Cristiana Silva Faria

Mestre em Microbiologia Aplicada

Development of cell factories for the efficient production of mannosylglycerate, a thermolyte with great potential in biotechnology

Dissertação para obtenção do Grau de Doutor em
Bioengenharia – MIT Portugal

Orientador: Doutora Maria Helena Dias dos Santos
Professora Catedrática, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa

Co-orientador: Doutora Isabel Cristina de Almeida Pereira da Rocha
Professora Auxiliar, Departamento de Engenharia Biológica, Universidade do Minho

Arguentes Doutor Eugénio Manuel de Faria Campos Ferreira
Professor Catedrático da Escola de Engenharia, Universidade do Minho

Doutor Fernando José Santos Rodrigues
Professor Associado, com Agregação da Escola de Ciências da Saúde, Universidade do Minho.

Vogais Doutor Gabriel António Amaro Monteiro
Professor Associado do Instituto Superior Técnico, Universidade de Lisboa

Doutora Paula Maria Theriaga Mendes Bernardo Gonçalves
Professora Auxiliar da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

outubro de 2017

Development of cell factories for the efficient production of mannosylglycerate, a thermolyte with great potential in biotechnology

Agradecimentos

Esta tese de doutoramento foi possível graças ao apoio e motivação de diversas pessoas. A todas estas pessoas, seguem umas palavras de agradecimento por tornarem possível a realização desta importante etapa da minha vida.

Em primeiro lugar quero agradecer à minha orientadora, a Professora Helena Santos, por me ter acolhido no seu grupo, primeiramente numa bolsa de investigação, seguida de mestrado e posteriormente doutoramento. Muito obrigada por todos os ensinamentos, sejam profissionais ou, também não menos importantes, pessoais que me estimularam a querer ser o melhor de mim sem nunca abdicar da minha humanidade. Agradeço-lhe por todas as sessões de discussão científica, pela paciência até nas aprendizagens mais básicas, pela motivação constante e pela sua capacidade em acreditar em mim, mesmo nos momentos mais difíceis. Também muito importante é a sua constante preocupação em garantir, em toda a sua extensão, o bom funcionamento do trabalho laboratorial.

Gostaria também de agradecer à Professora Isabel Rocha por ter aceitado o desafio de co-orientar a minha tese de doutoramento. Obrigada pela forma tão calorosa como me recebeu no seu grupo e por toda a atenção que dedicou para que eu, a bióloga, percebesse as metodologias do mundo *in silico*. Agradeço também toda a sua compreensão, o seu apoio e motivação para que nunca desistisse deste projeto e por ter sempre arranjado tempo para reunir comigo. O seu conhecimento nesta área tende para infinito, fazendo com que o meu "modo esponja" fosse sempre ligado antes de cada discussão de trabalho. Obrigada por acreditar em mim e por me dar espaço para que eu pudesse defender as minhas ideias.

Um obrigada muito especial ao Doutor Nuno Borges por me ter ensinado a arte de ser cientista. Obrigada por toda a paciência, pelas reuniões intermináveis, apoio constante e por me teres estimulado a pensar no que estava a fazer e a perceber todos os passos que compreendem o desenvolvimento de um trabalho laboratorial. Agradeço-te por teres ouvido todas as minhas inseguranças e lamúrias e por teres retribuído com palavras motivadoras. Obrigada por todos os momentos de amizade, foi um gosto ser tua aluna! Gostava de te deixar uma palavra de apreço pela ajuda na elaboração desta tese.

Gostaria de agradecer ao Professor Nunes da Ponte, ao Doutor José Silva Lopes e à Doutora Fátima Lopes por me terem ajudado durante este programa doutoral.

Um agradecimento a todos os elementos que estiveram no grupo Cell Physiology and NMR desde que eu entrei por todos os momentos de discussão e de amizade. Um obrigada especial à Carla Jorge por me ter ajudado nas façanhas da sinucleína.

Quero também agradecer aos membros do grupo BisBII por todos os bons momentos, bom ambiente e conselhos científicos. Obrigada por toda a amizade e diversão que vos é tão característico!

Obrigada a todos os elementos do Laboratório de Bioprocessos pela ajuda e constante bom ambiente! Vocês tornaram as maratonas no laboratório muito mais fáceis de suportar ☺.

Gostava de também agradecer ao Rui, Ester, Touni e Flor, a super-equipa de BA, pela constante preocupação e apoio, pelos inúmeros jantares e sessões de cinema. À Sara, ao Paulo e à Carla, pelas reuniões urgentes e inadiáveis, sejam presenciais ou por videoconferência. Vocês são o verdadeiro "Mood Booster"! À Teresa e à Paty pela sua eterna amizade! Às minhas "piquininas", Joana A., Joana S., Antónia e Sofia que sempre me alegraram o dia com toda a sua maluquice. À Sophia, Patrícia, Marlene, Joana R., Ana Cristina, João e ao Óscar por estarem sempre dispostos a ajudar-me e a animar-me!

Gostava de agradecer à minha família por todo o apoio! Esta tese não teria sido possível sem vocês porque me acolheram quando eu precisei e me construíram para que eu conseguisse

ganhar forças para acabar esta etapa! É por esta razão que esta tese é dedicada a vós. OBRIGADA!

Também agradeço à Fundação para a Ciência e Tecnologia pela atribuição da bolsa de doutoramento (SFRH / BD / 79552 / 2011), ao programa doutoral MIT Portugal e ao programa FEDER – Programa Operacional Factores de Competitividade – COMPETE.

Resumo

Manosilglicerato (MG) é um soluto compatível envolvido na resposta a stress osmótico ou térmico em microrganismos marinhos hiper/termofílicos. Ensaios com proteínas-modelo demonstraram uma notável capacidade estabilizadora *in vitro*, justificando a aprovação de várias patentes. No entanto, os elevados custos de produção inviabilizam aplicações industriais. Esta tese tem dois objetivos, um de carácter sobretudo fundamental e outro visando a aplicação: i) estudar a eficácia de MG como estabilizador de proteínas *in vivo*, para fundamentar conjeturas sobre o seu papel fisiológico; e ii) desenvolvimento de um bioprocesso para produzir MG a preços competitivos.

O primeiro objetivo foi conseguido utilizando um modelo de agregação proteica em levedura, em que a α-sinucleína, fundida com eGFP, foi concomitantemente expressa com a enzima bifuncional que sintetiza MG a partir de GDP-manose e 3-fosfoglicerato. Observou-se uma redução de 3,3× no número de células com focos fluorescentes em comparação com células sem MG e provou-se que esta diminuição não era devida a efeitos indiretos. Este trabalho abre caminhos para o desenvolvimento de drogas contra doenças degenerativas associadas a deficiências estruturais.

Para o segundo objetivo, implementaram-se estratégias, com complexidade crescente, na levedura *Saccharomyces cerevisiae*. Numa primeira abordagem, sobre-expressaram-se os genes *PMI40* e *PSA1* com o intuito de aumentar a disponibilidade de GDP-manose, um precursor da biossíntese de MG. Esta estratégia resultou num aumento de 2,2× na produção, atingindo-se um rendimento de 15,86 $mg_{MG}.g_{PS}^{-1}$ para células em bioreator. Um valor superior de produtividade (1,79 $mg_{MG}.g_{PS}^{-1}h^{-1}$), conseguiu-se em modo contínuo com taxa de diluição de 0,15 h^{-1}. Seguidamente, foi adotada uma abordagem holística usando modelos à escala genómica para pesquisar soluções *in silico* capazes de desviar mais