyl phosphate + ADP + Orthophosphate + CO2		OE1893F	PMID:16621811 PMID:3700337
R90140	Cyclic pyranopterin monophosphate => Molybdopterin + Pyrophosphate		OE1139F OE3595R OE1140R	PMID:15716436
R90001	Menaquinol + H+ + Halocyanin (oxidized) => Menaquinone + H+ (pmf associated) + Halocyanin (reduced)	1.10.2.2	OE1872R OE1874R OE1876R	PMID:9932647
R90002	Oxygen + H+ + Halocyanin (reduced) => H2O + H+ (pmf associated) + Halocyanin (oxidized)	1.9.3.1	OE1979R OE1984F OE1988R OE4070R OE4071R	PMID:2542239
R90141	Molybdopterin + Molybdenum => Molybdenum cofactor		OE1146R OE1143R	
R90005	H+ + Photon => H+ (pmf associated)		OE3106F	PMID:9729742
R04754	Decanoyl-CoA + FAD => trans-Dec-2-enoyl-CoA + FADH2	1.3.3.6 1.3.99.3 1.3.99.13	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R	
R04224	2-Methylprop-2-enoyl-CoA + H2O <=> (S)-3-Hydroxyisobutyryl-CoA	4.2.1.17	OE3846R	
R04225	3-Methyl-2-oxopentanoate + Lipoamide => S-(2-Methylbutanoyl)-dihydrolipoamide + CO2	1.2.4.4	OE4114F OE3712R OE4113F	
R90138	Hexadecanoic acid (E) + ATP + H2O => Hexadecanoic acid + ADP + Orthophosphate + H+			
R04326	5'-Phosphoribosylglycinamide + 5,10-Methenyltetrahydrofolate + H2O <=> 5'-Phosphoribosyl-N-formylglycinamide + Tetrahydrofolate	2.1.2.2	OE1620R	
R90139	GTP + S-Adenosyl-L-methionine => Cyclic pyranopterin monophosphate + S-Adenosyl-L-homocysteine		OE1142F OE2825F	PMID:15716436
R04325	10-Formyltetrahydrofolate + 5'-Phosphoribosylglycinamide => Tetrahydrofolate + 5'-Phosphoribosyl-N-formylglycinamide	2.1.2.2	OE1620R	
R90133	6-Deoxy-5-ketomannitol-1-phosphate + L-Aspartate 4-semialdehyde => Compound I + D-Glyceraldehyde 3-phosphate		OE1472F	PMID:15182204
R90134	Inosine (E) + H+ (pmf associated) => Inosine + H+			PMID:9457844
R03224	(3R)-3-Hydroxyacyl-CoA <=> cis-2,3-Dehydroacyl-CoA + H2O	4.2.1.17	OE3846R	
R90135	Menaquinol + Oxygen => Menaquinone + H2O2	1.10.3.-	OE7065F OE6185F OE7066F OE6186F	
R90136	Methanethiol <=> Methanethiol (E)			PMID:8206953

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R90130	4-Ketofructose-1,6-bisphosphate + NADH => 6-Deoxy-5-Ketofructose 1 phosphate + Orthophosphate + NAD+			PMID:15182204
R90131	2-amino-3,7-dideoxy-hept-6-ulsonic acid => 3-Dehydroquinate		OE1475F	PMID:15182204
R90132	6-Deoxy-5-Ketofructose 1 phosphate + NAD+ <=> 6-Deoxy-5-ketomannitol-1-phosphate + NADH			PMID:15182204
R04233	CTP + D-4'-Phosphopantothenate + L-Cysteine => CDP + Orthophosphate + (R)-4'-Phosphopantothenoyl-L-cysteine	6.3.2.5	OE1856R	
R04231	CTP + D-4'-Phosphopantothenate + L-Cysteine => CMP + Pyrophosphate + (R)-4'-Phosphopantothenoyl-L-cysteine	6.3.2.5	OE1856R	
R03314	L-Glutamate 5-semialdehyde => (S)-1-Pyrroline-5-carboxylate + H2O	-		PMID:9827332
R04230	ATP + D-4'-Phosphopantothenate + L-Cysteine => ADP + Orthophosphate + (R)-4'-Phosphopantothenoyl-L-cysteine	6.3.2.5	OE1856R	
R03217	O-Acetyl-L-homoserine + L-Cysteine <=> L-Cystathionine + Acetate	2.5.1.49	OE4398F	
R90124	L-Citrulline (E) + H+ (pmf associated) => L-Citrulline + H+			PMID:5903088
R90125	Succinate (E) + H+ (pmf associated) <=> Succinate + H+			
R90122	Guanine (E) + H+ (pmf associated) => Guanine + H+		OE3950R	PMID:9457844
R90123	Hypoxanthine (E) + H+ (pmf associated) => Hypoxanthine + H+		OE3950R	PMID:9457844
R05341	Lycopene <=> gamma-Carotene		OE3983R	
R90128	D-Ribose 1,5-bisphosphate + NAD+ <=> D-Ribulose 1,5-bisphosphate + NADH		OE4651F	PMID:15375115
R90129	D-Fructose 1,6-bisphosphate + NAD+ => 4-Ketofructose-1,6-bisphosphate + NADH			PMID:15182204
R90126	UDP-D-galactose + H2O => D-Galactose + UDP			
R01466	O-Phospho-L-homoserine + H2O => L-Threonine + Orthophosphate	4.2.3.1	OE1807R OE3941F OE4412R	PMID:3700337
R01465	L-Threonine + NAD+ => L-2-Amino-3-oxobutanoate + NADH + H+	1.1.1.103	OE4674F	
R90127	Adenosine => Adenosine (E)			
R01567	ATP + Thymidine => ADP + dTMP	2.7.1.21	OE3159R	
R04786	Phytoene <=> Phytofluene + Hydrogen	1.14.99.-	OE3381R OE3468R OE1808F	
R90120	Manganese (E) + ATP + H2O => Manganese + ADP + Orthophosphate + H+		OE5146R OE5147R	
R90121	Adenine (E) + H+ (pmf associated) => Adenine + H+		OE3950R	PMID:9457844
R01569	dTMP + H2O => Thymidine + Orthophosphate	3.1.3.5	OE3017R	
R01373	Prephenate <=> Phenylpyruvate + H2O + CO2	4.2.1.51	OE4117F	PMID:3700337
R04095	3-Methylbutanoyl-CoA + FAD <=> 3-Methylcrotonyl-CoA + FADH2	1.3.99.3	OE2710F OE2013R OE2138F OE3123R OE4500R OE1555F	
R04203	(2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA + NAD+ <=> 2-Methylacetoacetyl-CoA + NADH	1.1.1.35	OE2015R	
R04097	S-(3-Methylbutanoyl)-dihydrolipoamide + CoA => 3-Methylbutanoyl-CoA + Dihydrolipoamide	2.3.1.-		
R90111	Zinc => Zinc (E)		OE3619R	
R01286	L-Cystathionine + H2O => L-Homocysteine + NH3 + Pyruvate	4.4.1.8	OE2173F OE2681F	
R90112	Cadmium => Cadmium (E)		OE3619R	
R01288	O-Succinyl-L-homoserine + Sulfide <=> L-Homocysteine + Succinate	2.5.1.48	OE2173F OE2681F	
R90113	Potassium + H+ (pmf associated) => Potassium (E) + H+		OE3889R	
R01287	O-Acetyl-L-homoserine + Hydrogen sulfide => L-Homocysteine + Acetate	2.5.1.49	OE4398F	

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R90114	Phosphonate (E) + ATP + H2O + Orthophosphate => Phosphonate + ADP + H+		OE3908R OE3907R OE3910R	
R90116	Oligopeptide (E) + ATP + H2O => Oligopeptide + ADP + Orthophosphate + H+		OE4302R OE4317F	PMID:10064582
R90117	Dipeptide (E) + ATP + H2O => Dipeptide + ADP + Orthophosphate + H+		OE4303R OE4555F OE4304R OE4305R OE4550F OE4551F OE4552F	PMID:10064582
R90118	Ferrichrome (E) + ATP + H2O => Ferrichrome + ADP + Orthophosphate + H+		OE4577F OE4579F	
R01280	ATP + Hexadecanoic acid + CoA <=> AMP + Palmitoyl-CoA + Pyrophosphate	6.2.1.3	OE2912F OE3891R OE1271F	
R04204	2-Methylbut-2-enoyl-CoA + H2O <=> (2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA	4.2.1.17	OE3846R	
R01385	UDPglucuronate <=> UDP-D-galacturonate	5.1.3.6	OE2110R OE1113R	
R04209	1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate <=> Aminoimidazole ribotide + CO2	4.1.1.21	OE1951F OE1952F	
R04208	ATP + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine => ADP + Orthophosphate + Aminoimidazole ribotide	6.3.3.1	OE2292F	
R90110	Betaine (E) => Betaine			
R04993	2-Demethylmenaquinone + S-Adenosyl-L-methionine => Menaquinol + S-Adenosyl-L-homocysteine	2.1.1.-	OE3994F	
R04992	Succinate semialdehyde-thiamin diphosphate anion + Isochorismate => 2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate + Thiamin diphosphate + Pyruvate	2.5.1.64	OE2563R	
R04787	Phytofluene <=> zeta-Carotene	1.14.99.-	OE3381R OE3468R OE1808F	
R04212	L-Asparaginyl-tRNA(Asn) + L-Glutamate + Orthophosphate + ADP <=> L-Aspartyl-tRNA(Asn) + L-Glutamine + ATP	6.3.5.6	OE1522F OE2283F OE2284F	PMID:10993083
R90119	Zinc (E) + ATP + H2O => Zinc + ADP + Orthophosphate + H+		OE5146R OE5147R	
R01279	Palmitoyl-CoA + FAD => trans-Hexadec-2-enoyl-CoA + FADH2	1.3.3.6 1.3.99.3 1.3.99.13 2.7.4.14	OE2710F OE1555F OE2013R OE2138F OE3123R OE4500R OE3429F	
R00158	ATP + UMP <=> ADP + UDP			
R02022	dCTP + Oxidized thioredoxin + H2O <=> CTP + Thioredoxin	1.17.4.2	OE3328R OE4345R OE4346R	
R02023	dUTP + Oxidized thioredoxin + H2O <=> UTP + Thioredoxin	1.17.4.2	OE3328R OE4345R OE4346R	
R02024	Thioredoxin + CDP => dCDP + Oxidized thioredoxin + H2O	1.17.4.1	OE3328R OE4345R OE4346R	PMID:8990160
R00156	ATP + UDP <=> ADP + UTP	2.7.4.6	OE2667F	
R02971	ATP + Pantetheine => ADP + Pantetheine 4'-phosphate	2.7.1.33 2.7.1.34		
R01157	Agmatine + H2O => Putrescine + Urea	3.5.3.11	OE3486R	
R02020	dGTP + Oxidized thioredoxin + H2O <=> GTP + Thioredoxin	1.17.4.2	OE3328R OE4345R OE4346R	
R00985	Chorismate + NH3 => Anthranilate + Pyruvate + H2O	4.1.3.27	OE1568F OE1570F OE1573A1F OE3331R OE3332R	
R00986	Chorismate + L-Glutamine => Anthranilate + Pyruvate + L-Glutamate	4.1.3.27	OE1568F OE1570F OE1573A1F OE3331R OE3332R	
R02021	Thioredoxin + 3'-Phosphoadenylyl sulfate => Oxidized thioredoxin + Sulfite + Adenosine 3',5'-bisphosphate + H+	1.8.4.8		
R01015	D-Glyceraldehyde 3-phosphate <=> Glycerone phosphate	5.3.1.1	OE2500R	PMID:2256927 PMID:9827332

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R02164	Succinate + Menaquinone <=> Fumarate + Menaquinol	1.3.5.1	OE2865R OE2866R OE2867R OE2868R	PMID:11803024 PMID:4004256 PMID:5780095 PMID:9827332
R00299	ATP + D-Glucose => ADP + D-Glucose 6-phosphate	2.7.1.2 2.7.1.1	OE4691R	PMID:2256927 PMID:3255682 PMID:6642538
R02805	P1,P4-Bis(5'-xanthosyl) tetraphosphate + H2O => XTP + Xanthosine 5'-phosphate	3.6.1.17	OE1648R	
R90046	Glycine (E) + H+ (pmf associated) <=> Glycine + H+		OE2560R	PMID:889797
R90045	L-Leucine (E) + H+ (pmf associated) <=> L-Leucine + H+			PMID:10779875 PMID:889797
R00705	3-Oxopropanoate + CoA + NAD+ => Acetyl-CoA + CO2 + NADH + H+	1.2.1.18	OE4529F OE2367F OE2190R OE2133R	
R01547	ATP + dAMP <=> ADP + dADP	2.7.4.3	OE3425F	
R01548	dATP + Cytidine => dADP + CMP	2.7.1.48	OE2749F	
R00704	(R)-Lactate + NAD+ <=> Pyruvate + NADH + H+	1.1.1.28	OE4399F	PMID:8157586
R90048	L-Valine (E) + H+ (pmf associated) <=> L-Valine + H+			PMID:889797
R01549	dATP + Uridine => dADP + UMP	2.7.1.48	OE2749F	
R00703	(S)-Lactate + NAD+ <=> Pyruvate + NADH + H+	1.1.1.27	OE4036R OE4021F OE1778R OE2916F	PMID:7736359 PMID:8157586 PMID:9827332
R90047	L-Threonine (E) + H+ (pmf associated) => L-Threonine + H+			PMID:889797
R90049	L-Alanine (E) + H+ (pmf associated) <=> L-Alanine + H+			PMID:889797
R00708	(S)-1-Pyrroline-5-carboxylate + NADP+ + H2O => L-Glutamate + NADPH + H+	1.5.1.12	OE2133R OE2190R OE2367F OE4529F	PMID:9827332
R05595	Crotonoyl-CoA + H2O <=> 3-Hydroxybutanoyl-CoA	4.2.1.17	OE3846R	
R00707	(S)-1-Pyrroline-5-carboxylate + NAD+ + H2O => L-Glutamate + NADH + H+	1.5.1.12	OE2133R OE2190R OE2367F OE4529F	PMID:9827332
R02019	GDP + Thioredoxin => dGDP + Oxidized thioredoxin + H2O	1.17.4.1	OE3328R OE4345R OE4346R	PMID:8990160
R00161	ATP + FMN => Pyrophosphate + FAD	2.7.7.2		
R00160	FAD + H2O => AMP + FMN	3.6.1.9 3.6.1.18		
R90040	L-Lysine (E) + H+ (pmf associated) <=> L-Lysine + H+		OE1288F OE5101R	PMID:889797
R00702	trans,trans-Farnesyl diphosphate => Pyrophosphate + Presqualene diphosphate + H+	2.5.1.21	OE2014F	PMID:2857171
R02016	Oxidized thioredoxin + NADPH + H+ => Thioredoxin + NADP+	1.8.1.9	OE2805R OE4227F	
R90042	L-Asparagine (E) + H+ (pmf associated) <=> L-Asparagine + H+			PMID:889797
R90041	L-Phenylalanine (E) + H+ (pmf associated) => L-Phenylalanine + H+		OE2779F	PMID:889797
R90044	L-Serine (E) + H+ (pmf associated) <=> L-Serine + H+			PMID:35540 PMID:889797
R02018	Thioredoxin + UDP => dUDP + Oxidized thioredoxin + H2O	1.17.4.1	OE3328R OE4345R OE4346R	PMID:8990160
R02017	Thioredoxin + ADP => dADP + Oxidized thioredoxin + H2O	1.17.4.1	OE3328R OE4345R OE4346R	PMID:8990160
R01163	L-Histidinal + H2O + NAD+ => L-Histidine + NADH + H+	1.1.1.23	OE3071F	PMID:3700337
R90043	L-Isoleucine (E) + H+ (pmf associated) <=> L-Isoleucine + H+			PMID:889797
R00970	ITP + Uridine => IDP + UMP	2.7.1.48	OE2749F	
R00032	beta-Carotene + Oxygen => Retinal	1.14.99.36		PMID:11092896 PMID:11226271
R01561	Adenosine + Orthophosphate <=> Adenine + alpha-D-Ribose 1-phosphate	2.4.2.1		PMID:9457844
R01939	L-2-Aminoadipate + 2-Oxoglutarate <=> 2-Oxoadipate + L-Glutamate	2.6.1.39		PMID:3700337
R90037	L-Glutamine (E) + H+ (pmf associated) => L-Glutamine + H+			PMID:889797

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R90036	L-Glutamate (E) + H+ (pmf associated) <=> L-Glutamate + H+		OE5200R	PMID:889797
R90035	L-Aspartate (E) + H+ (pmf associated) => L-Aspartate + H+		OE5200R	PMID:35540 PMID:889797
R90039	L-Histidine (E) + H+ (pmf associated) <=> L-Histidine + H+		OE1288F OE5101R	PMID:889797
R01010	Glycerone phosphate + H2O => Glycerone + Orthophosphate	3.1.3.1	OE5192R	
R90038	L-Tryptophan (E) + H+ (pmf associated) <=> L-Tryptophan + H+		OE2779F	
R90033	L-Methionine (E) + H+ (pmf associated) => L-Methionine + H+			PMID:889797
R90032	L-Tyrosine (E) + H+ (pmf associated) <=> L-Tyrosine + H+		OE2779F	PMID:889797
R90031	L-Arginine (E) + H+ (pmf associated) => L-Arginine + H+		OE1288F OE5101R	PMID:889797
R00177	ATP + L-Methionine + H2O => Orthophosphate + Pyrophosphate + S-Adenosyl-L-methionine	2.5.1.6	OE2857F	
R90030	L-Cysteine (E) + H+ (pmf associated) => L-Cysteine + H+			
R00965	Orotidine 5'-phosphate => UMP + CO2	4.1.1.23	OE3363F	PMID:10672188
R01135	GTP + IMP + L-Aspartate => GDP + Orthophosphate + N6-(1,2-Dicarboxyethyl)-AMP	6.3.4.4	OE2579F	
R00966	Uracil + 5-Phospho-alpha-D-ribose 1-diphosphate => UMP + Pyrophosphate	2.4.2.9	OE4234R	
R00963	UMP + H2O => Uridine + Orthophosphate	3.1.3.5	OE3017R	
R01138	dADP + Phosphoenolpyruvate => dATP + Pyruvate	2.7.1.40	OE1495R	
R01137	ATP + dADP <=> ADP + dATP	2.7.4.6	OE2667F	
R01940	2-Oxoadipate + Lipoamide => S-Glutaryldihydrolipoamide + CO2	1.2.4.2		PMID:3700337
R00964	ATP + Uridine => ADP + UMP	2.7.1.48	OE2749F	
R00962	ITP + Cytidine => IDP + CMP	2.7.1.48	OE2749F	
R00969	P1,P4-Bis(5'-uridyl) tetraphosphate + H2O => UTP + UMP	3.6.1.17	OE1648R	
R01036	Glycerol + NAD+ <=> D-Glyceraldehyde + NADH + H+	1.1.1.1 1.1.1.21	OE3563R OE4674F OE1698R OE2451R OE2486F OE3312R	
R00967	UTP + Uridine => UDP + UMP	2.7.1.48	OE2749F	
R01039	Glycerone + NADPH + H+ <=> Glycerol + NADP+	1.1.1.156		
R00968	GTP + Uridine => GDP + UMP	2.7.1.48	OE2749F	
R01521	NADP+ + D-Glucose <=> D-Glucono-1,5-lactone + NADPH	1.1.1.47		PMID:10650712
R04109	L-Glutamyl-tRNA(Glu) + NADPH => (S)-4-Amino-5-oxopentanoate + tRNA(Glu) + NADP+ + H2O	1.2.1.-	OE3496R	
R01523	ATP + D-Ribulose 5-phosphate => ADP + D-Ribulose 1,5-bisphosphate	2.7.1.19		PMID:3255682
R90068	Potassium <=> Potassium (E)		OE1931R	
R90067	ATP + Potassium (E) + H2O => ADP + Orthophosphate + Potassium + H+		OE5051A1F OE5052F OE5053F OE5054F	
R00286	UDPglucose + H2O + NAD+ <=> UDPglucuronate + NADH + H+	1.1.1.22	OE1077R OE2524R	
R05578	tRNA(Glu) + L-Glutamate + ATP <=> L-Glutamyl-tRNA(Glu) + Pyrophosphate + AMP	6.1.1.17	OE2652F	
R90069	L-Ornithine (E) + H+ (pmf associated) => L-Ornithine + H+			PMID:6933439
R05577	tRNA(Asp) + L-Aspartate + ATP <=> L-Aspartyl-tRNA(Asp) + Pyrophosphate + AMP	6.1.1.12	OE1687F	
R00188	Adenosine 3',5'-bisphosphate + H2O => AMP + Orthophosphate	3.1.3.7		
R90064	2,3-Bis-O-(geranylgeranyl)glycerol 1-phosphate + H2O + NADPH + H+ => 2,3-Di-O-phytanyl-sn-glycerol + Orthophosphate + NADP+			PMID:16788058
				Continued on next page

Id	Definition	EC No.	Genes	Ref
R01528	6-Phospho-D-gluconate + NADP+ => D-Ribulose 5-phosphate + CO2 + NADPH + H+	1.1.1.44	OE4581F	PMID:3255682
R90063	HCO3- + H+ <=> H2O + CO2			
R90066	Sodium (E) => Sodium			
R90065	H+ (pmf associated) + Sodium + Potassium (E) => H+ + Potassium + Sodium (E) [pmf]		OE2841R OE3422R OE4599F OE5301F	
R90060	FADH2 + Menaquinone => FAD + Menaquinol	1.3.99.1	OE2865R OE2866R OE2867R OE2868R	
R00183	AMP + H2O => Adenosine + Orthophosphate	3.1.3.5	OE3017R	
R90062	L-Ornithine + L-Arginine (E) => L-Ornithine (E) + L-Arginine			PMID:6933439
R90061	L-Ornithine => L-Ornithine (E)			
R01127	IMP + H2O <=> 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	3.5.4.10	OE4329F	PMID:11844782
R01126	IMP + H2O => Inosine + Orthophosphate	3.1.3.5	OE3017R	
R01818	D-Mannose 6-phosphate <=> D-Mannose 1-phosphate	5.4.2.8	OE2318R OE4094F	
R00189	ATP + Deamino-NAD+ + NH3 => AMP + Pyrophosphate + NAD+	6.3.1.5	OE3843F	
R02058	Acetyl-CoA + D-Glucosamine 6-phosphate => CoA + N-Acetyl-D-glucosamine 6-phosphate	2.3.1.4		
R01954	ATP + L-Citrulline + L-Aspartate => AMP + Pyrophosphate + N-(L-Arginino)succinate	6.3.4.5	OE4420R	
R00289	UTP + D-Glucose 1-phosphate <=> Pyrophosphate + UDPglucose	2.7.7.9	OE1578F OE1016R OE1078F OE1114F OE2530F OE1014R	PMID:10760168
R01541	ATP + 2-Dehydro-3-deoxy-D-gluconate => ADP + 2-Dehydro-3-deoxy-6-phospho-D-gluconate	2.7.1.45	OE1266R	PMID:11271421 PMID:3255682
R00959	D-Glucose 1-phosphate <=> D-Glucose 6-phosphate	5.4.2.2	OE2318R OE4094F OE4190F	
R90150	Acetaldehyde (E) + H+ (pmf associated) <=> Acetaldehyde + H+			
R01034	Glycerone + NADH <=> Glycerol + NAD+	1.1.1.6	OE5160F	PMID:3255682
R00291	UDPglucose <=> UDP-D-galactose	5.1.3.2	OE2110R OE1113R	
R00004	Pyrophosphate + H2O => Orthophosphate	3.6.1.1	OE1407F	
R00190	Adenine + 5-Phospho-alpha-D-ribose 1-diphosphate => AMP + Pyrophosphate	2.4.2.7	OE1504F	PMID:9457844
R90059	Cobalt (E) + ATP + H2O => Cobalt + ADP + Orthophosphate + H+		OE3315R OE3317R OE3318R OE3319R OE4256F	
R90058	Biotin (E) + H+ (pmf associated) => Biotin + H+			
R00009	H2O2 => Oxygen + H2O	1.11.1.6	OE5186R	PMID:7814327
R90057	Thiamin (E) + ATP + H2O => Thiamin + ADP + Orthophosphate + H+		OE3641F OE4358F OE4359F	
R90056	Folate (E) + H+ (pmf associated) => Folate + H+			
R02147	Guanosine + Orthophosphate <=> Guanine + alpha-D-Ribose 1-phosphate	2.4.2.1		PMID:9457844
R90151	Glycerol => Glycerol (E)			
R90055	Sulfate (E) + ATP + H2O => Sulfate + ADP + Orthophosphate + H+		OE3269R	
R90054	Orthophosphate (E) + H+ (pmf associated) => Orthophosphate + H+		OE5132F	
R01130	IMP + NAD+ + H2O => Xanthosine 5'-phosphate + NADH + H+	1.1.1.205	OE2458R	
R00199	ATP + Pyruvate + H2O => AMP + Phosphoenolpyruvate + Orthophosphate	2.7.9.2	OE1500R	PMID:2256927 PMID:9827332
R90052	Nitrate (E) + H+ (pmf associated) => Nitrate + H+			
R90050	Iron (E) + H2O + ATP => Iron + ADP + H+ + Orthophosphate		OE4591R OE4593R OE2346R OE2348R	PMID:12196172
R00192	S-Adenosyl-L-homocysteine + H2O <=> Adenosine + L-Homocysteine	3.3.1.1	OE4159F	
R01131	ATP + Inosine => ADP + IMP	2.7.1.73	OE3606R	PMID:9457844
R01132	Hypoxanthine + 5-Phospho-alpha-D-ribose 1-diphosphate => IMP + Pyrophosphate	2.4.2.8	OE1840R	PMID:9457844

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R00066	6,7-Dimethyl-8-(1-D-ribityl)lumazine => Riboflavin + 4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine	2.5.1.9	OE1946R OE4683F	
R04030	ATP + 2-Succinylbenzoate + CoA => AMP + Pyrophosphate + 2-Succinylbenzoyl-CoA	6.2.1.26	OE2555R	
R02061	trans,trans-Farnesyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + Geranylgeranyl diphosphate	2.5.1.10 2.5.1.29 2.5.1.1	OE2650F OE4010F	PMID:3700337
R00945	Tetrahydrofolate + L-Serine <=> 5,10-Methylenetetrahydrofolate + Glycine + H2O	2.1.2.1	OE3036F	
R00258	L-Alanine + Oxaloacetate <=> Pyruvate + L-Aspartate	2.6.1.1	OE1755F OE1944R OE2619F	PMID:8157586 PMID:9827332
R04137	3-Hydroxyisovaleryl-CoA <=> 3-Methylcrotonyl-CoA + H2O	4.2.1.17	OE3846R	
R00257	ATP + Deamino-NAD+ + L-Glutamine + H2O => AMP + Pyrophosphate + NAD+ + L-Glutamate	6.3.5.1	OE3843F	
R04037	Phosphoribosyl-AMP + H2O <=> 5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-	3.5.4.19	OE4199R	PMID:3700337
R00940	Tetrahydrofolate + NADP+ => Folate + NADPH	1.5.1.3		PMID:17416665 PMID:15554970
R04035	Phosphoribosyl-ATP + H2O => Phosphoribosyl-AMP + Pyrophosphate	3.6.1.31	OE4641R	PMID:3700337
R02063	Geranylgeranyl diphosphate + NADPH + H+ => Phytyl diphosphate + NADP+	-	OE1657R OE1699R OE2720R OE4702F	PMID:3700337
R02257	Propane-1,2-diol + NAD+ <=> (S)-Lactaldehyde + NADH	1.1.1.77	OE4036R OE4021F OE1778R OE2916F	
R00943	Tetrahydrofolate + Formate + ATP => ADP + Orthophosphate + 10-Formyltetrahydrofolate	3.5.4.9 6.3.4.3	OE3038F	
R04031	2-Succinylbenzoate + H2O <=> 2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate	4.2.1.-	OE2558R	
R02065	Geranylgeranyl diphosphate => Pyrophosphate + Prephytoene diphosphate + H+	2.5.1.32	OE3093R OE3376F	PMID:2857171
R90091	6-Deoxy-5-Ketofructose 1 phosphate + L-Aspartate 4-semialdehyde => Compound I + Hydroxypyruvaldehyde 3-P		OE1472F	PMID:15182204
R90081	Thymidine (E) + H+ (pmf associated) => Thymidine + H+			
R90082	Cytidine (E) + H+ (pmf associated) => Cytidine + H+			
R00742	ATP + Acetyl-CoA + HCO3- => ADP + Orthophosphate + Malonyl-CoA	6.4.1.2	OE1939F OE3175F	
R90083	Guanosine (E) + H+ (pmf associated) => Guanosine + H+		OE3950R	PMID:9457844
R01123	Isopentenyl diphosphate <=> Dimethylallyl diphosphate	5.3.3.2	OE3560F OE6213R OE7093R	PMID:3700337
R90084	Deoxyadenosine (E) + H+ (pmf associated) => Deoxyadenosine + H+			
R00121	Oxygen + H2O + Sulfur <=> Hydrogen sulfide + Sulfite	1.13.99.-		
R00848	sn-Glycerol 3-phosphate + FAD => Glycerone phosphate + FADH2	1.1.99.5	OE3763F OE3764F OE3765F	PMID:3700337 PMID:5534921
R90085	Deoxyuridine (E) + H+ (pmf associated) => Deoxyuridine + H+			
R90086	Deoxycytidine (E) + H+ (pmf associated) => Deoxycytidine + H+			
R00849	sn-Glycerol 3-phosphate + NADP+ => Glycerone phosphate + NADPH	1.1.99.4 1.1.99.5	OE3763F OE3764F OE3765F	PMID:3700337 PMID:5534921
R90087	Deoxyguanosine (E) + H+ (pmf associated) => Deoxyguanosine + H+			
R90088	Adenosine-GDP-Cobinamide (E) + ATP + H2O => Adenosine-GDP-cobinamide + ADP + Orthophosphate + H+			
R00847	ATP + Glycerol => ADP + sn-Glycerol 3-phosphate	2.7.1.30	OE3762R	PMID:3255682 PMID:3700337 PMID:5534921

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R05553	4-amino-4-deoxychorismate <=> 4-Aminobenzoate + Pyruvate	4.1.3.38		
R90089	D-Glucose (E) + ATP + H2O => D-Glucose + ADP + H+ + Orthophosphate		OE1314F	Arch Microbiol 155:131-136
R05555	trans,trans-Farnesyl diphosphate + Isopentenyl diphosphate <=> trans,trans,cis-Geranylgeranyl diphosphate + Pyrophosphate	2.5.1.31 2.5.1.-	OE3503F OE3505R OE2559R OE2650F OE4010F	PMID:2857171
R00262	L-threo-3-Methylaspartate <=> L-Glutamate	5.4.99.1	OE4204F OE4206F	
R00261	L-Glutamate => 4-Aminobutanoate + CO2	4.1.1.19 4.1.1.15	OE1498R	
R00935	(S)-Methylmalonate semialdehyde + CoA + NAD+ => Propanoyl-CoA + CO2 + NADH + H+	1.2.1.27	OE4529F OE2367F OE2190R OE2133R	
R00937	Tetrahydrofolate + NAD+ => Folate + NADH	1.5.1.3		PMID:17416665 PMID:15554970
R00936	Dihydrofolate + NADH + H+ => Tetrahydrofolate + NAD+	1.5.1.3		PMID:17416665 PMID:15554970
R00939	Dihydrofolate + NADPH + H+ <=> Tetrahydrofolate + NADP+	1.5.1.3	OE2921R	PMID:15554970 PMID:17416665
R00267	Isocitrate + NADP+ <=> 2-Oxoglutarate + CO2 + NADPH + H+	1.1.1.42	OE3634F	PMID:4393394 PMID:5058691 PMID:5780095 PMID:5802601 PMID:9827332
R04125	Tetrahydrofolate + S-Aminomethyldihydrolipoylprotein <=> 5,10-Methylenetetrahydrofolate + NH3 + Dihydrolipoylprotein	2.1.2.10	OE3278R	
R00268	Oxalosuccinate <=> 2-Oxoglutarate + CO2	1.1.1.42	OE3634F	PMID:4393394 PMID:5058691 PMID:5780095 PMID:5802601 PMID:9827332
R00124	ATP + ADP <=> ADP + ATP	2.7.4.6	OE2667F	
R01106	Nitrate + Halocyanin (reduced) => Nitrite + Halocyanin (oxidized)	1.7.99.4		PMID:629538 PMID:4750772
R90080	Adenosine (E) + H+ (pmf associated) => Adenosine + H+		OE3950R	PMID:9457844
R00127	ATP + AMP <=> ADP	2.7.4.3	OE3425F	
R90072	Oxygen (E) <=> Oxygen			
R90073	Hydrogen (E) <=> Hydrogen			
R90070	H2O (E) <=> H2O			
R90071	CO2 (E) <=> CO2			
R00835	D-Glucose 6-phosphate + NADP+ <=> D-Glucono-1,5-lactone 6-phosphate + NADPH + H+	1.1.1.49		PMID:11271421 PMID:5780095 PMID:10650712
R90076	sn-Glycerol-3-P (E) + ATP + H2O => sn-Glycerol 3-phosphate + ADP + Orthophosphate + H+		OE5166F OE5168F OE5169F OE5170F	PMID:3700337
R00130	ATP + Dephospho-CoA => ADP + CoA	2.7.1.24	OE4442F	
R90077	(S)-Lactate (E) + H+ (pmf associated) <=> (S)-Lactate + H+		OE6063F	
R01519	D-Glucono-1,5-lactone + H2O => D-Gluconic acid	3.1.1.17		PMID:10650712 PMID:2256927
R01518	2-Phospho-D-glycerate <=> 3-Phospho-D-glycerate	5.4.2.1 5.4.2.4	OE3653R	PMID:2256927 PMID:9827332
R00734	L-Tyrosine + 2-Oxoglutarate <=> 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate	2.6.1.9 2.6.1.1 2.6.1.5 2.6.1.57	OE2507R OE1755F OE1944R OE2619F	
R90074	NH3 (E) <=> NH3			
R02272	(S)-4-Amino-5-oxopentanoate <=> 5-Aminolevulinate	5.4.3.8	OE4268F	
R90075	D-Ribose (E) + ATP + H2O => D-Ribose + ADP + Orthophosphate + H+		OE2314R OE2315R OE2316R	
R00275	O2.- + H+ <=> H2O2 + Oxygen	1.15.1.1	OE2708R OE2906R	PMID:3341765
R00833	(R)-2-Methyl-3-oxopropanoyl-CoA <=> Succinyl-CoA	5.4.99.2	OE1721R OE2005F OE1972F	

Continued on next page

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R90078	Oxalate (E) + Formate => Oxalate + Formate (E)		OE1265F	
R90079	Cl- => Cl- (E)			
R01513	3-Phospho-D-glycerate + NAD+ <=> 3-Phosphonooxypyruvate + NADH + H+	1.1.1.95	OE1165R OE3065R OE4408F	
R01512	ATP + 3-Phospho-D-glycerate <=> ADP + 3-Phospho-D-glyceroyl phosphate	2.7.2.3	OE2745R	PMID:2256927 PMID:9827332
R04158	Geranylgeranyl diphosphate + sn-Glycerol 3-phosphate <=> Pyrophosphate + sn-3-O-(Geranylgeranyl)glycerol 1-phosphate	2.5.1.41		
R00236	Acetyl adenylate + CoA => AMP + Acetyl-CoA	6.2.1.1	OE2450F OE1726F	
R00235	ATP + Acetate + CoA => AMP + Pyrophosphate + Acetyl-CoA	6.2.1.1	OE2450F OE1726F	
R00927	CoA + 2-Methylacetoacetyl-CoA <=> Propanoyl-CoA + Acetyl-CoA	2.3.1.16	OE3884F	
R00926	Propinol adenylate + CoA <=> AMP + Propanoyl-CoA	6.2.1.1	OE2450F OE1726F	
R00238	Acetyl-CoA <=> CoA + Acetoacetyl-CoA	2.3.1.9 2.3.1.16	OE3884F	PMID:12013440 PMID:3700337
R00924	Propanoyl-CoA + FAD <=> FADH2 + Propenoyl-CoA	1.3.99.3	OE1555F OE2013R OE2138F OE2710F OE3123R OE4500R	PMID:3700337
R02082	(S)-3-Hydroxy-3-methylglutaryl-CoA + NADPH <=> (R)-Mevalonate + CoA + NADP+	1.1.1.34	OE3637R	PMID:3700337
R02003	Geranyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + trans,trans-Farnesyl diphosphate	2.5.1.10 2.5.1.29 2.5.1.1	OE2650F OE4010F	PMID:3700337
R04150	2-Succinylbenzoyl-CoA <=> 1,4-Dihydroxy-2-naphthoate + CoA	4.1.3.36	OE2561R	
R02085	(S)-3-Hydroxy-3-methylglutaryl-CoA <=> 3-Methylglutaconyl-CoA + H2O	4.2.1.18		
R05145	Biotinyl-5'-AMP + Apo-[carboxylase] <=> AMP + Holo-[carboxylase]	6.3.4.15	OE3186F OE1987F	
R00137	ATP + Nicotinamide D-ribonucleotide => Pyrophosphate + NAD+	2.7.7.1	OE1462R	
R02088	dAMP + H2O => Deoxyadenosine + Orthophosphate	3.1.3.5	OE3017R	
R01900	Isocitrate <=> cis-Aconitate + H2O	4.2.1.3	OE4613F	PMID:5780095 PMID:9827332
R02984	dTDPglucose <=> dTDPgalactose	5.1.3.2	OE2110R OE1113R	
R02285	N-Formimino-L-glutamate + H2O => L-Glutamate + Formamide	3.5.3.8	OE2736F	
R00722	ATP + IDP <=> ADP + ITP	2.7.4.6	OE2667F	
R01701	3-Methyl-2-oxobutanoate + Lipoamide => S-(2-Methylpropanoyl)-dihydrolipoamide + CO2	1.2.4.4	OE4114F OE4113F OE3712R	
R02288	4-Imidazolone-5-propanoate + H2O => N-Formimino-L-glutamate	3.5.2.7	OE2738F	
R01702	4-Methyl-2-oxopentanoate + Lipoamide => S-(3-Methylbutanoyl)-dihydrolipoamide + CO2	1.2.4.4	OE4114F OE4113F OE3712R	
R01001	L-Cystathionine + H2O => L-Cysteine + NH3 + 2-Oxobutanoate	4.4.1.1	OE2173F OE2681F	
R00036	5-Aminolevulinate => Porphobilinogen + H2O	4.2.1.24	OE4262F	
R00243	L-Glutamate + NAD+ + H2O <=> 2-Oxoglutarate + NH3 + NADH + H+	1.4.1.2 1.4.1.3	OE1270F OE2728R	PMID:2917175 PMID:3700337 PMID:8157586 PMID:8605224 PMID:9827332 PMID:10076069 PMID:12052548
R00864	Sulfite + Sulfur <=> Thiosulfate			
R00245	L-Glutamate 5-semialdehyde + NAD+ + H2O <=> L-Glutamate + NADH + H+	1.5.1.12	OE2133R OE2190R OE2367F OE4529F	PMID:9827332
R02199	L-Isoleucine + 2-Oxoglutarate <=> (R)-2-Oxo-3-methylpentanoate + L-Glutamate	2.6.1.42 2.6.1.6	OE3959R	

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R04148	Nicotinate D-ribonucleotide + Dimethylbenzimidazole => Nicotinate + N1-(5-Phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole	2.4.2.21	OE3242F	
R02198	L-Isoleucine + 2-Oxoglutarate <=> 3-Methyl-2-oxopentanoate + L-Glutamate	2.6.1.42	OE3959R	
R02094	ATP + dTMP <=> ADP + dTDP	2.7.4.9 2.7.4.12	OE3715R	
R02014	dATP + Oxidized thioredoxin + H2O <=> ATP + Thioredoxin	1.17.4.2	OE3328R OE4345R OE4346R	
R02093	ATP + dTDP <=> ADP + dTTP	2.7.4.6	OE2667F	
R00248	L-Glutamate + NADP+ + H2O <=> 2-Oxoglutarate + NH3 + NADPH + H+	1.4.1.3 1.4.1.4	OE2728R OE1943F	PMID:1980084 PMID:3700337 PMID:8157586 PMID:9827332 PMID:12052548
R02091	dGTP + Cytidine => dGDP + CMP	2.7.1.48	OE2749F	
R04144	ATP + 5-Phosphoribosylamine + Glycine => ADP + Orthophosphate + 5'-Phosphoribosylglycinamide	6.3.4.13	OE2864F	
R02098	ATP + dUMP <=> ADP + dUDP	2.7.4.9 2.7.4.4	OE3715R	
R02097	dTTP + Uridine => dTDP + UMP	2.7.1.48	OE2749F	
R00149	ATP + NH3 + CO2 + H2O => ADP + Orthophosphate + Carbamoyl phosphate	6.3.4.16	OE3554F OE3556R	
R02096	dTTP + Cytidine => dTDP + CMP	2.7.1.48	OE2749F	
R04800	Neurosporene + Oxygen + NADPH => Lycopene + H2O + NADP+	1.14.99.30 1.14.99.-	OE1426F OE3381R OE3468R	
R04292	Iminoaspartate + Glycerone phosphate => Pyridine-2,3-dicarboxylate + H2O + Orthophosphate	4.1.99.-	OE3644F	
R05679	Glycerone phosphate + NADH + H+ => Glycerol 1-phosphate + NAD+	1.1.1.261	OE1602F	PMID:9973362 PMID:9419225
R02099	ATP + Deoxyuridine => ADP + dUMP	2.7.1.21	OE3159R	
R90098	Glycerol (E) + ATP + H2O => Glycerol + ADP + Orthophosphate + H+		OE5166F OE5168F OE5169F OE5170F	PMID:3700337
R00710	Acetaldehyde + NAD+ + H2O => Acetate + NADH + H+	1.2.1.3	OE2133R OE2190R OE2367F OE4529F	
R00858	Sulfite + NADPH => Hydrogen sulfide + NADP+ + H2O	1.8.1.2		
R00711	Acetaldehyde + NADP+ + H2O => Acetate + NADPH + H+	1.2.1.3	OE2133R OE2190R OE2367F OE4529F	
R90099	Cl- (E) + Photon => Cl-		OE1299R	PMID:9729742
R00713	Succinate semialdehyde + NAD+ + H2O => Succinate + NADH + H+	1.2.1.16 1.2.1.24	OE2190R OE2133R OE2367F OE4529F	
R90097	Formate => Formate (E)			
R01714	5-O-(1-Carboxyvinyl)-3-phosphoshikimate => Chorismate + Orthophosphate	4.2.3.5	OE2761R	PMID:3700337
R90094	D-Galactose (E) + ATP + H2O => D-Galactose + ADP + H+ + Orthophosphate		OE1314F	
R90095	Acetate (E) + H+ (pmf associated) <=> Acetate + H+			
R00150	ADP + Carbamoyl phosphate => ATP + NH3 + CO2	2.7.2.2	OE5206R	PMID:7868583 PMID:8759859
R90092	Compound I + NAD+ => 2-amino-3,7-dideoxy-hept-6-ulsonic acid + NADH + NH3		OE1475F	PMID:15182204
R02297	Xanthosine + Orthophosphate <=> Xanthine + alpha-D-Ribose 1-phosphate	2.4.2.1		PMID:9457844
R90093	Cobamide coenzyme (E) + ATP + H2O => Cobamide coenzyme + ADP + H+ + Orthophosphate		OE2952F OE2955F OE2951R	PMID:16109931
R05680	Glycerone phosphate + NADPH + H+ => Glycerol 1-phosphate + NADP+	1.1.1.261	OE1602F	PMID:9973362
R01717	Chorismate <=> Isochorismate	5.4.4.2	OE2566R	
R01716	Chorismate + L-Glutamine <=> 4-amino-4-deoxychorismate + L-Glutamate	6.3.5.8		
R01715	Chorismate => Prephenate	5.4.99.5	OE2784R	PMID:3700337

Table A.1 – continued from previous page

Id	Definition	EC No.	Genes	Ref
R02291	4-Phospho-L-aspartate + NADPH + H+ => L-Aspartate 4-semialdehyde + Orthophosphate + NADP+	1.2.1.11	OE3063F	
R00714	Succinate semialdehyde + NADP+ + H2O => Succinate + NADPH + H+	1.2.1.16	OE2190R OE4529F OE2133R OE2367F	
R02292	L-Aspartate 4-semialdehyde + Pyruvate <=> 2,3-Dihydrodipicolinate + H2O	4.2.1.52	OE1665R	
R04138	ATP + 3-Methylcrotonyl-CoA + HCO3- => ADP + Orthophosphate + 3-Methylglutaconyl-CoA	6.4.1.4	OE3177F	
R00253	ATP + L-Glutamate + NH3 => ADP + Orthophosphate + L-Glutamine	6.3.1.2	OE3922R	
R00716	L-Lysine + 2-Oxoglutarate + NADPH + H+ => N6-(L-1,3-Dicarboxypropyl)-L-lysine + NADP+ + H2O	1.5.1.8		PMID:3700337

Table A.2: *Natronomonas pharaonis* reaction list. In cases where there is ambiguity in the genetic evidence, all of the candidate genes are listed.

Id	Definition	EC No.	Genes
R04225	3-Methyl-2-oxopentanoate + Lipoamide => S-(2-Methylbutanoyl)-dihydrolipoamide + CO2	1.2.4.4	NP0560A NP2542A NP0558A
R03018	ATP + Pantothenate => ADP + D-4'-Phosphopantothenate	2.7.1.33	
R02065	Geranylgeranyl diphosphate => Pyrophosphate + Prephytoene diphosphate + H+	2.5.1.32	NP4770A
R01800	CDPdiacylglycerol + L-Serine => CMP + Phosphatidylserine	2.7.8.8	NP3230A
R03509	N-(5-Phospho-D-ribosyl)anthranilate <=> 1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate	5.3.1.24	NP3340A
R02649	ATP + N-Acetyl-L-glutamate => ADP + N-Acetyl-L-glutamate 5-phosphate	2.7.2.8	NP5262A
R03096	Indole-3-acetamide + H2O <=> Indole-3-acetate + NH3	3.5.1.4	NP0178A NP0374A NP4212A
R00669	N-Acetylornithine + H2O => Acetate + L-Ornithine	3.5.1.16 3.5.1.14	NP0200A NP5266A
R90079	Cl- => Cl- (E)		
R00344	ATP + Pyruvate + HCO3- => ADP + Orthophosphate + Oxaloacetate	6.4.1.1	NP4252A NP4368A
R01398	Carbamoyl phosphate + L-Ornithine <=> Orthophosphate + L-Citrulline	2.1.3.3	NP5268A
R05595	Crotonoyl-CoA + H2O <=> 3-Hydroxybutanoyl-CoA	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R90124	L-Citrulline (E) + H+ (pmf associated) => L-Citrulline + H+		
R00511	CMP + H2O => Cytidine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R01196	Oxidized ferredoxin + Pyruvate + CoA <=> Reduced ferredoxin + Acetyl-CoA + CO2	1.2.7.1	NP4044A NP4046A
R04144	ATP + 5-Phosphoribosylamine + Glycine <=> ADP + Orthophosphate + 5'-Phosphoribosylglycinamide	6.3.4.13	NP3972A
R00160	FAD + H2O => AMP + FMN	3.6.1.9 3.6.1.18	
R90082	Cytidine (E) + H+ (pmf associated) => Cytidine + H+		
R03035	ATP + Pantetheine 4'-phosphate => Pyrophosphate + Dephospho-CoA	2.7.7.3	NP3368A
R00485	L-Asparagine + H2O => L-Aspartate + NH3	3.5.1.1 3.5.1.38	NP2432A
R01056	D-Ribose 5-phosphate <=> D-Ribulose 5-phosphate	5.3.1.6	NP0786A
R00362	Citrate => Acetate + Oxaloacetate	4.1.3.6	
R01220	5,10-Methylenetetrahydrofolate + NADP+ <=> 5,10-Methenyltetrahydrofolate + NADPH + H+	1.5.1.5	NP2054A
R00066	6,7-Dimethyl-8-(1-D-ribityl)lumazine => Riboflavin + 4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine	2.5.1.9	NP1356A
			Continued on next page

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R04746	(S)-Hydroxyoctanoyl-CoA <=> trans-Oct-2-enoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A NP4746A
R01001	L-Cysteine + NH3 + 2-Oxobutanoate => L-Cystathionine + H2O	4.4.1.1	
R03217	O-Acetyl-L-homoserine + L-Cysteine => L-Cystathionine + Acetate	2.5.1.49	NP0280A NP0284A
R05046	Formamidopyrimidine nucleoside triphosphate + H2O => 2,5-Diaminopyrimidine nucleoside triphosphate + Formate	3.5.4.16	
R04639	2,5-Diamino-6-(5'-triphosphoryl-3',4'-trihydroxy-2'-oxopentyl)- <=> 2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H2O	3.5.4.16	
R90075	D-Ribose (E) + ATP + H2O => D-Ribose + ADP + Orthophosphate + H+		
R00127	ATP + AMP <=> ADP	2.7.4.3	NP4910A
R00999	O-Succinyl-L-homoserine + H2O => 2-Oxobutanoate + Succinate + NH3	2.5.1.48	NP4746A
R04125	Tetrahydrofolate + S-Aminomethyldihydrolipoylprotein => 5,10-Methylenetetrahydrofolate + NH3 + Dihydrolipoylprotein	2.1.2.10	NP4774A
R01900	Isocitrate <=> cis-Aconitate + H2O	4.2.1.3	NP0404A NP1994A
R03243	L-Histidinol phosphate + 2-Oxoglutarate <=> 3-(Imidazol-4-yl)-2-oxopropyl phosphate + L-Glutamate	2.6.1.9	NP2140A
R04148	Nicotinate D-ribonucleotide + Dimethylbenzimidazole => Nicotinate + N1-(5-Phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole	2.4.2.21	NP0736A
R00790	Nitrite + Reduced ferredoxin + H+ => NH3 + H2O + Oxidized ferredoxin	1.7.7.1	NP4224A NP1146A NP4004A
R01514	ATP + D-Glycerate => ADP + 3-Phospho-D-glycerate	2.7.1.31	NP1162A
R00667	L-Ornithine + 2-Oxoglutarate <=> L-Glutamate 5-semialdehyde + L-Glutamate	2.6.1.13	NP5264A
R90061	L-Ornithine => L-Ornithine (E)		
R90080	Adenosine (E) + H+ (pmf associated) => Adenosine + H+		NP1252A
R00571	ATP + UTP + NH3 => ADP + Orthophosphate + CTP	6.3.4.2	NP1968A
R00014	Thiamin diphosphate + Pyruvate => 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO2	2.2.1.6 1.2.4.1	NP1904A NP2200A NP2202A
R03503	ATP + 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine => AMP + 2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine	2.7.6.3	
R01177	Acetyl-CoA + Butanoyl-CoA <=> CoA + 3-Oxohexanoyl-CoA	2.3.1.9 2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R01968	dGMP + H2O => Deoxyguanosine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R02473	ATP + (R)-Pantoate + beta-Alanine => AMP + Pyrophosphate + Pantothenate	6.3.2.1	
R01285	Cystathionine + H2O => L-Homocysteine + NH3 + Pyruvate	4.4.1.8	NP4746A
R00573	ATP + UTP + L-Glutamine + H2O => ADP + Orthophosphate + CTP + L-Glutamate	6.3.4.2	NP1968A
R00858	Sulfite + NADPH => Hydrogen sulfide + NADP+ + H2O	1.8.1.2	
R05578	tRNA(Glu) + L-Glutamate + ATP => L-Glutamyl-tRNA(Glu) + Pyrophosphate + AMP	6.1.1.17	NP3694A
R04672	2-(alpha-Hydroxyethyl)thiamine diphosphate + Pyruvate => (S)-2-Acetolactate + Thiamin diphosphate	2.2.1.6	NP1904A NP2200A NP2202A
R05809	Cobalt-precorrin 3 + S-Adenosyl-L-methionine <=> Cobalt-precorrin 4 + S-Adenosyl-L-homocysteine	2.1.1.131	NP1114A NP1116A
R01137	ATP + dADP <=> ADP + dATP	2.7.4.6	NP3666A
R04620	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H2O => 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- + Orthophosphate	3.1.3.1	
R02331	ATP + dUDP <=> ADP + dUTP	2.7.4.6	NP3666A
R90105	ATP + H2O + Copper (E) => ADP + Orthophosphate + H+ + Copper		NP2668A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R01702	4-Methyl-2-oxopentanoate + Lipoamide => S-(3-Methylbutanoyl)-dihydrolipoamide + CO2	1.2.4.4	NP0560A NP2542A NP0558A
R04173	3-Phosphonooxypyruvate + L-Glutamate <=> O-Phospho-L-serine + 2-Oxoglutarate	2.6.1.52	NP2578A NP0884A
R90111	Zinc + H+ (pmf associated) => Zinc (E) + H+		NP3196A
R02325	dCTP + H2O => dUTP + NH3	3.5.4.13	NP5166A NP0954A
R04198	2,3,4,5-Tetrahydrodipicolinate + NAD+ <=> L-2,3-Dihydrodipicolinate + NADH + H+	1.3.1.26	NP1492A
R03824	gamma-Carotene + H+ + NADPH => beta-Carotene + NADP+		NP0652A
R04448	ATP + 5-(2-Hydroxyethyl)-4-methylthiazole => ADP + 4-Methyl-5-(2-phosphoethyl)-thiazole	2.7.1.50	NP4052A
R03457	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate <=> 3-(Imidazol-4-yl)-2-oxopropyl phosphate + H2O	4.2.1.19	NP2258A
R04741	(S)-3-Hydroxydodecanoyl-CoA + NAD+ <=> 3-Oxododecanoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R00377	dCTP => Pyrophosphate + dCMP	2.7.7.7	NP0584A NP0476A NP0482A NP2918A
R90107	L-Proline (E) + Sodium (E) [pmf] => L-Proline + Sodium		NP4534A NP6036A
R00833	(R)-2-Methyl-3-oxopropanoyl-CoA <=> Succinyl-CoA	5.4.99.2	NP1226A NP2320A NP2710A
R04560	10-Formyltetrahydrofolate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide <=> Tetrahydrofolate + 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	2.1.2.3	NP1662A
R00024	D-Ribulose 1,5-bisphosphate + CO2 + H2O => 3-Phospho-D-glycerate	4.1.1.39	NP2770A
R04993	2-Demethylmenaquinone + S-Adenosyl-L-methionine => Menaquinone + S-Adenosyl-L-homocysteine	2.1.1.-	NP5034A
R02082	(S)-3-Hydroxy-3-methylglutaryl-CoA + NADPH => (R)-Mevalonate + CoA + NADP+	1.1.1.34	NP0368A NP2422A
R00104	ATP + NAD+ => ADP + NADP+	2.7.1.23	NP2512A NP4558A
R00783	Nitrite + Ferrocytochrome c + H+ => Nitric oxide + H2O + Ferricytochrome c	1.7.2.1	NP1958A
R03663	ATP + L-Threonine + tRNA(Thr) => AMP + Pyrophosphate + L-Threonyl-tRNA(Thr)	6.1.1.3	NP2410A
R05553	4-amino-4-deoxychorismate <=> 4-Aminobenzoate + Pyruvate	4.1.3.38	
R02063	Geranylgeranyl diphosphate + H+ + NADPH <=> Phytyl diphosphate + NADP+		
R05807	Sirohydrochlorin + Cobalt <=> Cobalt-precorrin 2	4.99.1.3	NP1092A NP1108A NP1588A
R01229	Guanine + 5-Phospho-alpha-D-ribose 1-diphosphate => GMP + Pyrophosphate	2.4.2.7 2.4.2.22	NP1426A NP1254A
R01655	5,10-Methenyltetrahydrofolate + H2O <=> 10-Formyltetrahydrofolate + H+	3.5.4.9	NP2054A
R04475	N-Succinyl-LL-2,6-diaminoheptanedioate + 2-Oxoglutarate <=> N-Succinyl-2-L-amino-6-oxoheptanedioate + L-Glutamate	2.6.1.17	
R07771	Lipoyl-AMP + Apoprotein => Protein N6-(lipoyl)lysine + AMP	2.7.7.63	NP5116A
R01106	Nitrate + Ferrocytochrome c => Ferricytochrome c + Nitrite	1.7.99.4	NP1244A NP4226A
R90037	L-Glutamine (E) + H+ (pmf associated) => L-Glutamine + H+		
R90133	6-Deoxy-5-ketomannitol-1-phosphate + L-Aspartate 4-semialdehyde => Compound I + D-Glyceraldehyde 3-phosphate		NP3160A
R01717	Chorismate <=> Isochorismate	5.4.4.2	NP2724A
R02055	Phosphatidylserine => Phosphatidylethanolamine + CO2	4.1.1.65	NP4176A
R01954	ATP + L-Citrulline + L-Aspartate => AMP + Pyrophosphate + N-(L-Arginino)succinate	6.3.4.5	NP5252A
R01724	Nicotinate + 5-Phospho-alpha-D-ribose 1-diphosphate => Nicotinate D-ribonucleotide + Pyrophosphate	2.4.2.11	NP1174A
R90035	L-Aspartate (E) + H+ (pmf associated) => L-Aspartate + H+		
R05225	Adenosyl cobyrinate a,c diamide + L-Glutamine + ATP <=> Adenosyl cobyrinate hexaamide + L-Glutamate + Orthophosphate + ADP	6.3.5.10	NP5126A
			Continued on next page

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R04559	1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole <=> Fumarate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	4.3.2.2	NP1658A
R02292	L-Aspartate 4-semialdehyde + Pyruvate <=> L-2,3-Dihydrodipicolinate + H2O	4.2.1.52	NP1490A
R04748	(S)-Hydroxyhexanoyl-CoA + NAD+ <=> 3-Oxohexanoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R03665	ATP + L-Valine + tRNA(Val) => AMP + Pyrophosphate + L-Valyl-tRNA(Val)	6.1.1.9	NP0288A
R02017	Thioredoxin + ADP => dADP + Oxidized thioredoxin + H2O	1.17.4.1	NP3346A NP4434A NP5066A NP6166A NP6168A
R01214	L-Valine + 2-Oxoglutarate <=> 3-Methyl-2-oxobutanoic acid + L-Glutamate	2.6.1.42 2.6.1.6	NP0798A NP5036A
R03140	D-Ribulose 1,5-bisphosphate + Oxygen => 3-Phospho-D-glycerate + 2-Phosphoglycolate	4.1.1.39	NP2770A
R03346	Nicotinate D-ribonucleotide + H2O => Nicotinate D-ribonucleoside + Orthophosphate	3.1.3.5	NP1854A NP2590A NP4208A
R03223	2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate + 4-Methyl-5-(2-phosphoethyl)-thiazole => Pyrophosphate + Thiamin monophosphate	2.5.1.3	NP4054A
R01287	O-Acetyl-L-homoserine + Hydrogen sulfide => L-Homocysteine + Acetate	2.5.1.49	NP0280A NP0284A
R90044	L-Serine (E) + H+ (pmf associated) <=> L-Serine + H+		
R00156	ATP + UDP <=> ADP + UTP	2.7.4.6	NP3666A
R04365	Succinyl-CoA + 2,3,4,5-Tetrahydrodipicolinate + H2O <=> CoA + N-Succinyl-2-L-amino-6-oxoheptanedioate	2.3.1.117	NP1494A
R90003	ADP + Orthophosphate + H+ (pmf associated) <=> ATP + H2O + H+	3.6.3.14	NP0264A NP1018A NP1020A NP1022A NP1024A NP1026A NP1028A NP1030A NP1032A
R03459	2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine + H2O => 5-Amino-6-(5'-phosphoribosylamino)uracil + NH3	3.5.4.26	
R90084	Deoxyadenosine (E) + H+ (pmf associated) => Deoxyadenosine + H+		
R90147	Molybdenum (E) + ATP + H2O => Molybdenum + ADP + Orthophosphate + H+		
R02661	2-Methylpropanoyl-CoA + FAD => 2-Methylprop-2-enoyl-CoA + FADH2	1.3.99.3	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R10003	Geranylgeranyl diphosphate + Isopentenyl diphosphate => all-trans-Pentaprenyl diphosphate + Pyrophosphate	2.5.1.-	NP3696A NP0604A
R05071	(S)-2-Acetolactate <=> (R)-3-Hydroxy-3-methyl-2-oxobutanoate	1.1.1.86	NP2198A
R90048	L-Valine (E) + H+ (pmf associated) <=> L-Valine + H+		
R00369	L-Alanine + Glyoxylate => Pyruvate + Glycine	2.6.1.44	NP0884A NP2578A
R02236	Dihydrofolate + NADP+ <=> Folate + NADPH + H+	1.5.1.3	NP2922A
R00291	UDPglucose <=> UDP-D-galactose	5.1.3.2	NP3504A NP4112A NP4662A
R04405	5-Methyltetrahydropteroyltri-L-glutamate + L-Homocysteine => Tetrahydropteroyltri-L-glutamate + L-Methionine	2.1.1.14	NP3670A NP3672A
R03194	S-Adenosyl-L-methionine + Uroporphyrinogen III <=> S-Adenosyl-L-homocysteine + Precorrin 2	2.1.1.107	NP1328A
R03005	ATP + Nicotinate D-ribonucleotide => Pyrophosphate + Deamino-NAD+	2.7.7.1	NP0310A NP0908A
R90103	Vitamin B12 (E) + ATP + H2O => Vitamin B12 + ADP + Orthophosphate + H+		NP3814A NP3816A NP3818A
R01664	dCMP + H2O => Deoxycytidine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R00548	FMN + H2O => Riboflavin + Orthophosphate	3.1.3.2	NP4208A
R00703	(S)-Lactate + NAD+ <=> Pyruvate + NADH + H+	1.1.1.27	NP3764A NP1452A
R02739	alpha-D-Glucose 6-phosphate <=> beta-D-Glucose 6-phosphate	5.3.1.9	NP4992A
R06529	Adenosyl cobyrinate hexaamide + D-1-Aminopropan-2-ol O-phosphate + ATP <=> Adenosyl cobinamide phosphate + ADP + Orthophosphate	6.3.1.10	NP5308A
R00425	GTP + H2O => Formate + 2,5-Diamino-6-hydroxy-4-(5'-phosphoribosylamino)-pyrimidine + Pyrophosphate	3.5.4.25	

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R90149	UDP-D-galactose + H2O => D-Galactose + UDP		
R03321	beta-D-Glucose 6-phosphate <=> D-Fructose 6-phosphate	5.3.1.9	NP4992A
R02088	dAMP + H2O => Deoxyadenosine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R00674	L-Serine + Indole <=> L-Tryptophan + H2O	4.2.1.20	NP3162A NP3164A
R02872	Presqualene diphosphate + NADPH + H+ => Pyrophosphate + Squalene + NADP+	2.5.1.21	
R00442	CTP => Pyrophosphate + CMP	2.7.7.6	NP0114A NP0116A NP0112A NP0110A NP2834A NP5080A NP5078A NP3682A NP0108A NP2844A NP4162A NP2842A NP5234A
R00149	ATP + NH3 + CO2 + H2O => ADP + Orthophosphate + Carbamoyl phosphate	6.3.4.16	NP4506A NP4830A
R07461	Thiamine biosynthesis intermediate 3 + Thiamine biosynthesis intermediate 4 => Thiamine biosynthesis intermediate 5		NP3782A
R00557	L-Arginine + Oxygen + NADPH => Nitric oxide + L-Citrulline + NADP+	1.14.13.39	NP1908A
R90088	Adenosine-GDP-Cobinamide (E) + ATP + H2O => Adenosine-GDP-cobinamide + ADP + Orthophosphate + H+		
R01465	L-Threonine + NAD+ => L-2-Amino-3-oxobutanoic acid + NADH + H+	1.1.1.103	
R00004	Pyrophosphate + H2O => Orthophosphate	3.6.1.1	NP5192A
R07281	D-Ribulose 5-phosphate => 3, 4-Dihydroxy-2-butanone 4-phosphate + Formate	5.4.99.-	NP0074A
R90069	L-Ornithine (E) + H+ (pmf associated) => L-Ornithine + H+		
R02329	dTTP + alpha-D-Galactose 1-phosphate <=> Pyrophosphate + dTDPgalactose	2.7.7.32	NP4652A NP4656A NP4674A NP4680A NP2386A NP4276A
R00220	L-Serine => Pyruvate + NH3	4.3.1.19	NP1076A
R90145	D-Ribose 1,5-bisphosphate + NAD+ => D-Ribulose 1,5-bisphosphate + NADH		NP5174A NP3202A
R07459	ATP => Thiamine biosynthesis intermediate 4 + Pyrophosphate		
R05810	Cobalt-precorrin 4 + S-Adenosyl-L-methionine <=> Cobalt-precorrin 5 + S-Adenosyl-L-homocysteine	2.1.1.133	NP1120A
R90086	Deoxycytidine (E) + H+ (pmf associated) => Deoxycytidine + H+		
R03659	ATP + L-Methionine + tRNA(Met) => AMP + Pyrophosphate + L-Methionyl-tRNA	6.1.1.10	NP1192A
R00722	ATP + IDP <=> ADP + ITP	2.7.4.6	NP3666A
R90101	Sodium + H+ (pmf associated) => H+ + Sodium (E) [pmf]		NP4702A NP1832A
R01135	GTP + IMP + L-Aspartate => GDP + Orthophosphate + N6-(1,2-Dicarboxyethyl)-AMP	6.3.4.4	NP1530A
R04800	Neurosporene + Oxygen + NADPH => Lycopene + H2O + NADP+	1.14.99.30	NP4764A NP0204A NP1630A
R00429	ATP + GTP => AMP + Guanosine 3'-diphosphate 5'-triphosphate	2.7.6.5	NP3860A
R00248	L-Glutamate + NADP+ + H2O <=> 2-Oxoglutarate + NH3 + NADPH + H+	1.4.1.3 1.4.1.4	NP1582A NP1806A NP6184A
R04208	ATP + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine => ADP + Orthophosphate + Aminoimidazole ribotide	6.3.3.1	NP3940A
R06530	L-Threonine O-3-phosphate <=> D-1-Aminopropan-2-ol O-phosphate + CO2	4.1.1.81	
R01648	4-Aminobutanoate + 2-Oxoglutarate => Succinate semialdehyde + L-Glutamate	2.6.1.19	NP6188A
R00200	ADP + Phosphoenolpyruvate => ATP + Pyruvate	2.7.1.40	NP1746A
R90046	Glycine (E) + H+ (pmf associated) <=> Glycine + H+		NP3588A
R04218	Prephytoene diphosphate + H+ + NADPH => Pyrophosphate + NADP+ + Phytoene	2.5.1.32	NP4770A
R90052	Nitrate (E) + H+ (pmf associated) => Nitrate + H+		
R00177	ATP + L-Methionine + H2O => Orthophosphate + Pyrophosphate + S-Adenosyl-L-methionine	2.5.1.6	NP3100A
R90114	Phosphonate (E) + ATP + H2O + Orthophosphate => Phosphonate + ADP + H+		NP6126A NP6124A NP6122A
R04233	CTP + D-4'-Phosphopantothenate + L-Cysteine => CDP + Orthophosphate + (R)-4'-Phosphopantothenoyl-L-cysteine	6.3.2.5	NP1374A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R01993	(S)-Dihydroorotate + H2O <=> N-Carbamoyl-L-aspartate	3.5.2.3	NP1706A
R03187	ATP + N-Methylhydantoin + H2O => ADP + Orthophosphate + N-Carbamoylsarcosine	3.5.2.14	NP2954A NP6216A NP6218A
R00986	Chorismate + L-Glutamine <=> Anthranilate + Pyruvate + L-Glutamate	4.1.3.27	NP0800A NP0802A NP3342A NP3344A
R90119	Zinc (E) + ATP + H2O => Zinc + ADP + Orthophosphate + H+		
R04787	Phytofluene => zeta-Carotene	1.3.99.-	NP4764A NP0204A NP1630A
R04591	ATP + 1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate + L-Aspartate <=> ADP + Orthophosphate + 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole	6.3.2.6	NP4996A
R90112	Cadmium + H+ (pmf associated) => Cadmium (E) + H+		NP3196A
R05611	all-trans-Heptaprenyl diphosphate + Isopentenyl diphosphate => all-trans-Octaprenyl diphosphate + Pyrophosphate	2.5.1.11	NP0604A NP4556A
R00316	ATP + Acetate => Pyrophosphate + Acetyl adenylate	6.2.1.1	NP4240A NP4242A NP5128A NP5132A NP0164A
R01231	ATP + Xanthosine 5'-phosphate + L-Glutamine + H2O => AMP + Pyrophosphate + GMP + L-Glutamate	6.3.5.2	NP0620A NP1970A
R03651	ATP + L-Glutamate + tRNA(Gln) => AMP + Pyrophosphate + L-Glutamyl-tRNA(Gln)	6.1.1.17	
R90033	L-Methionine (E) + H+ (pmf associated) => L-Methionine + H+		
R03084	3-Dehydroquinate <=> 3-Dehydroshikimate + H2O	4.2.1.10 4.2.1.11	NP2240A
R00275	O2.- + H+ <=> H2O2 + Oxygen	1.15.1.1	NP4040A
R01070	D-Fructose 1,6-bisphosphate <=> Glycerone phosphate + D-Glyceraldehyde 3-phosphate	4.1.2.13	NP1594A NP3160A
R00412	Succinate + FAD <=> Fumarate + FADH2	1.3.99.1	NP4264A NP4266A NP4268A NP4270A
R90092	Compound I + NAD+ => 2-amino-3,7-dideoxy-hept-6-ulsonic acid + NADH + NH3		NP2238A
R02765	(S)-2-Methyl-3-oxopropanoyl-CoA <=> (R)-2-Methyl-3-oxopropanoyl-CoA	5.1.99.1	
R00150	ADP + Carbamoyl phosphate => ATP + NH3 + CO2	2.7.2.2	NP3632A
R06411	cis-2-Methyl-5-isopropylhexa-2,5-dienoyl-CoA + H2O <=> 3-Hydroxy-2,6-dimethyl-5-methylene-heptanoyl-CoA	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R00658	2-Phospho-D-glycerate <=> Phosphoenolpyruvate + H2O	4.2.1.11	NP2846A
R00923	(S)-2-Methyl-3-oxopropanoyl-CoA => Propanoyl-CoA + CO2	4.1.1.41	
R04035	Phosphoribosyl-ATP + H2O <=> Phosphoribosyl-AMP + Pyrophosphate	3.6.1.31	NP0470A
R00578	ATP + L-Aspartate + L-Glutamine + H2O => AMP + Pyrophosphate + L-Asparagine + L-Glutamate	6.3.5.4	NP2978A
R90094	D-Galactose (E) + ATP + H2O => D-Galactose + ADP + H+ + Orthophosphate		NP0530A
R90031	L-Arginine (E) + H+ (pmf associated) => L-Arginine + H+		NP0040A NP0042A NP0646A NP2382A NP2676A NP4372A NP6260A
R02662	2-Methylpropanoyl-CoA + Dihydrolipoamide => S-(2-Methylpropanoyl)-dihydrolipoamide + CoA	2.3.1.-	
R03082	(R)-Lactaldehyde + Glutathione + NAD+ <=> (R)-S-Lactoylglutathione + NADH	1.2.1.1	NP3020A NP1686A
R04520	Geranylgeranyl diphosphate + sn-3-O-(Geranylgeranyl)glycerol 1-phosphate <=> Pyrophosphate + 2,3-Bis-O-(geranylgeranyl)glycerol 1-phosphate	2.5.1.42	
R01385	UDPglucuronate <=> UDP-D-galacturonate	5.1.3.6	NP4662A NP3504A NP4112A
R90130	4-Ketofructose-1,6-bisphosphate + NADH => 6-Deoxy-5-Ketofructose 1 phosphate + Orthophosphate + NAD+		
R90050	Iron (E) + H2O + ATP => Iron + ADP + H+ + Orthophosphate		NP3816A NP3818A NP5004A NP5000A
R01134	IMP + NH3 + NADP+ <=> GMP + NADPH + H+	1.7.1.7	NP3080A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R00443	UTP => Pyrophosphate + UMP	2.7.7.6	NP0114A NP0116A NP0112A NP0110A NP2834A NP5080A NP5078A NP3682A NP0108A NP2844A NP4162A NP2842A NP5234A
R05221	Adenosyl cobinamide + ATP <=> Adenosyl cobinamide phosphate + ADP	2.7.1.156	
R00996	L-Threonine => 2-Oxobutanoate + NH3	4.3.1.19	NP1076A
R00375	dATP => Pyrophosphate + dAMP	2.7.7.7	NP0854A NP0476A NP0482A NP2918A
R01086	N-(L-Arginino)succinate <=> Fumarate + L-Arginine	4.3.2.1	NP5254A
R04467	N6-Acetyl-LL-2,6-diaminoheptanedioate + 2-Oxoglutarate <=> N-Acetyl-L-2-amino-6-oxopimelate + L-Glutamate	2.6.1.-	
R00708	(S)-1-Pyrroline-5-carboxylate + NADP+ + H2O => L-Glutamate + NADPH + H+	1.5.1.12	NP3020A NP1686A
R90116	Oligopeptide (E) + ATP + H2O => Oligopeptide + ADP + Orthophosphate + H+		NP0758A NP0760A NP0762A NP0764A NP0766A NP3578A NP3580A NP3582A NP3584A
R00519	Formate + NAD+ <=> H+ + CO2 + NADH	1.2.1.2	NP3710A NP4944A NP4946A NP4960A NP4962A
R00939	Dihydrofolate + NADPH + H+ => Tetrahydrofolate + NADP+	1.5.1.3	NP2922A
R01569	dTMP + H2O => Thymidine + Orthophosphate	3.1.3.5 3.1.3.35	NP1854A NP2590A
R01714	5-O-(1-Carboxyvinyl)-3-phosphoshikimate <=> Chorismate + Orthophosphate	4.2.3.5	NP3082A
R03026	(S)-3-Hydroxybutanoyl-CoA <=> Crotonoyl-CoA + H2O	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R05223	Adenosine-GDP-cobinamide + alpha-Ribazole => Cobamide coenzyme + GMP	2.7.8.26	NP5306A
R90098	Glycerol (E) + ATP + H2O => Glycerol + ADP + Orthophosphate + H+		
R07462	Thiamine biosynthesis intermediate 5 => C15815		
R06860	2-Oxoglutarate + Isochorismate => 2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate + Pyruvate + CO2	2.5.1.64	NP2726A
R04558	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + L-Glutamate + D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate <=> N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D- + L-Glutamine	2.4.2.-	NP0082A NP4158A
R00502	UTP + alpha-D-Galactose 1-phosphate <=> Pyrophosphate + UDP-D-galactose	2.7.7.10	NP4652A NP4656A NP4674A NP4680A NP2386A NP4276A
R01082	Fumarate + H2O <=> (S)-Malate	4.2.1.2	NP3778A
R02272	5-Aminolevulinate <=> (S)-4-Amino-5-oxopentanoate	5.4.3.8	NP1246A
R00671	L-Ornithine => L-Proline + NH3	4.3.1.12	NP0448A NP3802A
R01775	L-Homoserine + NADP+ <=> L-Aspartate 4-semialdehyde + NADPH + H+	1.1.1.3	NP0302A
R03858	Lauroyl-CoA + Acetyl-CoA <=> CoA + 3-Oxotetradecanoyl-CoA	2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R00704	(R)-Lactate + NAD+ <=> Pyruvate + NADH + H+	1.1.1.28	NP3764A NP1452A
R02340	Indole + D-Glyceraldehyde 3-phosphate <=> Indoleglycerol phosphate	4.2.1.20	NP3162A NP3164A
R90054	Orthophosphate (E) + H+ (pmf associated) => Orthophosphate + H+		NP1412A NP1270A NP4804A
R05811	Cobalt-precorrin 5 + S-Adenosyl-L-methionine + H2O <=> Cobalt-precorrin 6 + S-Adenosyl-L-homocysteine + Acetaldehyde	2.1.1.-	NP1118A
R04426	3-Isopropylmalate + NAD+ <=> 3-Carboxy-4-methyl-2-oxopentanoate + NADH + H+	1.1.1.85	NP2190A
R04594	N1-(5-Phospho-alpha-D-ribosyl)-5,6-dimethylbenzimidazole + H2O => alpha-Ribazole + Orthophosphate	3.1.3.73	NP2774A
			Continued on next page

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R03260	O-Succinyl-L-homoserine + L-Cysteine => L-Cystathionine + Succinate	2.5.1.48	NP4746A
R01728	Prephenate + NAD+ <=> 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADH + H+	1.3.1.12 1.3.1.52 1.3.1.43	NP3358A
R90056	Folate (E) + H+ (pmf associated) => Folate + H+		
R02093	ATP + dTDP <=> ADP + dTTP	2.7.4.6	NP3666A
R03012	L-Histidinol + NAD+ <=> L-Histidinal + NADH + H+	1.1.1.23	NP2876A
R02540	2-Phenylacetamide + H2O <=> Phenylacetic acid + NH3	3.5.1.4	NP0178A NP0374A NP4212A
R04137	3-Hydroxyisovaleryl-CoA <=> 3-Methylcrotonyl-CoA + H2O	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R00566	L-Arginine <=> Agmatine + CO2	4.1.1.19	NP4484A
R03777	Octanoyl-CoA + FAD => trans-Oct-2-enoyl-CoA + FADH2	1.3.99.13 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R03815	Dihydrolipoylprotein + NADP+ => Lipoylprotein + NADPH	1.8.1.4	NP0104A
R05069	(S)-2-Aceto-2-hydroxybutanoate <=> (R)-3-Hydroxy-3-methyl-2-oxopentanoate	5.4.99.3 1.1.1.86	NP2198A
R02300	5,10-Methenyltetrahydrofolate + H2O => 5-Formyltetrahydrofolate	2.1.2.10	NP0436A NP4774A
R90039	L-Histidine (E) + H+ (pmf associated) <=> L-Histidine + H+		NP0040A NP0042A NP0646A NP2382A NP2676A NP4372A NP6260A
R02019	GDP + Thioredoxin => dGDP + Oxidized thioredoxin + H2O	1.17.4.1	NP3346A NP4434A NP5066A NP6166A NP6168A
R00944	10-Formyltetrahydrofolate + H2O => Formate + Tetrahydrofolate	3.5.1.10	NP5006A
R04031	2-Succinylbenzoate + H2O <=> 2-Succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate	4.2.1.-	NP2736A
R04463	ATP + 5'-Phosphoribosyl-N-formylglycinamide + L-Glutamine + H2O => ADP + Orthophosphate + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine + L-Glutamate	6.3.5.3	NP2982A NP5008A NP5010A
R01563	N-Carbamoylsarcosine + H2O => Sarcosine + CO2 + NH3	3.5.1.59	NP1274A NP1170A
R00235	ATP + Acetate + CoA => AMP + Pyrophosphate + Acetyl-CoA	6.2.1.1	NP4240A NP4242A NP5128A NP5132A NP0164A
R02530	Glutathione + Methylglyoxal => (R)-S-Lactoylglutathione	4.4.1.5	NP1228A NP3948A NP4460A NP2650A
R00713	Succinate semialdehyde + NAD+ + H2O => Succinate + NADH + H+	1.2.1.16 1.2.1.24	
R04286	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) => 6-Pyruvoyltetrahydropterin + Triphosphate	4.2.3.12	NP1342A
R04751	Hexanoyl-CoA + FAD => trans-Hex-2-enoyl-CoA + FADH2	1.3.99.13 1.3.99.2 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R00942	ATP + Tetrahydrofolate + L-Glutamate <=> ADP + Orthophosphate + Tetrahydrofolyl-[Glu](2)	6.3.2.17	NP1478A
R02328	dTTP + D-Glucose 1-phosphate <=> Pyrophosphate + dTDPglucose	2.7.7.24	NP4652A NP4656A NP4674A NP4680A NP2386A NP4276A
R04440	(R)-2,3-Dihydroxy-3-methylbutanoate + NADP+ <=> (R)-3-Hydroxy-3-methyl-2-oxobutanoate + NADPH	1.1.1.86	NP2198A
R00158	ATP + UMP <=> ADP + UDP	2.7.4.14	NP4914A
R01325	Citrate <=> cis-Aconitate + H2O	4.2.1.3	NP0404A NP1994A
R90144	Isopentenyl phosphate + ATP => Isopentenyl diphosphate + ADP		NP2852A
R01090	L-Leucine + 2-Oxoglutarate <=> 4-Methyl-2-oxopentanoate + L-Glutamate	2.6.1.42	NP0798A NP5036A
R02485	Deoxycytidine + H2O => Deoxyuridine + NH3	3.5.4.5 3.5.4.14	NP3512A
R04743	(S)-Hydroxydecanoyl-CoA + NAD+ <=> 3-Oxodecanoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R90134	Inosine (E) + H+ (pmf associated) => Inosine + H+		NP1252A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R02283	N-Acetylornithine + 2-Oxoglutarate <=> N-Acetyl-L-glutamate 5-semialdehyde + L-Glutamate	2.6.1.11	NP5264A
R00963	UMP + H2O => Uridine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R02135	Thiamin monophosphate + H2O => Thiamin + Orthophosphate	3.1.3.-	NP2774A
R01226	5,10-Methylenetetrahydrofolate + 3-Methyl-2-oxobutanoic acid + H2O => Tetrahydrofolate + 2-Dehydropantoate	2.1.2.11	NP4334A
R01518	2-Phospho-D-glycerate <=> 3-Phospho-D-glycerate	5.4.2.1	NP1332A NP1964A
R03313	L-Glutamyl 5-phosphate + NADPH + H+ <=> L-Glutamate 5-semialdehyde + Orthophosphate + NADP+	1.2.1.41	NP3978A
R00965	Orotidine 5'-phosphate => UMP + CO2	4.1.1.23	NP1734A
R90100	Nitrate (E) + Photon => Nitrate		NP0642A
R04292	Iminoaspartate + Glycerone phosphate => Pyridine-2,3-dicarboxylate + H2O + Orthophosphate	4.1.99.-	NP2418A
R90142	Molybdenum cofactor + GTP => MoCo dinucleotide (Guanine) + Pyrophosphate		NP1536A NP2138A
R00352	ADP + Orthophosphate + Acetyl-CoA + Oxaloacetate <=> ATP + Citrate + CoA	2.3.3.8	
R01072	L-Glutamine + 5-Phospho-alpha-D-ribose 1-diphosphate + H2O => 5-Phosphoribosylamine + Pyrophosphate + L-Glutamate	2.4.2.14	NP4312A
R03425	Glycine + Lipoylprotein => S-Aminomethyldihydrolipoylprotein + CO2	1.4.4.2	NP4594A NP4596A
R00239	ATP + L-Glutamate => ADP + L-Glutamyl 5-phosphate	2.7.2.11	NP3976A
R90121	Adenine (E) + H+ (pmf associated) => Adenine + H+		NP1252A
R02734	N-Succinyl-LL-2,6-diaminoheptanedioate + H2O <=> Succinate + LL-2,6-Diaminoheptanedioate	3.5.1.18	
R02026	O-Acetyl-L-homoserine + Thiosulfate + Thioredoxin + H+ => L-Homocysteine + Sulfite + Oxidized thioredoxin + Acetate	2.5.1.49	
R00405	ATP + Succinate + CoA <=> ADP + Orthophosphate + Succinyl-CoA	6.2.1.5	NP4354A NP4356A
R05813	Cobalt-dihydro-precorrin 6 + S-Adenosyl-L-methionine <=> Cobalt-precorrin 8 + S-Adenosyl-L-homocysteine + CO2	2.1.1.132	NP1088A NP1124A
R03530	ATP + dIDP <=> ADP + dITP	2.7.4.6	NP3666A
R04744	(S)-Hydroxydecanoyl-CoA <=> trans-Dec-2-enoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R02537	Phenylacetaldehyde + NADP+ + H2O <=> Phenylacetic acid + NADPH	1.2.1.5	NP1686A
R02199	L-Isoleucine + 2-Oxoglutarate <=> (3R)-3-Methyl-2-oxopentanoic acid + L-Glutamate	2.6.1.42	NP0798A NP5036A
R04385	ATP + Holo-[carboxylase] + HCO3- <=> ADP + Orthophosphate + Carboxybiotin-carboxyl-carrier protein	6.3.4.14	NP4252A NP4368A
R01975	(S)-3-Hydroxybutanoyl-CoA + NAD+ <=> Acetoacetyl-CoA + NADH	1.1.1.35	NP1906A NP2754A NP2630A NP4322A
R00271	Acetyl-CoA + H2O + 2-Oxoglutarate => 2-Hydroxybutane-1,2,4-tricarboxylate + CoA	2.3.3.14	NP2994A
R90000	NADH + Menaquinone + H+ => NAD+ + Menaquinol + H+ (pmf associated)	1.6.5.3	NP2292A NP2294A NP2296A NP2298A NP2300A NP2304A NP2306A NP2308A NP2310A NP2312A NP2314A NP2406A NP3508A
R00694	L-Phenylalanine + 2-Oxoglutarate <=> Phenylpyruvate + L-Glutamate	2.6.1.9 2.6.1.1 2.6.1.5 2.6.1.57	NP0824A NP1666A NP2140A NP4024A NP4410A
R00480	ATP + L-Aspartate <=> ADP + 4-Phospho-L-aspartate	2.7.2.4	NP0550A
R00479	Isocitrate => Succinate + Glyoxylate	4.1.3.1	NP4432A
R02984	dTDPglucose <=> dTDPgalactose	5.1.3.2	NP4662A NP3504A NP4112A
R00086	ATP + H2O => ADP + Orthophosphate	3.6.1.15 3.6.3.14	NP4426A NP0264A NP1018A NP1020A NP1022A NP1024A NP1026A NP1028A NP1030A NP1032A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R00568	CTP + H2O => UTP + NH3	3.5.4.13	NP5166A NP0954A
R05815	Cobyrinate + L-Glutamine + ATP + H2O <=> Cob(II)yrinate a,c diamide + L-Glutamate + ADP + Orthophosphate	6.3.1.-	NP0738A
R00927	CoA + 2-Methylacetoacetyl-CoA => Propanoyl-CoA + Acetyl-CoA	2.3.1.16	NP2214A
R90071	CO2 (E) <=> CO2		
R00258	L-Alanine + 2-Oxoglutarate <=> Pyruvate + L-Glutamate	2.6.1.2	NP0824A NP1666A NP4024A NP4410A
R90060	FADH2 + Menaquinone => FAD + Menaquinol	1.3.99.1	NP4264A NP4266A NP4268A NP4270A
R00303	H2O + alpha-D-Glucose 6-phosphate => D-Glucose + Orthophosphate	3.1.3.9	
R01051	ATP + D-Ribose => ADP + D-Ribose 5-phosphate	2.7.1.15	NP3184A
R00354	Acetyl-CoA + Oxaloacetate => (3S)-Citryl-CoA	4.1.3.34	NP4244A
R04738	(S)-3-Hydroxyhexadecanoyl-CoA <=> trans-Hexadec-2-enoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R04037	Phosphoribosyl-AMP + H2O <=> 5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-	3.5.4.19	NP5154A
R00524	Formamide + H2O => Formate + NH3	3.5.1.49	NP6204A
R04150	2-Succinylbenzoyl-CoA <=> 1,4-Dihydroxy-2-naphthoate + CoA	4.1.3.36	NP2730A
R01561	Adenosine + Orthophosphate <=> Adenine + D-Ribose 1-phosphate	2.4.2.1	
R03471	ATP + 4-Amino-5-hydroxymethyl-2-methylpyrimidine => ADP + 4-Amino-2-methyl-5-phosphomethylpyrimidine	2.7.1.49	NP0564A
R04325	10-Formyltetrahydrofolate + 5'-Phosphoribosylglycinamide <=> Tetrahydrofolate + 5'-Phosphoribosyl-N-formylglycinamide	2.1.2.2	NP1662A
R03504	2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- <=> Glycolaldehyde + 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine	4.1.2.25	
R00188	Adenosine 3',5'-bisphosphate + H2O => AMP + Orthophosphate	3.1.3.7	
R02412	ATP + Shikimate <=> ADP + Shikimate 3-phosphate	2.7.1.71	NP3982A
R02024	Thioredoxin + CDP => dCDP + Oxidized thioredoxin + H2O	1.17.4.1	NP3346A NP4434A NP5066A NP6166A NP6168A
R01512	ATP + 3-Phospho-D-glycerate <=> ADP + 3-Phospho-D-glyceroyl phosphate	2.7.2.3	NP4130A
R00768	L-Glutamine + D-Fructose 6-phosphate => L-Glutamate + D-Glucosamine 6-phosphate	2.6.1.16	NP2384A NP4654A
R01057	D-Ribose 1-phosphate <=> D-Ribose 5-phosphate	5.4.2.2	NP4660A NP4676A NP5158A NP3514A
R04204	2-Methylbut-2-enoyl-CoA + H2O => (2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R01015	D-Glyceraldehyde 3-phosphate <=> Glycerone phosphate	5.3.1.1	NP2182A NP3716A
R04158	Geranylgeranyl diphosphate + Glycerol 1-phosphate => Pyrophosphate + sn-3-O-(Geranylgeranyl)glycerol 1-phosphate	2.5.1.41	
R03045	3-Hydroxypropionyl-CoA <=> Propenoyl-CoA + H2O	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R02748	Deoxyinosine + Orthophosphate => Hypoxanthine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.4	NP3958A
R90138	Hexadecanoic acid (E) + ATP + H2O => Hexadecanoic acid + ADP + Orthophosphate + H+		
R90058	Biotin (E) + H+ (pmf associated) => Biotin + H+		
R05048	2,5-Diaminopyrimidine nucleoside triphosphate => 2,5-Diamino-6-(5'-triphosphoryl-3',4'-trihydroxy-2'-oxopentyl)-	6.-.-.- 3.5.4.16	

Id	Definition	EC No.	Genes
R00084	Porphobilinogen + H2O <=> Hydroxymethylbilane + NH3	2.5.1.61	NP1326A
R00867	ATP + D-Fructose => ADP + D-Fructose 6-phosphate	2.7.1.1	NP3184A
		2.7.1.4	
R00286	UDPglucose + H2O + NAD+ <=> UDPglucuronate + NADH + H+	1.1.1.22	NP2322A NP4668A
R01737	ATP + D-Gluconic acid => ADP + 6-Phospho-D-gluconate	2.7.1.12	
R00131	Urea + H2O => CO2 + NH3	3.5.1.5	NP2008A NP2010A NP2012A NP2014A NP2016A NP2018A NP2020A
R00137	ATP + Nicotinamide D-ribonucleotide => Pyrophosphate + NAD+	2.7.7.1	NP0310A NP0908A
R90073	Hydrogen (E) <=> Hydrogen		
R00935	(S)-Methylmalonate semialdehyde + CoA + NAD+ => Propanoyl-CoA + CO2 + NADH + H+	1.2.1.27	NP1686A NP3020A
R01858	dGTP + Pyruvate => dGDP + Phosphoenolpyruvate	2.7.1.40	NP1746A
R02147	Guanosine + Orthophosphate <=> Guanine + D-Ribose 1-phosphate	2.4.2.1	
R01395	ATP + Carbamate => ADP + Carbamoyl phosphate	2.7.2.2	NP3632A
R04001	3-Isopropylmalate <=> 2-Isopropylmaleate + H2O	4.2.1.33	NP2192A NP2194A
R04457	4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine + 3, 4-Dihydroxy-2-butanone 4-phosphate => 6,7-Dimethyl-8-(1-D-ribityl)lumazine + Orthophosphate + H2O	2.5.1.9	NP4412A
R90081	Thymidine (E) + H+ (pmf associated) => Thymidine + H+		
R04859	O-Acetyl-L-serine + Thiosulfate + Thioredoxin + H+ => L-Cysteine + Sulfite + Oxidized thioredoxin + Acetate	2.5.1.47 2.5.1.49	NP3116A NP4748A
R04747	Hexanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxooctanoyl-CoA	2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R90125	Succinate (E) + H+ (pmf associated) <=> Succinate + H+		
R01801	CDPdiacylglycerol + Glycerol 1-phosphate => CMP + Phosphatidylglycerophosphate	2.7.8.5	NP1334A NP2144A
R00190	Adenine + 5-Phospho-alpha-D-ribose 1-diphosphate => AMP + Pyrophosphate	2.4.2.7	NP1426A NP1254A
R03038	ATP + L-Alanine + tRNA(Ala) => AMP + Pyrophosphate + L-Alanyl-tRNA	6.1.1.7	NP0710A
R90146	AMP + Orthophosphate => D-Ribose 1,5-bisphosphate + Adenine		NP3958A
R00472	Acetyl-CoA + H2O + Glyoxylate => (S)-Malate + CoA	2.3.3.9	NP4430A
R00259	Acetyl-CoA + L-Glutamate => CoA + N-Acetyl-L-glutamate	2.3.1.1	
R03222	Protoporphyrinogen IX + Oxygen <=> Protoporphyrin + H2O	1.3.3.4	NP4520A
R03629	Melatonin + Reduced flavoprotein + Oxygen <=> 6-Hydroxymelatonin + Oxidized flavoprotein + H2O	1.14.14.1	NP2540A
R04745	(S)-Hydroxyoctanoyl-CoA + NAD+ <=> 3-Oxooctanoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R01213	Acetyl-CoA + 3-Methyl-2-oxobutanoic acid + H2O => 2-Isopropylmalate + CoA	2.3.3.13	NP0624A NP2206A NP2994A
R00355	L-Aspartate + 2-Oxoglutarate <=> Oxaloacetate + L-Glutamate	2.6.1.1	NP0824A NP1666A NP4024A NP4410A
R03220	Coproporphyrinogen III + Oxygen <=> Protoporphyrinogen IX + CO2 + H2O	1.3.3.3	
R00351	Acetyl-CoA + H2O + Oxaloacetate => Citrate + CoA	2.3.3.1 2.3.3.3	NP1314A
R03269	(R)-4'-Phosphopantothenoyl-L-cysteine => Pantetheine 4'-phosphate + CO2	4.1.1.36	NP1374A
R90110	Betaine (E) => Betaine		NP3588A
R02326	ATP + dCDP <=> ADP + dCTP	2.7.4.6	NP3666A
R03968	2-Isopropylmalate <=> 2-Isopropylmaleate + H2O	4.2.1.33	NP2192A NP2194A
R00702	trans,trans-Farnesyl diphosphate => Pyrophosphate + Presqualene diphosphate + H+	2.5.1.21	NP4556A NP0604A
R04199	2,3,4,5-Tetrahydrodipicolinate + NADP+ <=> L-2,3-Dihydrodipicolinate + NADPH + H+	1.3.1.26	NP1492A
R03758	Aminoacetone + CO2 <=> L-2-Amino-3-oxobutanoic acid	-	
R01513	3-Phospho-D-glycerate + NAD+ <=> 3-Phosphonooxypyruvate + NADH + H+	1.1.1.95	NP0272A NP4720A
			Continued on next page

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R01397	Carbamoyl phosphate + L-Aspartate => Orthophosphate + N-Carbamoyl-L-aspartate	2.1.3.2	NP3518A NP3520A
R01859	ATP + Propanoyl-CoA + HCO3- => ADP + Orthophosphate + (S)-2-Methyl-3-oxopropanoyl-CoA	6.4.1.3	NP4250A NP4364A NP3464A
R01701	3-Methyl-2-oxobutanoic acid + Lipoamide => S-(2-Methylpropanoyl)-dihydrolipoamide + CO2	1.2.4.4	NP0560A NP2542A NP0558A
R02090	ATP + dGMP <=> ADP + dGDP	2.7.4.8	
R00378	dTTP => Pyrophosphate + dTMP	2.7.7.7	NP0854A NP0476A NP0482A NP2918A
R03990	Tetradecanoyl-CoA + FAD => trans-Tetradec-2-enoyl-CoA + FADH2	1.3.99.13 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R90123	Hypoxanthine (E) + H+ (pmf associated) => Hypoxanthine + H+		NP1252A
R00451	meso-2,6-Diaminoheptanedioate <=> L-Lysine + CO2	4.1.1.20	NP1646A
R01244	Adenine + H2O => Hypoxanthine + NH3	3.5.4.2	NP0972A NP2032A NP4788A
R02198	L-Isoleucine + 2-Oxoglutarate <=> 3-Methyl-2-oxopentanoate + L-Glutamate	2.6.1.42	NP0798A NP5036A
R01197	Oxidized ferredoxin + 2-Oxoglutarate + CoA <=> Reduced ferredoxin + Succinyl-CoA + CO2	1.2.7.3	NP1232A NP1234A NP3012A
R01157	Agmatine + H2O <=> Putrescine + Urea	3.5.3.11	NP3022A NP4754A
R00428	GTP + H2O => Formamidopyrimidine nucleoside triphosphate	3.5.4.16	
R00226	Pyruvate => (S)-2-Acetolactate + CO2	2.2.1.6	NP2200A NP2202A
R04638	2-Amino-4-hydroxy-6-(erythro-1,2,3-trihydroxypropyl) + H2O => Dihydroneopterin phosphate + Pyrophosphate	3.6.1.-	
R90148	Glycine + Tetrahydropteroyltri-L-glutamate => NH3 + CO2 + H+ + 5-Methyltetrahydropteroyltri-L-glutamate		
R03508	1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate <=> Indoleglycerol phosphate + CO2 + H2O	4.1.1.48	NP3166A
R00257	ATP + Deamino-NAD+ + L-Glutamine + H2O => AMP + Pyrophosphate + NAD+ + L-Glutamate	6.3.5.1	NP0054A NP3828A
R00529	ATP + Sulfate <=> Pyrophosphate + Adenylylsulfate	2.7.7.4	NP4570A
R01658	Dimethylallyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + Geranyl diphosphate	2.5.1.1 2.5.1.10 2.5.1.29	NP3696A NP4556A NP0604A
R03197	Uroporphyrinogen III <=> Coproporphyrinogen III + CO2	4.1.1.37	
R01570	Thymidine + Orthophosphate => Thymine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.4	NP3958A
R05064	(S)-3-Hydroxyisobutyryl-CoA + H2O => CoA + (S)-3-Hydroxyisobutyrate	3.1.2.4	
R03877	ATP + Protoporphyrin + Magnesium + H2O <=> ADP + Orthophosphate + Magnesium protoporphyrin + H+	6.6.1.1	NP1094A
R90076	ATP + H2O + Glycerol 1-phosphate (E) => ADP + Orthophosphate + H+ + Glycerol 1-phosphate		
R90127	Adenosine => Adenosine (E)		
R06558	Adenosyl cobinamide + GTP <=> Adenosyl cobinamide phosphate + GDP	2.7.1.156	
R90108	Cl- (E) + H+ (pmf associated) => Cl- + H+		
R01665	ATP + dCMP <=> ADP + dCDP	2.7.4.14	NP4914A
R01163	L-Histidinal + H2O + NAD+ <=> L-Histidine + NADH + H+	1.1.1.23	NP2876A
R00754	Ethanol + NAD+ <=> Acetaldehyde + NADH + H+	1.1.1.1	NP1260A
R05226	Adenosyl cobyrinate hexaamide + 1-Aminopropan-2-ol <=> Adenosyl cobinamide	6.3.1.10	NP5308A
R90062	L-Ornithine + L-Arginine (E) => L-Ornithine (E) + L-Arginine		
R90078	Oxalate (E) + Formate => Oxalate + Formate (E)		NP3798A
R02472	(R)-Pantoate + NADP+ <=> 2-Dehydropantoate + NADPH	1.1.1.169	NP0942A
R01515	3-Phospho-D-glyceroyl phosphate + H2O => 3-Phospho-D-glycerate + Orthophosphate	3.6.1.7	NP3750A
R90129	D-Fructose 1,6-bisphosphate + NAD+ => 4-Ketofructose-1,6-bisphosphate + NADH		
R90041	L-Phenylalanine (E) + H+ (pmf associated) => L-Phenylalanine + H+		

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R00161	ATP + FMN => Pyrophosphate + FAD	2.7.7.2	
R90140	Cyclic pyranopterin monophosphate => Molybdopterin + Pyrophosphate		NP2274A NP2500A NP3946A NP5020A NP0052A
R90064	2,3-Bis-O-(geranylgeranyl)glycerol 1-phosphate + H2O + NADPH + H+ => 2,3-Di-O-phytanyl-sn-glycerol + Orthophosphate + NADP+		
R05577	tRNA(Asp) + L-Aspartate + ATP => L-Aspartyl-tRNA(Asp) + Pyrophosphate + AMP	6.1.1.12	NP1168A
R01138	dATP + Pyruvate => dADP + Phosphoenolpyruvate	2.7.1.40	NP1746A
R02291	L-Aspartate 4-semialdehyde + Orthophosphate + NADP+ <=> 4-Phospho-L-aspartate + NADPH + H+	1.2.1.11	NP1988A
R04798	zeta-Carotene => Neurosporene + Hydrogen	1.14.99.30	NP4764A NP0204A NP1630A
R02685	(3S)-3-Hydroxyacyl-CoA <=> trans-2,3-Dehydroacyl-CoA + H2O	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R01867	(S)-Dihydroorotate + Oxygen <=> Orotate + H2O2	1.3.3.1	NP0428A
R01884	Creatinine + H2O <=> Creatine	3.5.2.10	NP2120A
R07770	ATP + Lipoate => Pyrophosphate + Lipoyl-AMP	2.7.7.63	NP5116A
R05812	Cobalt-precorrin 6 + NADPH + H+ <=> Cobalt-dihydro-precorrin 6 + NADP+	1.3.1.54	
R00615	Thiamin diphosphate + H2O => Thiamin monophosphate + Orthophosphate	3.6.1.15	NP4426A
R04439	(R)-2,3-Dihydroxy-3-methylbutanoate + NADP+ <=> (S)-2-Acetolactate + NADPH + H+	1.1.1.86	NP2198A
R01958	Formylkynurenine + H2O <=> Formate + L-Kynurenine	3.5.1.49 3.5.1.9	NP6204A
R02085	(S)-3-Hydroxy-3-methylglutaryl-CoA <=> 3-Methylglutaconyl-CoA + H2O	4.2.1.18	
R00698	L-Phenylalanine <=> 2-Phenylacetamide	1.11.1.7	NP2708A
R02245	ATP + (R)-Mevalonate => ADP + (R)-5-Phosphomevalonate	2.7.1.36	NP2850A
R02568	D-Fructose 1-phosphate <=> Glycerone phosphate + D-Glyceraldehyde	4.1.2.13	NP1594A NP3160A
R90102	3-Phospho-D-Glycerate (E) + ATP + H2O => 3-Phospho-D-glycerate + ADP + Orthophosphate + H+		NP5220A
R00742	ATP + Acetyl-CoA + HCO3- => ADP + Orthophosphate + Malonyl-CoA	6.4.1.2	NP4250A NP4364A NP3464A
R90038	L-Tryptophan (E) + H+ (pmf associated) <=> L-Tryptophan + H+		
R90002	Oxygen + H+ + Halocyanin (reduced) => H2O + H+ (pmf associated) + Halocyanin (oxidized)	1.9.3.1	NP2448A NP2450A NP2456A NP2962A NP2964A NP2966A NP2968A NP4294A NP4296A
R01870	Orotidine 5'-phosphate + Pyrophosphate <=> Orotate + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.10	NP1256A
R00342	(S)-Malate + NAD+ <=> Oxaloacetate + NADH + H+	1.1.1.37	NP0498A
R04030	ATP + 2-Succinylbenzoate + CoA => AMP + Pyrophosphate + 2-Succinylbenzoyl-CoA	6.2.1.26	NP2738A
R05679	Glycerol 1-phosphate + NAD+ <=> Glycerone phosphate + NADH + H+	1.1.1.261	NP4492A
R90045	L-Leucine (E) + H+ (pmf associated) <=> L-Leucine + H+		
R01541	ATP + 2-Dehydro-3-deoxy-D-gluconate <=> ADP + 2-Dehydro-3-deoxy-6-phospho-D-gluconate	2.7.1.45	NP3184A
R03314	L-Glutamate 5-semialdehyde <=> (S)-1-Pyrroline-5-carboxylate + H2O	-	
R00261	L-Glutamate => 4-Aminobutanoate + CO2	4.1.1.19 4.1.1.15	NP1194A
R05066	(S)-3-Hydroxyisobutyrate + NAD+ <=> (S)-Methylmalonate semialdehyde + NADH + H+	1.1.1.35	NP1906A NP2754A NP2630A NP4322A
R01130	IMP + NAD+ + H2O => Xanthosine 5'-phosphate + NADH + H+	1.1.1.205	NP3384A NP3080A
R00705	3-Oxopropanoate + CoA + NAD+ => Acetyl-CoA + CO2 + NADH + H+	1.2.1.18	NP1686A NP3020A
R04097	3-Methylbutanoyl-CoA + Dihydrolipoamide => S-(3-Methylbutanoyl)-dihydrolipoamide + CoA	2.3.1.-	
R02094	ATP + dTMP <=> ADP + dTDP	2.7.4.9	NP2578A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R01863	Inosine + Orthophosphate <=> Hypoxanthine + D-Ribose 1-phosphate	2.4.2.1	
R00734	L-Tyrosine + 2-Oxoglutarate <=> 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate	2.6.1.9 2.6.1.1 2.6.1.5 2.6.1.57	NP0824A NP1666A NP2140A NP4024A NP4410A
R90085	Deoxyuridine (E) + H+ (pmf associated) => Deoxyuridine + H+		
R00441	GTP => Pyrophosphate + GMP	2.7.7.6	NP0114A NP0116A NP0112A NP0110A NP2834A NP5080A NP5078A NP3682A NP0108A NP2844A NP4162A NP2842A NP5234A
R01049	ATP + D-Ribose 5-phosphate <=> AMP + 5-Phospho-alpha-D-ribose 1-diphosphate	2.7.6.1	NP0606A
R00955	UDPglucose + alpha-D-Galactose 1-phosphate <=> D-Glucose 1-phosphate + UDP-D-galactose	2.7.7.12	NP4652A NP4656A NP4674A NP4680A NP2386A NP4276A
R90049	L-Alanine (E) + H+ (pmf associated) <=> L-Alanine + H+		
R01288	O-Succinyl-L-homoserine + Sulfide => L-Homocysteine + Succinate	2.5.1.48	NP4746A
R00617	ATP + Thiamin monophosphate => ADP + Thiamin diphosphate	2.7.4.16	NP2280A
R00435	ATP => Pyrophosphate + AMP	2.7.7.6	NP0114A NP0116A NP0112A NP0110A NP2834A NP5080A NP5078A NP3682A NP0108A NP2844A NP4162A NP2842A NP5234A
R01251	(S)-1-Pyrroline-5-carboxylate + NADPH + H+ <=> L-Proline + NADP+	1.5.1.2	NP3974A
R03656	ATP + L-Isoleucine + tRNA(Ile) => AMP + Pyrophosphate + L-Isoleucyl-tRNA(Ile)	6.1.1.5	NP0610A
R00203	Methylglyoxal + NAD+ + H2O => Pyruvate + NADH + H+	1.2.1.22 1.2.1.23	NP1686A NP3020A
R00707	(S)-1-Pyrroline-5-carboxylate + NAD+ + H2O => L-Glutamate + NADH + H+	1.5.1.12	NP3020A NP1686A
R90132	6-Deoxy-5-Ketofructose 1 phosphate + NAD+ <=> 6-Deoxy-5-ketomannitol-1-phosphate + NADH		
R03537	2',3'-Cyclic AMP + H2O => 3'-AMP	3.1.4.16	NP1854A NP2590A
R03458	5-Amino-6-(5'-phosphoribosylamino)uracil + NADPH => 5-Amino-6-(5'-phosphoribitylamino)uracil + NADP+	1.1.1.193	NP2672A
R03947	Precorrin 2 + NAD+ <=> Sirohydrochlorin + NADH + H+	1.3.1.76	NP4500A
R04364	2,3,4,5-Tetrahydrodipicolinate + Acetyl-CoA <=> N-Acetyl-L-2-amino-6-oxopimelate + CoA	2.3.1.- 2.3.1.89	
R04209	1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate <=> Aminoimidazole ribotide + CO2	4.1.1.21	NP2286A NP2290A
R90036	L-Glutamate (E) + H+ (pmf associated) <=> L-Glutamate + H+		
R02016	Oxidized thioredoxin + NADPH + H+ => Thioredoxin + NADP+	1.8.1.9	NP2680A
R00216	(S)-Malate + NADP+ => Pyruvate + CO2 + NADPH	1.1.1.40	NP1772A NP0132A
R01322	ADP + Orthophosphate + (3S)-Citryl-CoA => ATP + Citrate + CoA	6.2.1.18	
R01126	IMP + H2O => Inosine + Orthophosphate	3.1.3.5	NP2590A NP1854A
R90104	Arsenite (E) + ATP + H2O => Arsenite + ADP + Orthophosphate + H+		NP1164A NP1382A NP1818A NP6058A
R90066	Sodium (E) => Sodium		
R01786	ATP + alpha-D-Glucose <=> ADP + alpha-D-Glucose 6-phosphate	2.7.1.1 2.7.1.2	NP3184A
R00582	O-Phospho-L-serine + H2O => L-Serine + Orthophosphate	3.1.3.3	NP0274A
R03067	2-Amino-7,8-dihydro-4-hydroxy-6-(diphosphooxymethyl)pteridine + 4-Aminobenzoate => Pyrophosphate + Dihydropteroate	2.5.1.15	NP1478A NP3770A
R01221	5,10-Methylenetetrahydrofolate + NH3 + CO2 + NADH + H+ <=> Glycine + Tetrahydrofolate + NAD+	2.1.2.10	NP4774A
R05575	(Hydroxymethylphenyl)succinyl-CoA + NAD+ <=> Benzoylsuccinyl-CoA + NADH + H+	1.1.1.35	NP1906A NP2754A NP2630A NP4322A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R00575	ATP + L-Glutamine + HCO3- + H2O => ADP + Orthophosphate + L-Glutamate + Carbamoyl phosphate	6.3.5.5	NP4506A NP4830A
R00032	beta-Carotene + Oxygen => Retinal	1.14.99.36	
R00268	Oxalosuccinate <=> 2-Oxoglutarate + CO2	1.1.1.42	NP2430A
R04095	3-Methylbutanoyl-CoA + FAD => 3-Methylcrotonyl-CoA + FADH2	1.3.99.3	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R07465	C15815 + L-Tyrosine + Iminoglycine => 4-Methyl-5-(2-phosphoethyl)-thiazole	ThiG	NP2394A NP2398A
R00549	ATP + Riboflavin => ADP + FMN	2.7.1.26	
R90083	Guanosine (E) + H+ (pmf associated) => Guanosine + H+		NP1252A
R02864	Siroheme + H+ <=> Iron + Sirohydrochlorin	4.99.1.4	NP4500A
R05808	Cobalt-precorrin 2 + S-Adenosyl-L-methionine <=> Cobalt-precorrin 3 + S-Adenosyl-L-homocysteine	2.1.1.130	NP1122A
R90043	L-Isoleucine (E) + H+ (pmf associated) <=> L-Isoleucine + H+		
R05220	Cob(I)yrinate a,c diamide + ATP <=> Adenosyl cobyrinate a,c diamide + Pyrophosphate + Orthophosphate	2.5.1.17	NP4974A
R01286	L-Cystathionine + H2O => L-Homocysteine + NH3 + Pyruvate	4.4.1.8	NP4746A
R03013	L-Histidinol phosphate + H2O <=> L-Histidinol + Orthophosphate	3.1.3.15	
R90004	Reduced ferredoxin + Menaquinone + H+ => Oxidized ferredoxin + Menaquinol + H+ (pmf associated)	1.6.5.3	NP2292A NP2294A NP2296A NP2298A NP2300A NP2304A NP2306A NP2308A NP2310A NP2312A NP2314A NP2406A NP3508A
R00430	GTP + Pyruvate => GDP + Phosphoenolpyruvate	2.7.1.40	NP1746A
R07460	L-Cysteine => Thiamine biosynthesis intermediate 3 + L-Alanine	2.8.1.7	NP0936A NP0790A
R01063	D-Glyceraldehyde 3-phosphate + Orthophosphate + NADP+ <=> 3-Phospho-D-glyceroyl phosphate + NADPH + H+	1.2.1.59	NP0012A
R02740	alpha-D-Glucose 6-phosphate <=> D-Fructose 6-phosphate	5.3.1.9	NP4992A
R01969	Deoxyguanosine + Orthophosphate <=> Guanine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.4	NP3958A
R03443	N-Acetyl-L-glutamate 5-semialdehyde + Orthophosphate + NADP+ <=> N-Acetyl-L-glutamate 5-phosphate + NADPH + H+	1.2.1.38	NP5260A
R03662	ATP + L-Serine + tRNA(Ser) => AMP + Pyrophosphate + L-Seryl-tRNA(Ser)	6.1.1.11	NP1692A
R03224	(3R)-3-Hydroxyacyl-CoA <=> cis-2,3-Dehydroacyl-CoA + H2O	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R00124	ATP + ADP <=> ADP + ATP	2.7.4.6	NP3666A
R01771	ATP + L-Homoserine => ADP + O-Phospho-L-homoserine	2.7.1.39	NP4524A
R02237	ATP + Dihydropteroate + L-Glutamate => ADP + Orthophosphate + Dihydrofolate	6.3.2.17	NP1478A
R04673	2-Oxobutanoate + 2-(alpha-Hydroxyethyl)thiamine diphosphate => (S)-2-Aceto-2-hydroxybutanoate + Thiamin diphosphate	2.2.1.6	NP1904A NP2200A NP2202A
R90118	Ferrichrome (E) + ATP + H2O => Ferrichrome + ADP + Orthophosphate + H+		NP3816A NP3818A
R00009	H2O2 => Oxygen + H2O	1.11.1.6	NP2708A
R90087	Deoxyguanosine (E) + H+ (pmf associated) => Deoxyguanosine + H+		
R02297	Xanthosine + Orthophosphate <=> Xanthine + D-Ribose 1-phosphate	2.4.2.1	
R04109	L-Glutamyl-tRNA(Glu) + NADPH <=> (S)-4-Amino-5-oxopentanoate + tRNA(Glu) + NADP+ + H2O	1.2.1.70	NP4502A
R05218	Cob(II)yrinate a,c diamide + H+ + NADH <=> Cob(I)yrinate a,c diamide + NAD+	1.16.8.1	

Id	Definition	EC No.	Genes
R04749	(S)-Hydroxyhexanoyl-CoA <=> trans-Hex-2-enoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R00410	Succinyl-CoA + Acetoacetate <=> Succinate + Acetoacetyl-CoA	2.8.3.5	NP6032A NP6034A
R01280	ATP + Hexadecanoic acid + CoA <=> AMP + Palmitoyl-CoA + Pyrophosphate	6.2.1.3	NP2626A NP3386A NP3422A NP4150A NP4174A NP4256A
R07280	5-Amino-6-(5'-phosphoribitylamino)uracil + H2O => 4-(1-D-Ribitylamino)-5-amino-2,6-dihydroxypyrimidine + Orthophosphate	3.1.3.-	NP2774A
R00512	ATP + CMP <=> ADP + CDP	2.7.4.14	NP4914A
R90068	Potassium (E) => Potassium		NP0220A NP0330A NP0332A NP3602A NP3604A NP4902A NP4904A NP4906A NP6110A NP6112A
R04336	L-2-Amino-6-oxoheptanedioate <=> 2,3,4,5-Tetrahydrodipicolinate + H2O	-	
R00192	S-Adenosyl-L-homocysteine + H2O <=> Adenosine + L-Homocysteine	3.3.1.1	NP0968A
R90089	D-Glucose (E) + ATP + H2O => D-Glucose + ADP + H+ + Orthophosphate		NP0530A
R02750	Deoxyribose + ATP => 2-Deoxy-D-ribose 5-phosphate + ADP	2.7.1.15	NP3184A
R01388	Hydroxypyruvate + NADH + H+ => D-Glycerate + NAD+	1.1.1.81	NP1162A
R90117	Dipeptide (E) + ATP + H2O => Dipeptide + ADP + Orthophosphate + H+		NP0758A NP0760A NP0762A NP0764A NP0766A NP3578A NP3580A NP3582A NP3584A
R02101	dUMP + 5,10-Methylenetetrahydrofolate => Dihydrofolate + dTMP	2.1.1.45	NP2924A
R00509	ATP + Adenylylsulfate => ADP + 3'-Phosphoadenylyl sulfate	2.7.1.25	
R00199	ATP + Pyruvate + H2O => AMP + Phosphoenolpyruvate + Orthophosphate	2.7.9.2	NP1196A
R90095	Acetate (E) + H+ (pmf associated) <=> Acetate + H+		
R03857	Lauroyl-CoA + FAD => 2-trans-Dodecenoyl-CoA + FADH2	1.3.99.13 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A NP4534A NP6036A
R90109	Pantothenate (E) + Sodium (E) [pmf] => Pantothenate + Sodium		
R02021	Thioredoxin + 3'-Phosphoadenylyl sulfate => Oxidized thioredoxin + Sulfite + Adenosine 3',5'-bisphosphate + H+	1.8.4.8	
R90139	GTP + S-Adenosyl-L-methionine => Cyclic pyranopterin monophosphate + S-Adenosyl-L-homocysteine		NP2822A NP3746A NP1040A
R00959	D-Glucose 1-phosphate <=> alpha-D-Glucose 6-phosphate	5.4.2.2	NP4660A NP4676A NP5158A NP3514A
R03991	Tetradecanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxopalmitoyl-CoA	2.3.1.16 2.3.1.154	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R04378	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + Pyrophosphate <=> 5-Amino-4-imidazolecarboxyamide + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.7	NP1426A NP1254A
R05070	(R)-2,3-Dihydroxy-3-methylpentanoate <=> (3R)-3-Methyl-2-oxopentanoic acid + H2O	4.2.1.9	NP4926A
R04780	H2O + D-Fructose 1,6-bisphosphate => Orthophosphate + D-Fructose 6-phosphate	3.1.3.11	NP1592A
R01778	(3S)-3-Hydroxyacyl-CoA + NAD+ <=> 3-Oxoacyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R03165	Hydroxymethylbilane <=> Uroporphyrinogen III + H2O	4.2.1.75	NP1330A
R01326	ATP + D-Mannose <=> ADP + D-Mannose 6-phosphate	2.7.1.1 2.7.1.7	NP3184A
R00585	L-Serine + Pyruvate <=> Hydroxypyruvate + L-Alanine	2.6.1.51	NP0884A NP2578A
R90091	6-Deoxy-5-Ketofructose 1 phosphate + L-Aspartate 4-semialdehyde => Compound I + Hydroxypyruvaldehyde 3-P		NP3160A
R04138	ATP + 3-Methylcrotonyl-CoA + HCO3- => ADP + Orthophosphate + 3-Methylglutaconyl-CoA	6.4.1.4	

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R00489	L-Aspartate => beta-Alanine + CO2	4.1.1.15	NP1194A
R06531	L-Threonine <=> L-Threonine O-3-phosphate		
R01248	(S)-1-Pyrroline-5-carboxylate + NADH + H+ <=> L-Proline + NAD+	1.5.1.2	NP3974A
R03472	Aminoimidazole ribotide => 4-Amino-5-hydroxymethyl-2-methylpyrimidine		NP2210A
R02003	Geranyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + trans,trans-Farnesyl diphosphate	2.5.1.1 2.5.1.10 2.5.1.29	NP3696A NP4556A NP0604A
R90042	L-Asparagine (E) + H+ (pmf associated) <=> L-Asparagine + H+		
R03654	ATP + Glycine + tRNA(Gly) => AMP + Pyrophosphate + Glycyl-tRNA(Gly)	6.1.1.14	NP0328A
R01715	Chorismate <=> Prephenate	5.4.99.5	NP3980A
R04230	ATP + D-4'-Phosphopantothenate + L-Cysteine => ADP + Orthophosphate + (R)-4'-Phosphopantothenoyl-L-cysteine	6.3.2.5	NP1374A
R00317	Acetyl phosphate + H2O => Acetate + Orthophosphate	3.6.1.7	NP3750A
R90047	L-Threonine (E) + H+ (pmf associated) => L-Threonine + H+		
R07463	Glycine => Iminoglycine	1.4.3.19	
R03083	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate <=> 3-Dehydroquinate + Orthophosphate	4.2.3.4	NP2238A
R02508	Cystathionine + Succinate <=> O-Succinyl-L-homoserine + L-Cysteine	2.5.1.48	NP4746A
R01411	5-Methylcytosine + H2O <=> Thymine + NH3	3.5.4.1	NP0972A NP4788A
R06412	trans-2-Methyl-5-isopropylhexa-2,5-dienoyl-CoA + H2O <=> 3-Hydroxy-2,6-dimethyl-5-methylene-heptanoyl-CoA	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R90070	H2O (E) <=> H2O		
R04786	Phytoene => Phytofluene + Hydrogen	1.3.99.-	NP4764A NP0204A NP1630A
R02029	Phosphatidylglycerophosphate + H2O => Phosphatidylglycerol + Orthophosphate	3.1.3.27	
R10001	D-Glucose + NADP+ => D-Glucono-1,5-lactone + NADPH	1.1.1.119	NP0126A NP0254A
R04170	(S)-3-Hydroxydodecanoyl-CoA <=> 2-trans-Dodecenoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R01127	IMP + H2O <=> 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	3.5.4.10	NP0732A
R90097	Formate => Formate (E)		
R01899	Isocitrate + NADP+ <=> Oxalosuccinate + NADPH + H+	1.1.1.42	NP2430A
R00332	ATP + GMP <=> ADP + GDP	2.7.4.8	
R00253	ATP + L-Glutamate + NH3 => ADP + Orthophosphate + L-Glutamine	6.3.1.2	NP0076A NP4376A
R04740	(S)-3-Hydroxytetradecanoyl-CoA <=> trans-Tetradec-2-enoyl-CoA + H2O	4.2.1.74 4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R01776	Acetyl-CoA + L-Homoserine => CoA + O-Acetyl-L-homoserine	2.3.1.31	NP0282A
R01716	Chorismate + L-Glutamine <=> 4-amino-4-deoxychorismate + L-Glutamate	6.3.5.8	
R04509	ATP + 4-Amino-2-methyl-5-phosphomethylpyrimidine => ADP + 2-Methyl-4-amino-5-hydroxymethylpyrimidine diphosphate	2.7.4.7	NP0546A
R90131	2-amino-3,7-dideoxy-hept-6-ulsonic acid => 3-Dehydroquinate		NP2238A
R01083	N6-(1,2-Dicarboxyethyl)-AMP <=> Fumarate + AMP	4.3.2.2	NP1658A
R00920	ATP + Propanoate + CoA => ADP + Orthophosphate + Propanoyl-CoA	6.2.1.13	NP2858A
R00376	dGTP => Pyrophosphate + dGMP	2.7.7.7	NP0854A NP0476A NP0482A NP2918A

<div align="center">Table A.2 – continued from previous page</div>

Id	Definition	EC No.	Genes
R03778	Octanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxodecanoyl-CoA	2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R00183	AMP + H2O => Adenosine + Orthophosphate	3.1.3.5	NP2590A NP1854A
R05068	(R)-2,3-Dihydroxy-3-methylpentanoate + NADP+ <=> NADPH + (R)-3-Hydroxy-3-methyl-2-oxopentanoate	1.1.1.86	NP2198A
R90113	Potassium + H+ (pmf associated) => Potassium (E) + H+		NP0826A NP0828A NP0830A NP0832A NP0834A NP0836A NP0838A NP0840A NP0842A NP5056A
R90055	Sulfate (E) + ATP + H2O => Sulfate + ADP + Orthophosphate + H+		NP1412A NP3040A NP3042A NP3044A NP3046A NP3928A NP3930A NP3932A
R02018	Thioredoxin + UDP => dUDP + Oxidized thioredoxin + H2O	1.17.4.1	NP3346A NP4434A NP5066A NP6166A NP6168A
R06895	Coproporphyrinogen III + S-Adenosyl-L-methionine <=> Protoporphyrinogen IX + CO2 + L-Methionine + 5'-Deoxyadenosine	1.3.99.22	
R05814	Cobalt-precorrin 8 <=> Cobyrinate	5.4.1.2	NP1090A
R90053	Orthophosphate (E) + ATP + H2O => ADP + H+ + Orthophosphate		NP1978A NP1980A NP1982A NP1984A NP6098A NP6100A NP6102A NP6104A
R01071	Phosphoribosyl-ATP + Pyrophosphate <=> ATP + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.17	NP1062A
R03657	ATP + L-Leucine + tRNA(Leu) => AMP + Pyrophosphate + L-Leucyl-tRNA	6.1.1.4	NP0452A
R02719	Xanthosine 5'-phosphate + H2O => Xanthosine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R01066	2-Deoxy-D-ribose 5-phosphate <=> D-Glyceraldehyde 3-phosphate + Acetaldehyde	4.1.2.4	NP3200A
R05612	all-trans-Hexaprenyl diphosphate + Isopentenyl diphosphate => all-trans-Heptaprenyl diphosphate + Pyrophosphate	2.5.1.30	NP0604A NP4556A
R01676	Guanine + H2O => Xanthine + NH3	3.5.4.3	NP0972A NP2032A NP4788A
R90093	Cobamide coenzyme (E) + ATP + H2O => Cobamide coenzyme + ADP + H+ + Orthophosphate		NP3814A NP3816A NP3818A
R02735	LL-2,6-Diaminoheptanedioate <=> meso-2,6-Diaminoheptanedioate	5.1.1.7	
R04754	Decanoyl-CoA + FAD => trans-Dec-2-enoyl-CoA + FADH2	1.3.99.13 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R03348	Pyridine-2,3-dicarboxylate + 5-Phospho-alpha-D-ribose 1-diphosphate => Nicotinate D-ribonucleotide + Pyrophosphate + CO2	2.4.2.19	NP2414A
R05222	Adenosyl cobinamide phosphate + GTP <=> Adenosine-GDP-cobinamide + Pyrophosphate	2.7.7.62	
R01324	Citrate <=> Isocitrate	4.2.1.3	NP0404A NP1994A
R00966	UMP + Pyrophosphate => Uracil + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.9	NP1408A
R01547	ATP + dAMP <=> ADP + dADP	2.7.4.3	NP4910A
R90001	Menaquinol + H+ + Halocyanin (oxidized) => Menaquinone + H+ (pmf associated) + Halocyanin (reduced)	1.10.2.2	
R03650	ATP + L-Cysteine + tRNA(Cys) => AMP + Pyrophosphate + L-Cysteinyl-tRNA(Cys)	6.1.1.16	NP4440A
R04203	(2S,3S)-3-Hydroxy-2-methylbutanoyl-CoA + NAD+ <=> 2-Methylacetoacetyl-CoA + NADH	1.1.1.35	NP1906A NP2754A NP2630A NP4322A
R01878	Cytidine + H2O => Uridine + NH3	3.5.4.5	NP3512A
R04441	(R)-2,3-Dihydroxy-3-methylbutanoate <=> 3-Methyl-2-oxobutanoic acid + H2O	4.2.1.9	NP4926A
R01354	ATP + Propanoate => Pyrophosphate + Propinol adenylate	6.2.1.1	NP4240A NP4242A NP5128A NP5132A NP0164A
R04621	Dihydroneopterin phosphate + H2O => 2-Amino-4-hydroxy-6-(D-erythro-1,2,3-trihydroxypropyl)-7,8- + Orthophosphate	3.6.1.-	
R04737	(S)-3-Hydroxyhexadecanoyl-CoA + NAD+ <=> 3-Oxopalmitoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A

<div align="right">Continued on next page</div>

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R01054	ADPribose + H2O => AMP + D-Ribose 5-phosphate	3.6.1.13	NP0392A
R00985	Chorismate + NH3 <=> Anthranilate + Pyruvate + H2O	4.1.3.27	NP0800A NP0802A NP3342A NP3344A
R01734	Dimethylbenzimidazole <=> Riboflavin	-	
R00357	L-Aspartate + H2O + Oxygen => Oxaloacetate + NH3 + H2O2	1.4.3.16	NP2416A
R02301	ATP + 5-Formyltetrahydrofolate => ADP + Orthophosphate + 5,10-Methenyltetrahydrofolate	6.3.3.2	NP2438A
R01857	ATP + dGDP <=> ADP + dGTP	2.7.4.6	NP3666A
R01528	6-Phospho-D-gluconate + NADP+ => D-Ribulose 5-phosphate + CO2 + NADPH + H+	1.1.1.44	NP0286A
R01073	N-(5-Phospho-D-ribosyl)anthranilate + Pyrophosphate <=> Anthranilate + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.18	NP0532A NP3062A NP3338A
R02731	D-Fructose 2,6-bisphosphate + H2O => Orthophosphate + D-Fructose 6-phosphate	3.1.3.46	NP1332A
R00570	ATP + CDP <=> ADP + CTP	2.7.4.6	NP3666A
R90141	Molybdopterin + Molybdenum => Molybdenum cofactor		NP1440A NP1442A
R00289	UTP + D-Glucose 1-phosphate <=> Pyrophosphate + UDPglucose	2.7.7.9	NP4652A NP4656A NP4674A NP4680A NP2386A NP4276A
R90057	Thiamin (E) + ATP + H2O => Thiamin + ADP + Orthophosphate + H+		
R90040	L-Lysine (E) + H+ (pmf associated) <=> L-Lysine + H+		NP0040A NP0042A NP0646A NP2382A NP2676A NP4372A NP6260A
R00330	ATP + GDP <=> ADP + GTP	2.7.4.6	NP3666A
R02557	Deoxyadenosine + Orthophosphate <=> Adenine + 2-Deoxy-D-ribose 1-phosphate	2.4.2.4	NP3958A
R00586	L-Serine + Acetyl-CoA <=> O-Acetyl-L-serine + CoA	2.3.1.30	NP4172A
R01736	(R)-S-Lactoylglutathione + H2O => Glutathione + (R)-Lactate	3.1.2.6	NP0980A NP1136A NP1688A NP1898A NP1974A NP2670A NP2870A NP3944A NP3970A NP4032A NP0092A NP0642A
R90099	Cl- (E) + Photon => Cl-		
R03174	2-Methylbutanoyl-CoA + Dihydrolipoamide => S-(2-Methylbutanoyl)-dihydrolipoamide + CoA	2.3.1.-	
R06836	D-Ribose 1,5-bisphosphate + ATP <=> 5-Phospho-alpha-D-ribose 1-diphosphate + ADP	2.7.4.23	
R00481	L-Aspartate + Oxygen => Iminoaspartate + H2O2	1.4.3.16	NP2416A
R05341	Lycopene => gamma-Carotene		NP0652A
R02061	trans,trans-Farnesyl diphosphate + Isopentenyl diphosphate => Pyrophosphate + Geranylgeranyl diphosphate	2.5.1.29	NP3696A NP4556A NP0604A
R00230	Acetyl-CoA + Orthophosphate <=> CoA + Acetyl phosphate	2.3.1.8	NP1772A NP0132A
R07769	Lipoyl-[acp] + Apoprotein => Protein N6-(lipoyl)lysine + Acyl-carrier protein	2.3.1.181	NP0556A
R90143	(R)-5-Phosphomevalonate + ATP => Isopentenyl phosphate + ADP + Orthophosphate + CO2		NP1580A
R03180	4-Guanidinobutanamide + H2O <=> 4-Guanidinobutanoate + NH3	3.5.1.4	NP0178A NP0374A NP4212A
R00238	Acetyl-CoA <=> CoA + Acetoacetyl-CoA	2.3.1.9 2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A NP1632A NP4464A NP5280A
R01466	O-Phospho-L-homoserine + H2O => L-Threonine + Orthophosphate	4.2.3.1	
R01519	D-Glucono-1,5-lactone + H2O <=> D-Gluconic acid	3.1.1.17	
R05613	all-trans-Pentaprenyl diphosphate + Isopentenyl diphosphate => all-trans-Hexaprenyl diphosphate + Pyrophosphate	2.5.1.33	NP4556A NP0604A
R03647	ATP + L-Aspartate + tRNA(Asn) => AMP + Pyrophosphate + L-Aspartyl-tRNA(Asn)	6.1.1.12	NP1168A
R00036	5-Aminolevulinate <=> Porphobilinogen + H2O	4.2.1.24	NP0920A
R01227	GMP + H2O => Guanosine + Orthophosphate	3.1.3.5	NP1854A NP2590A
R00897	O-Acetyl-L-serine + Hydrogen sulfide => L-Cysteine + Acetate	2.5.1.47	NP3116A NP4748A
R03172	2-Methylbutanoyl-CoA + FAD => 2-Methylbut-2-enoyl-CoA + FADH2	1.3.99.3	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R04742	Decanoyl-CoA + Acetyl-CoA <=> CoA + 3-Oxododecanoyl-CoA	2.3.1.16	NP2214A NP2260A NP2606A NP2612A NP3420A NP3438A NP3650A NP4580A
R01373	Prephenate <=> Phenylpyruvate + H2O + CO2	4.2.1.51	NP0106A
R00908	beta-Alanine + 2-Oxoglutarate => 3-Oxopropanoate + L-Glutamate	2.6.1.19	NP6188A
R90030	L-Cysteine (E) + H+ (pmf associated) => L-Cysteine + H+		
R00236	Acetyl adenylate + CoA => AMP + Acetyl-CoA	6.2.1.1	NP4240A NP4242A NP5128A NP5132A NP0164A
R00229	ATP + Acetate + CoA => ADP + Acetyl-CoA + Orthophosphate	6.2.1.13	NP2858A
R01652	4-Methyl-2-oxopentanoate + CO2 <=> 3-Carboxy-4-methyl-2-oxopentanoate	-	
R90074	NH3 (E) <=> NH3		
R01123	Isopentenyl diphosphate <=> Dimethylallyl diphosphate	5.3.3.2	NP4826A NP5124A NP0360A
R02058	Acetyl-CoA + D-Glucosamine 6-phosphate => CoA + N-Acetyl-D-glucosamine 6-phosphate	2.3.1.4	
R02413	Shikimate + NADP+ <=> 3-Dehydroshikimate + NADPH + H+	1.1.1.25	NP0806A
R90059	Cobalt (E) + ATP + H2O => Cobalt + ADP + Orthophosphate + H+		NP1396A NP1394A NP1398A
R01279	Palmitoyl-CoA + FAD => trans-Hexadec-2-enoyl-CoA + FADH2	1.3.99.13 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R01978	(S)-3-Hydroxy-3-methylglutaryl-CoA + CoA <=> Acetyl-CoA + H2O + Acetoacetyl-CoA	2.3.3.10	NP2608A NP4836A
R02536	Phenylacetaldehyde + NAD+ + H2O <=> Phenylacetic acid + NADH + H+	1.2.1.5 1.2.1.39	NP1686A NP3020A
R07768	Octanoyl-[acp] + Sulfur + S-Adenosyl-L-methionine => Lipoyl-[acp] + L-Methionine + 5'-Deoxyadenosine	2.8.1.8	NP0562A
R02484	Deoxyuridine + Orthophosphate => Uracil + Deoxy-ribose 1-phosphate	2.4.2.4	NP3958A
R01392	Hydroxypyruvate + NADPH + H+ => D-Glycerate + NADP+	1.1.1.81	NP1162A
R90122	Guanine (E) + H+ (pmf associated) => Guanine + H+		NP1252A
R00089	ATP => 3',5'-Cyclic AMP + Pyrophosphate	4.6.1.1	NP3092A
R05617	1,4-Dihydroxy-2-naphthoate + all-trans-Octaprenyl diphosphate => 2-Demethylmenaquinone + Pyrophosphate + CO2	2.5.1.-	NP2734A
R03460	Phosphoenolpyruvate + Shikimate 3-phosphate <=> Orthophosphate + 5-O-(1-Carboxyvinyl)-3-phosphoshikimate	2.5.1.19	NP3078A
R01818	D-Mannose 6-phosphate <=> D-Mannose 1-phosphate	5.4.2.8	NP4660A NP4676A NP5158A NP3514A
R90032	L-Tyrosine (E) + H+ (pmf associated) <=> L-Tyrosine + H+		
R02749	2-Deoxy-D-ribose 1-phosphate <=> 2-Deoxy-D-ribose 5-phosphate	5.4.2.7	NP3514A
R00130	ATP + Dephospho-CoA => ADP + CoA	2.7.1.24	NP0870A
R04224	2-Methylprop-2-enoyl-CoA + H2O => (S)-3-Hydroxyisobutyryl-CoA	4.2.1.17	NP0032A NP0488A NP0916A NP1462A NP2256A NP2632A NP2754A NP2758A NP2936A NP3434A NP3830A NP4248A NP4322A
R00926	Propinol adenylate + CoA <=> AMP + Propanoyl-CoA	6.2.1.1	NP4240A NP4242A NP5128A NP5132A NP0164A
R04640	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)- <=> N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D-	5.3.1.16	NP1268A
R90120	Manganese (E) + ATP + H2O => Manganese + ADP + Orthophosphate + H+		
R01876	Uridine + Orthophosphate <=> Uracil + D-Ribose 1-phosphate	2.4.2.3	NP3180A NP3510A
R00710	Acetaldehyde + NAD+ + H2O => Acetate + NADH + H+	1.2.1.3	NP1686A
R00945	Tetrahydrofolate + L-Serine => 5,10-Methylenetetrahydrofolate + Glycine + H2O	2.1.2.1	NP2050A
R00691	L-Arogenate <=> L-Phenylalanine + H2O + CO2	4.2.1.51	NP0106A
R00866	ATP + D-Fructose => ADP + D-Fructose 1-phosphate	2.7.1.3	NP3184A

Table A.2 – continued from previous page

Id	Definition	EC No.	Genes
R02098	ATP + dUMP <=> ADP + dUDP	2.7.4.9	NP2478A
R00714	Succinate semialdehyde + NADP+ + H2O => Succinate + NADPH + H+	1.2.1.16	NP3020A NP1686A
R90072	Oxygen (E) <=> Oxygen		
R00924	Propanoyl-CoA + FAD <=> FADH2 + Propenoyl-CoA	1.3.99.3	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A NP4924A NP2700A
R00227	Acetyl-CoA + H2O => CoA + Acetate	3.1.2.1	
R04739	(S)-3-Hydroxytetradecanoyl-CoA + NAD+ <=> 3-Oxotetradecanoyl-CoA + NADH	1.1.1.35 1.1.1.211	NP1906A NP2754A NP2630A NP4322A
R02142	Xanthine + 5-Phospho-alpha-D-ribose 1-diphosphate => Xanthosine 5'-phosphate + Pyrophosphate	2.4.2.22	NP1426A NP1254A
R01175	Butanoyl-CoA + FAD => FADH2 + Crotonoyl-CoA	1.3.99.2 1.3.99.3 1.3.3.6	NP1596A NP1754A NP1870A NP2620A NP2628A NP2934A NP3004A NP3460A NP4050A NP4214A NP4254A NP6180A
R01290	L-Serine + L-Homocysteine => L-Cystathionine + H2O	4.2.1.22	NP3116A NP4748A

Bibliography

A, A. R., Hoffmann, E., Hagemann, M., and Berg, G. (2005). Synthesis of the compatible solutes glucosylglycerol and trehalose by saltstressed cells of *Stenotrophomonas strains*. *FEMS Microbiol. Lett.*, **243**, 219–226.

Aitken, D. M. and Brown, A. D. (1969). Citrate and glyoxylate cycles in the halophil, *Halobacterium salinarium*. *Biochim. Biophys. Acta*, **177**, 351–354.

Aivaliotis, M., Gevaert, K., Falb, M., Tebbe, A., Konstantinidis, K., Bisle, B., Klein, C., Martens, L., Staes, A., amd J Van Damme, E. T., Siedler, F., Pfeiffer, F., Vandekerckhove, J., and Oesterhelt, D. (2007). Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*. *J. Proteome Res.*, **6**, 2195–2204.

Alam, M. and Oesterhelt, D. (1984). Morphology, Function and Isolation of Halobacterial Flagella. *J. Mol. Biol.*, **176**, 459–475.

Albert, R. (2005). Scale-free networks in cell biology. *J. Cell. Sci.*, **118**, 4947–4957.

Albert, R., Jeong, H., and Barabasi, A. (2000). Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.

Allers, T. and Mevarech, M. (2005). Archeal genetics - the third way. *Nature Reviews Genetics*, **6**, 58–73.

Altekar, W. and Rangaswamy, V. (1992). Degradation of endogenous fructose during catabolism of sucrose and mannitol in halophilic archaebacteria. *Archives of Microbiol.*, **158**, 356–363.

Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Altschul, S., Madden, T., Schffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Alvarez-Vasquez, F., Sims, K., Cowart, L. A., Okamoto, Y., Voit, E., and Hannun, Y. (2004). Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae. *Nature*, **433**, 425–430.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.

Bamberg, E., Tittor, J., and Oesterhelt, D. (1993). Light-driven proton or chloride pumping by halorhodopsin. *Proc. Natl. Acad. Sci. USA*, **90**, 693–643.

Barnhardt, D., Koek, W., Juchem, T., Hampp, N., Coupland, J., and Halliwell, N. (2004). Bacteriorhodopsin as a high-resolution, high-capacity buffer for digital holographic measurements. *Meas. Sci. Technol.*, **15**, 639–646.

Beard, S. J., Hayes, P. K., and Walsby, A. E. (1997). Growth competition between *Halobacterium salinarum* strain PHH1 and mutants affected in gas vesicle synthesis. *Microbiology UK*, **143**, 467–473.

Besemer, J. and Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. *Nucleic Acids Res.*, **27**, 39113920.

Bhaumik, S. R. and Sonawat, H. M. (1994). Pyruvate metabolism in *Halobacterium salinarium* studied by intracellular 13C nuclear magnetic resonance spectroscopy. *J. Bacteriol.*, **176**, 2172–2176.

Birkeland, N. K. and Ratkje, S. K. (1985). Active uptake of glutamate in vesicles of *Halobacterium salinarum*. *Membr. Biochem.*, **6**, 1–17.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.

Bogomolni, R. and Spudich, J. L. (1982). Identification of a third rhodopsin-like pigment in phototactic *Halobacterium halobium*. *Proc. Natl. Acad. Sci. USA*, **79**, 6250–6254.

Bolhuis, H., Poele, E. M., and Rodriguez-Valera, F. (2004). Isolation and cultivation of Walsby's square archaeon. *Environ. Microbiol.*, **6**, 1287–91.

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F., and Oesterhelt, D. (2006). The genome of the square archaeon Haloquadratum walsbyi: life at the limits of water activity. *BMC Genomics*, **7**, 169.

Bork, P., Dandekar, T., Diaz-Lazcoz, Y., Eisenhaber, F., Huynen, M., and Yuan, Y. (1998). Predicting function: fron genes to

genomes and back. *J. Mol. Biol.*, **283**, 707–725.

Boyce, S. and Tipton, K. F. (2000). History of the enzyme nomenclature system. *Bioinformatics*, **16**, 34–40.

Brandt, K. and Ingvorsen, K. (1997). *Desulfobacter halotolerans* sp. nov., a halotolerant acetate-oxidizing sulfate-reducing bacterium isolated from sediments of Great Salt Lake, Utah. *Syst. Appl. Microbiol.*, **20**, 366–371.

Brandt, K., Patel, B., and Ingvorsen, K. (1999). *Desulfocella halophila* gen. nov., sp. nov., a halophilic, fatty-acid-oxidizing, sulfate-reducing bacterium isolated from sediments of Great Salt Lake, Utah. *Int. J. Syst. Bacteriol.*, **49**, 193–200.

Brown, A. D. and Simpson, J. R. (1972). Water relations of sugar-tolerant yeasts: the role of intracellular polyols. *J. Gen. Microbiol.*, **72**, 589–591.

Buchalo, A. S., Nevo, E., Wasser, S. P., Oren, A., and Molitoris, H. P. (1998). Fungal life in the extremely hypersaline water of the Dead Sea: first records. *Proc. R. Soc. London B*, **265**, 1461–1465.

Buchalo, A. S., Nevo, E., Wasser, S. P., Volz, H. P. M. P. A., Kurchenko, I., Lauer, I., and Rawal, B. (1999). Species diversity and biology of funji isolated from the Dead Sea. In S. P. Wasser, editor, *Evolutionary theory and processes: modern perspectives. Papers in honor of Eviatar Nevo*, pages 283–300. Kluwer Academic Publishers, Dordecht.

Burge, C. and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.*, **268**, 78–94.

Burns, D., Janssen, P., Itoh, T., Kamekura, M., Li, Z., Rodriguez-Valera, G. J. F., Bolhuis, H., and Dyall-Smith, M. (2007). *Haloquadratum walsbyi* gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.*, **57**, 387–392.

Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S. Y., Tissier, C., Zhang, P., and Karp, P. D. (2006). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.*, **34**, D511D516.

Chan, E. (2005). Advances in sequencing technology. *Mutation Research*, **573**, 13–40.

Choquet, C. G., Richards, J. C., Patel, G. B., and Sprott, G. D. (1994). Ribose biosynthesis in methanogenic bacteria. *Arch. Microbiol.*, **161**, 481–488.

Costilow, R. and Laycock, L. (1971). Ornithine Cyclase (Deaminating): Purification of a protein that converts ornithine to proline and definition of the optimal assay conditions. *J. Biol. Chem.*, **246**, 6655–6660.

Covert, M. and Palsson, B. (2003). Constraints-based models: Regulation of Gene Expression Reduces the Steady-state Solution Space. *J. Theor. Biol.*, **221**, 309–325.

Csete, M. and Doyle, J. (2004). Bow ties, metabolism and disease. *Trends in Biotech.*, **22**, 446–450.

Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998). Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, **23**, 324–328.

Dandekar, T., Schuster, S., Snel, B., Huynen, M., and Bork, P. (1999). Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.*, **343**, 115–124.

DasSarma, S., Fleischmann, E. M., and Rodriguez-Valera, F. (1995). Media for halophiles. In S. DasSarma, F. T. Robb, A. R. Place, K. R. Sowers, H. J. Schreier, and E. M. Fleischmann, editors, *Archaea: a laboratory manualhalophiles*, pages 225–230. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

De Rosa, M., Gambacorta, A., Nicolaus, B., Ross, H. N. M., and Grant, W. D. (1982). An asymmetric archaebacterial diether lipid from alkaliphilic halophiles. *J. Gen. Microbiol.*, **128**, 343–348.

De Rosa, M., Gambacorta, A., Nicolaus, B., and Grant, W. D. (1983). A $C_{25,25}$ diether core lipid from archaebacterial haloalkaliphiles. *J. Gen. Microbiol.*, **129**, 2333–2337.

Denner, E., McGenity, T., Busse, H., Grant, W., Wanner, G., and Stan-Lotter, H. (1994). *Halococcus salifodinae* sp. nov., an archaeal isolate from an Austrian salt mine. *Int. J. Syst. Bacteriol.*, **44**, 774–780.

Desmarais, D., Jablonski, P. E., Fedarko, N. S., and Roberts, M. F. (1997). 2-Sulfotrehalose, a novel osmolyte in haloalkaliphilic archaea. *J. Bacteriol.*, **179**, 3146–3153.

Devos, D. and Valencia, A. (2000). Practical Limits of Function Prediction. *Proteins*, **41**, 98–107.

Devos, D. and Valencia, A. (2001). Intrinsic errors in genome annotation. *Trends Genet.*, **17**, 429–431.

Dobson, S. and Franzmann, P. (1996). Unification of the genera *Deleya*, *Halomonas*, *Halovibrio*, and the species *Paracoccus halodenitrificans* in the family *Halomonadaceae*. *Int. J. Syst. Bacteriol.*, **46**, 550–558.

Dobson, S., James, S., Franzmann, P., and McMeekin, T. (1990). Emended description of *Halomonas halmophila* (NCBM 1971[T]). *Int. J. Syst. Bacteriol.*, **40**, 462–463.

D'Souza, S. E. and Altekar, W. (1998). A Class II fructose-1,6-bisphosphate aldolase from a halophilic archaebacterium *Haloferax mediterranei*. *J. Gen. Appl. Microbiol.*, **44**, 235–241.

Dundas, I. E. and Halvorson, H. O. (1966). Arginine Metabolism in *Halobacterium Salinarum*, an Obligately Halophilic Bacterium.

J. Bacteriol., **91**, 113–119.

Dundas, I. E., Sarinivasan, V. R., and Halvorson, H. O. (1963). A chemically defined medium for *Halobacterium salinarum* strain I. *Can. J. Microbiol.*, **9**, 619–624.

Dyall-Smith, M. K. M. (1995). Taxonomy of the family *Halobacteriaceae* and the description of two new genera *Halorubrobacterium* and *Natrialba*. *J. Gen. Appl. Microbiol.*, **41**, 333–350.

Edwards, J. and Palsson, B. (1999). System properties of the *Haemophilus influezae* Rd metabolic genotype. *J. Biol. Chem.*, **274**, 17410–17416.

Edwards, J., Ibarra, R., and Palsson, B. (2000). *In silico* predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, **19**, 125–130.

Eisen, J. (1999). Going to extremes. Archaea: bridging the gap between bacteria and eukarya (A Keystone Symposia), Taos, NM, USA, 9-14 January 1999. *Trends Genet.*, **15**, 218–9.

Ekiel, I., Sprott, D., and Smith, I. (1986). Mevalonic Acid Is Partially Synthesized from Amino Acids in *Halobacterium cutirubrum*: a ^{13}C Nuclear Magnetic Resonance Study. *J. Bacteriol.*, **166**, 559–564.

Engelhard, M., Hess, B., Metz, G., Kreutz, W., Siebert, F., Soppa, J., and Oesterhelt, D. (1990). High Resolution 13C-Solid State NMR of Bacteriorhodopsin: Assignment of Specific Aspartic Acids and Structural Implications of Single Site Mutations. *Eur. Biophys. J.*, **18**, 17–24.

Enright, A., Ilipoulos, I., Kyrpides, N., and Ouzounis, C. (1999). Protein interaction maps for omplete genomes based on gene fusion events. *Nature*, **402**, 86–90.

Essen, L. O., Siegert, R., Lehmann, W. D., and Oesterhelt, D. (1998). Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. *Proc. Natl. Acad. Sci. USA*, **95**, 11673–11678.

Falb, M. (2005). *Computational Genome and Pathway Analysis of Halophilic Archaea*. PhD thesis, Ludwig-Maximilians-Universität München, Fakultät für Chemie und Pharmazie.

Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J., and Oesterhelt, D. (2005). Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res.*, **15**, 1336–1343.

Falb, M., Müller, K., Königsmaier, L., Oberwinkler, T., Horn, P., von Gronau, S., Gonzalez, O., Pfeiffer, F., Bornberg-Bauer, E., and Oesterhelt, D. (2008). Metabolism of halophilic archaea. *Extremophiles*, **12**, 177–196.

Feist, A. M., Scholten, J. C. M., Palsson, B. O., Brockman, F. J., and Ideker, T. (2006). Modeling methanogenesis with a genome scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol. Syst. Biol.* doi:10.1038/msb4100046.

Fendrich, C. (1988). *Halovibrio variabilis* gen. nov. sp. nov., *Pseudomonas halophila* sp. nov. and a new halophilic aerobic coccoid eubacterium from Great Salt Lake, Utah, USA. *Syst. Appl. Microbiol.*, **11**, 36–43.

Fischer, R. S., Bonner, C. A., Boon, D. R., and Jensen, R. A. (1993). Clues from a halophilic methanogen about aromatic amino acid biosynthesis in archaebacteria. *Arch. Microbiol.*, **160**, 440–446.

Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K. B., Bairoch, A., Schomburg, D., Tipton, K. F., and Apweiler, R. (2004). IntEnz, the integrated relational enzyme database. *Nucleic Acids Res.*, **32**, D434–D437.

Galinski, E. A. (1995). Osmoadaptation in bacteria. *Adv. Microb. Physiol.*, **37**, 272–328.

Galinski, E. A., Pfeiffer, H. P., and Trüper, H. G. (1985). 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus *Ectothiorhodospira*. *Eur. J. Biochem.*, **149**, 135–139.

Gavrieli, I., Beyth, M., and Yechieli, Y. (1999). The Dead Sea - A terminal lake in the Dead Sea rift: a short overview. In A. Oren, editor, *Microbiology and biogeochemistry of hypersaline environments*, pages 121–127. CRC Press, Boca Raton.

Ghosh, M. and Sonawat, H. M. (1998). Kreb's TCA cycle in *Halobacterium salinarum* investigated by C-13 nuclear magnetic resonance spectroscopy. *Extremophiles*, **2**, 427–433.

Ginzburg, M., Sachs, L., and Ginzburg, B. (1970). Ion Metabolism in a *Halobacterium*. I. Influence of age culture on intracellular concentrations. *J. Gen. Physiol.*, **55**, 187–207.

Gochnauer, M. B. and Kushner, D. J. (1969). Growth and nutrition of extremely halophilic bacteria. *Canad. J. Microbiol.*, **15**, 1157–1165.

Gonzalez, O., Küper, C., Jung, K., Naval, P., and Mendoza, E. (2007). Parameter estimation using Simulated Annealing for S-system models of biochemical networks. *Bioinformatics*, **23**, 480–486.

Gonzalez, O., Gronau, S., Falb, M., Pfeiffer, F., Mendoza, E., Zimmer, R., and Oesterhelt, D. (2008). Reconstruction, Modeling & Analysis of *Halobacterium salinarum* R-1 Metabolism. *Molecular Biosystems*, **4**, 148–159.

Grant, W. D. and Norton, C. F. (1998). Survival of halobacteria within fluid inclusions in salt crystals. *J. Gen. Mirocbiol.*, **134**, 1365–1373.

Grant, W. D., Gemmel, R. T., and McGenity, T. J. (1998). Halobacteria: the evidence for longevity. *Extremophiles*, **2**, 279–287.

Graupner, M., Xu, H., and White, R. (2002). New Class of IMP Cyclohydrolases in *Methanococcus jannaschii*. *J. Bacteriol.*, **184**, 1471–1473.

Green, M. L. and Karp, P. D. (2005). Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. *Nucleic Acids Res.*, **33**, 40354039.

Grey, V. L. and Fitt, P. S. (1976). An improved synthetic growth medium for *Halobacterium cutirubrum*. *Can. J. Microbiol.*, **22**, 440–442.

Grigorieff, N., Ceska, T., Downing, K., Baldwin, J., and Henderson, R. (1996). Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.*, **259**, 393–421.

Grochowski, L. L., Xu, H., and White, R. H. (2005). Ribose-5-Phosphate Biosynthesis in Methanocaldococcus jannaschii Occurs in the Absence of a Pentose-Phosphate Pathway. *J. Bacteriol.*, **187**, 7382–7389.

Grochowski, L. L., Xu, H., and White, R. H. (2006). *Methanocaldococcus jannaschii* Uses a Modified Mevalonate Pathway for Biosynthesis of Isopentenyl Diphosphate. *J. Bacteriol.*, **188**, 31923198.

Halladay, J. T., Jones, J. G., Lin, F., MacDonald, A. B., and DasSarma, S. (1993). The Rightward Gas Vesicle Operon in *Halobacterium* Plasmid pNRC100: Identification of the *gvpA* and *gvpC* Gene Products by Use of Antibody Probes and Genetic Analysis of the Region Downstream of *gvpC*. *J. Bacteriol.*, **175**, 684–692.

Hampp, N. and Neebe, M. (2006). Bacteriorhodopsin-based Multi-level Optical Security Features. In R. van Renesse, editor, *Optical Security and Counterfeit Deterrence Techniques VI*, pages 6075 60750M, 1–9. SPIE, Bellingham, USA.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J. M., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S., Fisk, D. G., Hirschman, J. E., Hong, E. L., Nash, R. S., Sethuraman, A., Theesfeld, C. L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S. Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E. M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., and White, R. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–61.

Hartmann, R., Sickinger, H., and Oesterhelt, D. (1977). Quantitative Aspects of Energy Conversion in Halobacteria. *Febs Letters*, **82**, 1–6.

Hartmann, R., Sickinger, H., and Oesterhelt, D. (1980). Anaerobic growth of halobacteria. *Proc. Natl Acad. Sci. USA*, **77**, 3821–3825.

Haupts, U., Tittor, J., and Oesterhelt, D. (1999). Closing in on Bacteriorhodopsin: Progress in Understanding the Molecule. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 367–399.

Heinrich, R. and Rapoport, T. A. (1974). A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.*, **42**, 89–95.

Higuchi, S., Kawashima, T., and Suzuki, M. (1999). Comparison of pathways for amino acid biosynthesis in archaebacteria using their genomic DNA sequences. *Proc. Japan Acad.*, **75**, 241–245.

Hildebrand, E. and Dencher, N. (1980). Two photosystems controlling behavioral responses of *Halobacterium halobium*. *Proc. Natl Acad. Sci. USA*, **77**, 3821–3825.

Hoff, W. D., Jung, K., and Spudich, J. L. (1997). Molecular Mechanism of Photosignaling by Archaeal Sensory Rhodopsins. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 223–258.

Hou, S., Larsen, R. W., Boudko, D., Riley, C. W., Karatan, E., Zimmer, M., Ordal, G. W., and Alam, M. (2000). Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature*, **203**, 540–544.

Iliopoulos, I., Tsoka, S., Andrade, M. A., Janssen, P., Audit, B., Tramontano, A., Valencia, A., Leroy, C., Sander, C., and Ouzounis, C. A. (2001). Genome sequences and great expectations. *Genome Biol.* 2(1):interactions0001.10001.3.

Imhoff, J. F. and Rodriguez-Valera, F. (1984). Betaine is the main compatible solute of halophilic eubacteria. *J. Bacteriol.*, **160**, 478–479.

Irvine, D. H. and Savageau, M. A. (1990). Efficient solution of nonlinear ordinary differential equations expressed in S-system canonical form. *SIAM J. Numer. Anal.*, **27**, 704–735.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A. (2000). The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

Juez, F. R.-V. G. and Kushner, D. J. (1983). *Halobacterium mediterranei* spec. nov., a new carbohydrate-utilizing extreme halophile. *Syst. Appl. Microbiol.*, **4**, 369–381.

Kacser, H. and Burns, J. A. (1973). The control of flux. *Symp. Soc. Exp. Biol.*, **27**, 65104.

Kakinuma, K., Yamagishi, Y., Fujimoto, Y., Ikekawa, N., and Oshima, T. (1988). Stereochemistry of the biosynthesis of sn-2,3-O-diphytanyl glycerol, membrane lipid of archaebacteria *Halobacterium halobium*. *J. Am. Chem. Soc.*, **110**, 48614863.

Kamekura, M. (1993). Lipids of extreme halophiles. In R. H. Vreeland and L. I. Hochstein, editors, *The biology of halophilic bacteria*. CRC Press, Boca Raton.

Kamekura, M. (1998). Diversity of extremely halophilic bacteria. *Extremophiles*, **2**, 289–295.

Kamekura, M. and Kates, M. (1988). Lipids of halophilic ahrchaebacteria. In F. Rodriguez-Valera, editor, *Halophilic bacteria*. CRC Press, Boca Raton.

Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Res.*, **30**, 42–46.

Kates, M., Wassef, M. K., and Kushner, D. J. (1968). Radioisotopic studies on the biosynthesis of the glycerol diether lipids of *Halobacterium cutirubrum*. *Can. J. Biochem.*, **46**, 971–977.

Kato, N., Yurimoto, H., and Thauer, R. K. (2006). The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Biosci. Biotechnol. Biochem.*, **70**, 10–21.

Kelly, M., Norgard, S., and Liaaen-Jensen, S. (1970). Bacterial carotenoids. XXXI. C_{50} carotenoids. 5. Carotenoids of *Halobacterium salinarium*, especially bacterioruberin. *Acta Chem. Scand.*, **24**, 2169–2182.

Kerscher, K. and Oesterhelt, D. (1977). Ferredoxin is the coenzyme of α-ketoacid oxidoreductases in *Halobacterium halobium*. *FEBS Lett.*, **83**, 197–201.

Kessel, M., Wildhaber, I., Cohen, S., and Baumeister, W. (1988). Three-dimensional structure of the regular surface glycoprotein layer of *Halobacterium volcanii* from the Dead Sea. *EMBO J.*, **7**, 1549–1554.

Kevbrina, M. V., Zvyagintseva, I. S., and Plakunov, V. K. (1989). The uptake of [14] acetate in *Natrococcus occultus*. *Mikrobiologiya*, **58**, 892–896.

Kim, E. and Fitt, P. (1977). Partial Purification and Properties of *Halobacterium cutirubrum* L-Alanine Dehydrogenase. *Biochem. J.*, **161**, 313–320.

Kimura, S., Ide, K., Kashihara, A., Kano, M., Hatakeyama, M., Masui, R., Nakagawa, N., Yokoyama, S., Kuramitsu, S., and Konagaya, A. (2005). Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm. *Bioinformatics*, **21**, 1154–1163.

Klein, C., Garcia-Rizo, C., Bisle, B., Scheffer, B., Zischka, H., Pfeiffer, F., Siedler, F., and Oesterhelt, D. (2005). The membrane proteome of *Halobacterium salinarum*. *Proteomics*, **5**, 180–197.

Koch, M. (2005). *Investigations on halobacterial transducers with respect to membrane potential sensing and adaptive methylation*. PhD thesis, Ludwig-Maximilians-Universität München, Fakultät für Chemie und Pharmazie.

Kolbe, M., Besir, H., Essen, L., and Oesterhelt, D. (2000). Structure of the Light-Driven Chloride Pump Halorhodopsin at 1.8 Å Resolution. *Science*, **288**, 1390–1396.

Königsmaier, L. (2006). *Analysen des Fettsäuremetabolismus in halophilen Archaea*. Master thesis, University of Salzburg.

Konstantinidis, K., Tebbe, A., Klein, C., Scheffer, B., Aivaliotis, M., Bisle, B., Falb, M., Pfeiffer, F., Siedler, F., and Oesterhelt, D. (2007). Genome-wide proteomics of Natronomonas pharaonis. *J Proteome Res.*, **6**, 185–193.

Krah, M., Marwan, W., Vermeglio, A., and Oesterhelt, D. (1994). Phototaxis of *Halobacterium salinarium* requires a signalling complex of sensory rhodopsin I and its methyl-accepting transducer HtrI. *EMBO J*, **13**, 2150–2155.

Kushawa, S. C., Kramer, J. K. G., and Kates, M. (1975). Isolation and characterization of C_{50} carotenoid pigments and other polar isoprenoids from *Halobacterium cutirubrum*. *Biochim. Biophys. Acta*, **398**, 303–313.

Lai, M. C., Sowers, K. R., Robertson, D. E., Roberts, M. F., and Gunsalus, R. P. (1991). Distribution of compatible solutes in the halophilic methanogenic archaebacteria. *J. Bacteriol.*, **173**, 5352–5358.

Lanyi, J. K., Renthal, R., and MacDonald, R. E. (1976). Light-induced glutamate transport in *Halobacterium salinarum* cell vesicles. II. Evidence that the driving force is a light-dependent sodium gradient. *Biochem.*, **15**, 1603–1610.

Lechner, J. and Sumper, M. (1987). The Primary Structure of a Procaryotic Glycoprotein. *J. Biol. Chem.*, **262**, 9724–9729.

Linke, B., McHardy, A., Neuweger, H., Krause, L., and Meyer, F. (2006). REGANOR: A Gene Prediction Server for Prokaryotic Genomes and a Database of High Quality Gene Predictions for Prokaryotes. *Appl. Bioinformatics*, **5**, 193–198.

Liu, P. and Wang, F. (2008). Inference of biochemical network models in S-system using multiobjective optimization approach. *Bioinformatics*, **24**, 1085–1092.

Lobyreva, L. B. and Plakunov, V. K. (1987). Kinetic characteristics of transport of aromatic amino acids by *Halobacterium salinarum* cells. *Microbiol.*, **56**, 136–140.

Lobyreva, L. B., Kokoeva, M. V., and Plakunov, V. K. (1994). Physiological role of tyrosine transport systems in *Halobacterium salinarium*. *Arch. Microbiol.*, **162**, 126–130.

Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P., and Lanyi, J. K. (1999). Structure of Bacteriorhodopsin at 1.55 Å Resolution. *J. Mol. Biol.*, **291**, 899–911.

Ma, H. and Zeng, A. (2003). Reconstruction of metabolic networks from genome data and analyis of their global structure for various organisms. *Bioinf.*, **19**, 270–277.

MacDonald, R. E. and Lanyi, J. K. (1975). Light-induced transport in *Halobacterium halobium* envelope vesicles: a chemiosmotic system. *Biochem.*, **14**, 2882–2889.

MacDonald, R. E., Greene, R. V., and Lanyi, J. K. (1977). Light-activated amino acid transport systems in Halobacterium halobium envelope vesicles: role of chemical and electrical gradients. *Biochemistry*, **16**, 3227–3235.

Madden, T. L., Tatusov, R. L., and Zhang, J. (1996). Applications of network BLAST server. *Meth. Enzymol.*, **266**, 131–141.

Madern, D., Ebel, C., and Zaccai, G. (2000). Halophilic adaptation of enzymes. *Extremophiles*, **4**, 91–98.

Mahadevan, R. and Schilling, C. (2003). The effects of alternate optimal solutions in constraints-based genome-scale metabolic models. *Met. Eng.*, **5**, 264–276.

Mahadevan, R., Edwards, J., and Doyle, F. (2002). Dynamic Flux Balance Analysis of Diauxic Growth in *E. coli. Biophys. J.*, **83**, 1331–1340.

Ma'or, Z., Henis, Y., Alon, Y., Orlov, E., Sorensen, K., and Oren, A. (2006). Antimicrobial properties of Dead Sea black mineral mud. *International Journal of Dermatology*, **45**, 504–511.

Marcotte, E., Pellegrini, M., Ng, H., Rice, W., Yeates, T., and Eisenberg, D. (1999). Detecting protein funciton and protein-protein interactions from genome sequences. *Science*, **285**, 751–753.

Marwan, W., Alam, M., and Oesterhelt, D. (1991). Rotation and Switching of the flagellar motor assembly in *Halobacterium halobium. J. Bacteriol.*, **173**, 1971–1977.

Matheson, A. T., Sprott, G. D., McDonald, I. J., and Tessier, H. (1976). Some properties of an unidentified halophile: growth characteristics, internal salt concentrations, and morphology. *Can. J. Microbiol.*, **22**, 780–786.

Matsubara, T., Iida-Tanaka, N., Kamekura, M., Moldoveanu, N., Ishizuka, I., Onishi, H., Hayashi, A., and Kates, M. (1994). Polar lipids of non-alkaliphilic extremely halophilic archaebacterium strain 172: a novel bis-sulfated glycolipid. *Biochim. Biophys. Acta*, **1214**, 97–108.

Mescher, M. and Strominger, J. (1976). Purification and Characterization of a Prokaryotic Glycoprotein from the Cell Envelope of *Halobacterium salinarium. J. Biol. Chem.*, **251**, 2005–2014.

Miyauchi, S., Komatsubara, M., and Kamo, K. (1992). In archaebacteria, there is a doxorubicin efflux pump similar to mammalian P-glycoprotein. *Biochim. Biophys. Acta*, **1110**, 144–150.

Miyauchi, S., Tanabu, S., Abe, A., Okumura, R., and Kamo, N. (1997). Culture in the presence of sugars increases activity of multi-drug efflux transporter on *Haloferax vaolcanii. Microb. Drug Resist.*, **3**, 359–363.

Montalvo-Rodriguez, R., Vreeland, R. H., Oren, A., Kessel, M., Betancourt, C., and Lopez-Garriga, J. (1998). *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int. J. Syst. Bacteriol.*, **48**, 1305–1312.

Morita, M., Yamaguchi, N., Eguchi, T., and Kakinuma, K. (1998). Structural diversity of the membrane core lipids of extreme halophiles. *Biosci. Biotechnol. Biochem*, **62**, 596–598.

Mullakhanbhai, M. F. and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.*, **104**, 207–214.

Müller, J. and DasSarma, S. (2005). Genomic Analysis of Anaerobic Respiration in the Archaeon Halobacterium sp. Strain NRC-1: Dimethyl Sulfoxide and Trimethylamine N-Oxide as Terminal Electron Acceptors. *J. Bacteriol.*, **187**, 1659–1667.

Mwatha, W. E. and Grant, W. D. (1993). *Natronobacterium vacuolata*, a haloalkaliphilic archaeon isolated from Lake Magadi, Kenya. *Int. J. Syst. Bacteriol.*, **43**, 401–404.

Ng, W. V., Kennedy, S., Mahalras, G., Berquist, B., Pan, M., Shukla, H., Lasky, S., Baliga, N., Thorsson, V., Sbronga, J., Swartzell, S., Weir, D., Hall, J., Dahl, T., Welti, R., Goo, Y., Leithauser, B., Keller, K., Cruz, R., Danson, M., Hough, D., Maddocks, D., Jablonski, P., Krebs, M., Angevine, C., Dale, H., Isenbarger, T., Peck, R., Pohlschroder, M., Spudich, J., Jung, K., Alam, M., Freitas, T., Hou, S., Daniels, C., Dennis, P., Omer, A., Ebhardt, H., Lowe, T., Liang, P., Riley, M., Hood, L., and DasSarma, S. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. (USA)*, **97**, 12176–12181.

Ni, T. and Savageau, M. (1995). Application of Biochemical Systems Theory to Metabolism in Human Red Blood Cells. *J. Biol. Chem.*, **271**, 7927–7941.

Nishihara, M., Yamazaki, T., Oshima, T., and Koga, Y. (1999). *sn*-Glycerol-1-Phosphate-Forming Activities in Archaea: Separation of Archaeal Phospholipid Biosynthesis and Glycerol Catabolism by Glycerophosphate Enantiomers. *J. Bacteriol.*, **181**, 1330–1333.

Norton, C., McGenity, T., and Grant, W. (1993). Archaeal halophiles from two British salt mines. *J. Gen. Microbiol.*, **139**, 1077–1081.

O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., and Apweiler, R. (2002). High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief. Bioinform.*, **3**, 275–284.

Oesterhelt, D. (1988). The Structure and Mechanism of the Family of Retinal Proteins from Halophilic Archaea. *Curr. Op. Struct. Biol.*, **8**, 489–500.

Oesterhelt, D. and Krippahl, G. (1973). Light Inhibition of Respiration in *Halobacterium halobium. FEBS Letters*, **36**, 72–76.

Onishi, H., McCance, M. E., and Gibbons, N. E. (1965). A synthetic medium for extremely halophilic bacteria. *Can. J. Microbiol.*, **11**, 365–373.

Oren, A. (1983a). *Halobacterium sodomense* sp. nov., a Dead Sea halobacterium with extremely high magnesium requirement and tolerance. *Int. J. Syst. Bacteriol.*, **33**, 381–386.

Oren, A. (1983b). Population dynamics of halobacteria in the Dead Sea water column. *Limnol. Oceanogr.*, **28**, 1094–1103.

Oren, A. (1999). Life at high salt concentrations. In M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, and E. Stackebrandt, editors, *The prokaryotes. A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications*. Springer-Verlag, New York, N.Y., 3rd edition.

Oren, A. (2002). *Halophilic Microorganisms and their Environments*, volume 5 of *Cellular Origin and Life in Extreme Habitats*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Oren, A. (2006). The Order Halobacteriales. *Prokaryotes*, **3**, 113–164.

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems*, **4**, 2.

Oren, A. and Gurevich, P. (1995a). Dynamics of a bloom of halophilic archaea in the Dead Sea. *Hydrobiologia*, **315**, 149–158.

Oren, A. and Gurevich, P. (1995b). Isocitrate lyase activity in halophilic archaea. *FEMS Microbiol. Lett.*, **130**, 91–95.

Oren, A. and Gurevich, P. (1995c). Occurrence of the methylglyoxal bypass in halophilic Archaea. *FEMS Microbiol. Lett.*, **125**, 83–88.

Oren, A. and Rodriguez-Valera, F. (2001). The contribution of halophilic Bacteria to the red coloration of saltern crystallizer ponds1. *FEMS Microbiology Ecology*, **36**, 123–130.

Oren, A. and Shilo, M. (1982). Population dynamics of *Dunaliella parva* in the Dead Sea. *Limnol. Oceanogr.*, **27**, 201–211.

Oren, A. and Trüpper, H. G. (1990). Anaerobic growth of halophilic archaeobacteria by reduction of dimethylsulfoxide and trimethylamine N-oxide. *FEMS Microbiol. Lett.*, **70**, 33–36.

Oren, A., Weisburg, W. G., Kessel, M., and Woese, C. R. (1984). *Halobacteroides halobius* gen. nov., sp. nov., a moderately halophilic anaerobic bacterium from the bottom sediments of the Dead Sea. *Syst. Appl. Microbiol.*, **5**, 58–69.

Oren, A., Pohla, H., and Stackebrandt, E. (1987). Transfer of *Clostridium lortetii* to a new genus *Sporohalobacter* gen.nov. as *Sporohalobacter lortetii* comb. nov., and description of *Sporohalobacter marismortui* sp.nov. *Syst. Appl. Microbiol.*, **9**, 239–246.

Oren, A., Stambler, N., and Dubinsky, Z. (1992). On the red colorization of saltern crystallizer ponds. *Int. J. Salt Lake Res.*, **1**, 77–89.

Oren, A., , and Dubinsky, Z. (1994). On the red colorization of saltern crystallizer ponds. II. Additional evidence for the contribution of halobacterial pigments. *Int. J. Salt Lake Res.*, **3**, 9–13.

Oren, A., Gurevich, P., Gemmel, R. T., and Teske, A. (1995). *Halobacterium gomorrense* gen. nov., sp. nov., a novel extremely halophilic archaeon from the Dead Sea. *Int. J. Syst. Bacteriol.*, **45**, 747–754.

Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T., and Sakai, Y. (2006). The Ribulose Monophosphate Pathway Substitutes for the Missing Pentose Phosphate Pathway in the Archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.*, **188**, 4689–4704.

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G., and Maltsev, N. (1999). The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. (USA)*, **96**, 2896–2901.

Palsson, B. O. (2004). Two-dimensional annotation of genomes. *Nature Biotechnol.*, **22**, 1218–1219.

Pan, N. B. M., Goo, Y., Yi, E., Goodlet, D., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W., and Hood, L. (2002). Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Proc. Natl Acad. Sci. USA*, **99**, 14913–14918.

Paterek, J. and Smith, P. (1985). Isolation and Characterization of a Halophilic Methanogen from Great Salt Lake. *Appl. Environ. Microbiol.*, **50**, 877–881.

Paterek, J. and Smith, P. (1988). *Methanohalophilus mahii* gen. nov., sp. nov., a methyltrophic halophilic methanogen. *Int. J. Syst.Bacteriol.*, **38**, 122–123.

Paul, G., Lottspeich, F., and Wieland, F. (1986). Asparaginyl-N-acetylgalactosamine: Linkage unit of halobacterial gycosaminoglycan. *J. Biol. Chem.*, **261**, 1020–1024.

Pelligrini, M., Marcotte, E., Thompson, M., Eisenberg, D., and Yeates, T. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci. (USA)*, **96**, 4285–4288.

Pfeiffer, F., Schuster, S. C., Broicher, A., Falb, M., Palm, P., Rodewald, K., Ruepp, A., Soppa, J., Tittor, J., and Oesterhelt, D.

(2008a). Evolution in the laboratory: The genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics*, **91**, 335–346.

Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejia, J., Rampp, M., and Oesterhelt, D. (2008b). Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.*, **DOI 10.1007/s00203-008-0389-z**.

Plugge, C. M., Leeuwen, J. M., Hummelen, T., Balk, M., and Stams, A. J. M. (2001). Elucidation of the pathways of catabolic glutamate conversion in three thermophilic anaerobic bacteria. *Archives of Microbiol.*, **176**, 29–36.

Post, F. (1977a). The microbial ecology of the Great Salt Lake. *Microb. Ecol.*, **3**, 143–165.

Post, F. (1977b). The microbial ecology of the Great Salt Lake north arm. In D. Greer, editor, *Deseric terminal lakes. Proceedings of the international conference on deseric terminal lakes, Weber State College, Ogden, Utah*. Utah Water Research Laboratory, Logan.

Pugh, E. L. and Kates, M. (1994). Acylation of proteins of the archaebacteria Halobacterium cutirubrum and Methanobacterium thermoautotrophicum. *Biochim. et Biophys. Acta*, **1196**, 38–44.

Pugh, E. L., Wassef, M. K., and Kates, M. (1971). Inhibition of fatty acid synthetase in *Halobacterium cutirubrum* and *Escherichia coli* by high salt concentrations. *Can. J. Biochem.*, **49**, 953–958.

Rafaeli-Eshkol, D. (1968). Studies on halotolerance in a moderately halophilic bacterium. Effect of growth conditions on salt resistance of the respiratory system. *Biochem. J.*, **109**, 679–685.

Ramakrishna, R., Edwards, J., McCulloch, A., and Palsson, B. (2001). Flux-Balance analysis of mitochondrial energy meabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, **280**, R695–R704.

Rawal, N., Kelkar, S. M., and Altekar, W. (1988). Alternative routes of carbohydrate metabolism in halophilic archaebacteria. *Indian J. Biochem. Biophys.*, **25**, 674–686.

Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.*, **4**, R54.

Reese, M. G., Kulp, D., Tammana, H., and Haussler, D. (2000). Genie - gene finding in *Drosophila melanogaster*. *Genome Res.*, **10**, 529–538.

Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997). In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: II. Mathematical Model. *Biotechnol. Bioeng.*, **55**, 592608.

Rodriguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1980). Isolation of extremely halophilic bacteria able to grow in defined inorganic media with single carbon sources. *J. Gen. Microbiol.*, **119**, 535–538.

Roth, J. R., Lawrence, J. G., and Bobik, T. A. (1996). COBALAMIN (COENZYME B12): Synthesis and Biological Significance. *Annu. Rev. Microbiol.*, **50**, 13781.

Saier, Jr., M. H. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.*, **64**, 354–411.

Salzberg, S. L., Delcher, A. L., Kasif, S., and White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Research*, **26**, 544–548.

Sartorius-Neef, S. and Pfeifer, F. (2004). In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon Halobacterium salinarum. *Mol. Microbiol.*, **51**, 579–588.

Sauer, U. (2004). High-throughput phenomics: experimental methods for mapping fluxomes. *Curr. Opin. Biotechnol.*, **15**, 5863.

Sauer, U. (2006). Metabolic networks in motion: ^{13}C-based flux analysis. *Mol. Syst. Biol.*, **2**, 62.

Savageau, M. (1969a). Biochemical Systems Analysis, I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.*, **25**, 365–369.

Savageau, M. (1969b). Biochemical Systems Analysis, II. The steady-state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.*, **25**, 370–379.

Savageau, M. (1970). Biochemical Systems Analysis, III. Dynamic solutions using a power-law approximation. *J. Theor. Biol.*, **26**, 215–226.

Scharf, B. and Engelhard, M. (1993). Halocyanin, an Archaebacterial Blue Copper Protein (Type I) from *Natronobacterium pharaonis*. *Biochemistry*, **32**, 12894–12900.

Scharf, B., Wittenberg, R., and Engelhard, M. (1997). Electron Transfer Proteins from the Haloalkaliphilic Archaeon *Natronobacterium pharaonis*: Possible Components of the Respiratory Chain Include Cytochrome *bc* and a Terminal Oxidase Cytochrome ba_3. *Biochemistry*, **36**, 4471–4479.

Schilling, C. H. and Palsson, B. O. (2000). Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.*, **203**, 249–283.

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S., and Palsson, B. O. (2002). Genome-scale metabolic model

of *Helicobacter pylori* 26695. *J. Bacteriol.*, **184**, 4582–4593.

Schobert, B. and Lanyi, J. K. (1982). Halorhodopsin Is a Light-driven Chloride Pump. *J. Biol. Chem.*, **257**, 10306–10313.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., and Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.*, **32**, D431–D433.

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli. Mol. Syst. Biol.*, **3**, 119.

Severina, L. O., Pimenov, N. V., and Plakunov, V. K. (1991). Glucose transport into the extremely halophilic archaebacteria. *Arch. Microbiol.*, **155**, 131–136.

Shand, R. F. and Perez, A. M. (1999). Haloarchaeal growth physiology. In J. Seckbach, editor, *Enigmatic microorganisms and life in extreme environments*, pages 414–424. Kluwer Academic Publishers, Dordrecht.

Shukla, H. D. and DasSarma, S. (2004). Complexity of Gas Vesicle Biogenesis in *Halobacterium* sp. Strain NRC-1: Identification of Five New Proteins. *J. Bacteriol.*, **186**, 3182–3186.

Silva, Z., Borges, N., Martins, L. O., Wait, R., da Costa, M. S., and Santos, H. (1999). Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by *Rhodothermus marinus* and textitRhodothermus obamensis. *Extremophiles*, **3**, 163–172.

Simon, R. D. (1981). Morphology and protein composition of gas vesicles from wild type and gas vacuole defective strains of *Halobacterium salinarum* strain 5. *J. Gen. Microbiol.*, **125**, 103–111.

Snel, B., Bork, P., and Huynen, M. (2000). Genome evolution: Gene fusion versus gene fission. *Trends Genet.*, **16**, 9–11.

Soliman, G. and Truper, H. (1982). *Halobacterium pharaonis* sp. nov., a new, extremely haloalkaliphilic archaebacterium with low magnesium requirement. *Zentralblatt Fur Bakteriologie Mikrobiologie Und Hygiene I Abteilung Originale C*, **3**, 318–329.

Solovyev, V., Salamov, A., and Lawrence, C. (1994). Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.*, **22**, 51565163.

Sonawat, H. M., Srivastava, S., Swaminathan, S., and Govil, G. (1990). Glycolysis and Entner-Doudoroff pathways in Halobacterium halobium: some new observations based on 13C NMR spectroscopy. *Biochem. Biophys. Res. Commun.*, **173**, 358–362.

Sowers, K. R., amd D Noll, D. E. R., Gunsalus, R. P., and Roberts, M. F. (1990). Nϵ-acetyl-β-lysine: an osmolyte synthesized by methanogenic bacteria. *Proc. Natl. Acad. Sci. USA*, **87**, 9083–9087.

Sparnins, V. L. and Chapman, P. J. (1976). Catabolism of L-Tyrosine by the Homoprotocatechuate Pathway in Gram-Positive Bacteria. *J. Bacteriol.*, **127**, 362–366.

Spudich, J. L. (1998). Variations on a molecular switch: transport and sensory signalling by archaeal rhodopsins. *Mol. Microbiol.*, **28**, 10511058.

Sreeramulu, K., Schmidt, C., Schäfer, G., and Anemüller, S. (1998). Studies of the Electron Transport Chain of the Euryarcheon *Halobacterium salinarum*: Indications for a Type II NADH Dehydrogenase and a Complex III Analog. *Journal of Bioenergetics and Biomembranes*, **30**, 443–453.

Stein, L. (2001). Genome Annotation: From Sequence to Biology. *Nature Rev. Genet.*, **2**, 493–503.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–3.

Storch, K., Rudolph, J., and Oesterhelt, D. (1999). Car: a cytoplasmic sensor responsible for arginine chemotaxis in the archaeon *Halobacterium salinarum*. *EMBO J.*, **18**, 1146–1158.

Stuer-Lauridsen, B. and Nygaard, P. (1998). Purine salvage in two halophilic archaea: characterization of salvage pathways and isolation of mutants resistant to purine analogs. *J. Bacteriol*, **180**, 457–463.

Sumper, M. (1987). Halobacterial glycoprotein biosynthesis. *Biochimica et Biophysica Acta*, **906**, 69–80.

Takano, J., Kaidoh, K., and Kamo, N. (1995). Fructose transport by *Haloferax volcanii*. *Can. J. Microbiol.*, **41**, 241–246.

Takashina, T., Hamamoto, T., Otozai, K., Grant, W. D., and Horikoshi, K. (1990). *Haloarcula japonica* sp. nov., a new triangular halophilic archaebacterium. *Syst. Appl. Microbiol.*, **13**, 177–181.

Tanaka, M., Mukohata, Y., and Yuasa, S. (2000). Differential transport properties of D-leucine and L-leucine in the archaeon, *Halobacterium salinarum*. *Can. J. Microbiol.*, **46**, 376382.

Tanaka, R. (2005). Scale-Rich Metabolic Networks. *Phys. Rev. Lett.*, **94**, 168101.

Tansil, B. (1984). *Bergey's manual of systematic bacteriology*. Baltimore: Williams & Wilkins.

Tatusov, R., Koonin, E., and Lipman, D. (1997). A Genomic Perspective on Protein Families. *Science*, **278**, 631–637.

Tatusov, R., Natale, D., Garkavstev, I., Tatusova, T., Shankavaram, U., Rao, B., Kiryutin, B., Galperin, M., Fedorova, N., and Koonin, E. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.*, **29**, 22–28.

Tawara, E. and Kamo, N. (1991). Glucose transport of *Haloferax volcanii* requires the Na$^+$-electrochemical potential gradient and inhibitors for the mammalian glucose transporter inhibit the transport. *Biochim. Biophys. Acta*, **1070**, 293–299.

Tebbe, A., Klein, C., Bisle, B., Siedler, F., Scheffer, B., Garcia-Rizo, C., Wolfertz, J., Hickmann, V., Pfeiffer, F., and Oesterhelt, D. (2005). Analysis of the cytosolic proteome of Halobacterium salinarum and its implication for genome annotation. *Proteomics*, **5**, 168–79.

Teusink, B., Passarge, J., Reijenga, C., Esgalhado, E., van der Weijden, C., Schepper, M., Walsh, M., Bakker, B., van Dam, K., Westerhoff, H., and Snoep, J. (2000). Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochem.*, **267**, 5313–5329.

Teusink, B., Wiersma, A., Molenaar, D., Francke, C., de Vos, W., Siezen, R., and Smid, E. (2006). Analysis of Growth of *Lactobacillus plantarum* WCSF1 on a Complex Medium Using a Genome-scale Metabolic Model. *J. Biol. Chem.*, **281**, 40041–40048.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M., and Reuss, M. (1997). In Vivo Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: I. Experimental Observations. *Biotechnol. Bioeng.*, **55**, 305316.

Tindall, B. J., Ross, H. N. M., and Grant, W. D. (1984). Natronobacterium gen. nov. and Natronococcus gen. nov., two new genera of haloalkaliphilic archaebacteria. *System. Appl. Microbiol.*, **5**, 41–57.

Tipton, K. F. and Boyce, S. (2000). Enzyme Classification and Nomenclature. In *Nature Encyclopedia of Life Sciences*. Macmillan Reference Co. UK. doi:10.1038/npg.els.0000710.

Torarinsson, E., Klenk, H. P., and Garrett, R. A. (2005). Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.*, **7**, 47–54.

Torres, N. and Voit, E. (2002). *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press, Cambridge, UK.

Tsai, C., Garcia, J., Patel, B., Cayol, J., Baresi, L., and Mah, R. (1995). *Haloanaerobium alcaliphilum* sp. nov., an anaerobic moderate halophile from the sediments of Great Salt Lake, Utah. *Int. J. Syst. Bacteriol.*, **45**, 301–307.

Tumbula, D. L., Teng, Q., Barkett, M. G., and Whitman, W. B. (1997). Ribose biosynthesis and evidence for an alternative first step in the common aromatic amino acid pathway in *Methanococcus maripaludis*. *J. Bacteriol.*, **179**, 6010–6013.

Uberacher, E. and Mural, R. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensorneural network approach. *Proc. Natl Acad. Sci. USA*, **88**, 1126111265.

Van Auken, O. and McNulty, I. (1973). The Effect of Environmental Factors on the Growth of a Halophilic Species of Algae. *Biol. Bull.*, **145**, 210–222.

Varma, A. and Palsson, B. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.*, **60**, 3724–3731.

Ventosa, A., Gutierrez, M. C., Garcia, M. T., and Ruiz-Berraquero, F. (1989). Classification of *Chromobacterium marismortui* in a new genus, *Chromohalobacter* gen. nov., as *Chromohalobacter marismortui* comb. nov. *Int. J. Syst. Bacteriol.*, **39**, 392–396.

Visser, D. and Heijnen, J. J. (2002). Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metab. Eng.*, **5**, 164176.

Voit, E. (2000). *Computational Analysis of Biochemical Systems*. Cambridge University Press, Cambridge, UK.

Voit, E. (2006). E Voit and J Almeida and S Marino and R Lall and G Goel and A Neves and H Santos. *IEE Proc.-Syst. Biol.*, **153**, 286–298.

Vreeland, R., Rosenzweig, W., and Powers, D. (2000). Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature*, **407**, 897–900.

Wagner, G., Hartmann, R., and Oesterhelt, D. (1978). Potassium Uniport and ATP Synthesis in *Halobacterium halobium*. *Eur. J. Biochem.*, **89**, 169–179.

Waino, M., Tindall, B., and Ingvorsen, K. (1999). *Gracilibacillus* gen. nov., with description of *Gracilibacillus halotolerans* gen. nov., sp. nov.: transfer of *Bacillus dipsosauri* to *Gracibacillus dipsosauri* comb. nov., and *Bacillus salexigens* to the genus *Salibacillus* gen. nov., as *Salibacillus salexigens* comb. nov. *Int. J. Syst. Evol. Microbiol.*, **49**, 821–831.

Walsbyi, A. E. (1980). A square bacterium. *Nature*, **283**, 69–71.

Wassef, M. K., Sarner, J., and Kates, M. (1970). Stereospecificity of the glycerol kinase and the glycerophosphate dehydrogenase in Halobacterium cutirubrum. *Can. J. Biochem.*, **48**, 69–73.

White, R. (2004). L-Aspartate semialdehyde and 6-deoxy-5-keto-hexose-1-phosphate are the precursors to the aromatic amino acids in *Methanocaldococcus jannaschii*. *Biochemistry*, **43**, 7618–7627.

Wieland, F., Heitzer, R., and Schaefer, W. (1983). Asparaginylglucose: Novel type of carbohydrate linkage. *Proc. Natl. Acad. Sci. USA*, **80**, 5470–5474.

Woese, C. and Fox, G. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.*

USA, **74**, 5088–5090.

Yao, N., Lei, M., Ren, L., Menke, N., Wang, Y., Fischer, T., and Hampp, N. (2005). Polarization multiplexed write-one-readmany optical data storage in bacteriorhodopsin films. *Opt. Lett.*, **30**, 3060–3062.

Yu, J. P., Ladapo, J., and Whitman, W. B. (1994). Pathway of glycogen metabolism in Methanococcus maripaludis. *J. Bateriol.*, **176**, 325–332.

Zaccai, G. and Eisenberg, H. (1991). A model for the stabilization of a halophilic protein. In G. di Prisco, editor, *Life under extreme conditions*, pages 125–137. Springer-Verlag, Berlin.

Zeikus, J. (1983). Metabolic communication between biodegradative populations in nature. In J. H. Slater, E. Whittenbury, and J. W. T. Wimpenny, editors, *Microbes in their natural environments. Society of General Microbiology Symposium 34*, pages 423–462. Cambridge University Press, Cambridge.

Zeikus, J. G., Hegge, P. W., Thompson, T. E., Phelps, T. J., and Langworthy, T. A. (1983). Isolation and description of *Haloanaerobium praevalens* gen. nov. and sp. nov., an obligately anaerobic halophile common to Great Salt Lake sediments. *Curr. Microbiol.*, **9**, 225–234.

Zhang, J. and Madden, T. L. (1997). PowerBLAST: A New Network BLAST Application for Interactive or Automated Sequence Analysis andAnnotation. *Genome Res.*, **7**, 649–656.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.*, **7**, 203–214.

Zoratti, M. and Lanyi, J. K. (1987). Phosphate transport in *Halobacterium halobium depends on cellular ATP levels*. *J. Bacteriol.*, **169**, 5755–5760.

Zvyagintseva, I. S., Gerasimenko, L. M., Kostrikina, N. A., Bulygina, E. S., and Zavarin, G. A. (1995). Interaction of the halobacteria and cyanobacteria in a halophilic cyanobacterial community. *Mikrobiologiya*, **64**, 252–258.

Acknowledgements

Finally, this dissertation would not have been possible without the extensive support of family and friends. I would like to express my sincere gratitude to my wife Criselda who has been very supportive from the start. I was able to concentrate on my work because she took care of the rest. I would also like to thank my parents who have been nothing but encouraging. And lastly, my appreciation to our friends here in Munich who have helped make living in a foreign land much more enjoyable.

Curriculum vitae

Personal Information

Name	Orland Gonzalez
Address	Troppauerstrasse 10, 80937 Munich
Birth	13.01.1981, in Manila Philippines

Education and Professional Experience

2005 -	PhD student:
	Ludwig-Maximilians-University Munich
	Teaching and Research Unit for Practical Informatics and Bioinformatics, chair: Prof. Ralf Zimmer
	Research fellow:
	Max Planck Insititute of Biochemistry
	Membrane Biochemistry, chair (Prof. Oesterhelt)
2004	Instructor (numerical methods)
	University of the Philippines
	Department of Computer Science
2002-2004	Lecturer (data structures)
	University of the Philippines
	Department of Computer Science
2001-2002	Junior Associate
	Azesu Systems Ltd.
2002-2004	University of the Philippines, Diliman
	MS Computer Science
	Thesis title: *Parameter Estimation Using Simulated Annealing for S-System Models of Biochemical Networks*
1997-2001	University of the Philippines, Diliman
	BS Computer Science

Journal Publications

- Gonzalez O, Küper C, Jung K, Naval P, Mendoza E (2007), Parameter estimation using Simulated Annealing for S-system models of biochemical networks, *Bioinformatics*, 23:480-486.

- Gonzalez O, Gronnau S, Falb M, Mendoza E, Zimmer R, Oesterhelt D (2008), Reconstruction, Modeling & Analysis of *H. salinarum* R-1 Metabolism, *Molecular BioSystems* 4:148-159.

- Falb M, Müller K, Königsmayer L, Oberwinkler T, Horn P, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D (2007), Metabolism of Halophilic Archaea, *Extremophiles* 12:177-196.

- Gonzalez O, Zimmer R (2008), Assigning functional linkages to proteins using phylogenetic profiles and continuous phenotypes, *Bioinformatics* 24:1257-1263.

- Gonzalez O, Gronnau S, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D (2008), Systems analysis of bioenergetics and growth of the extreme halophile *Halobacterium salinarum*, *PLOS Computational Biology*, accepted.

- Gonzalez O, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D (2008), Metabolic Reconstruction and Analysis of the Polyextremophile *Natronomonas pharaonis*, In preparation.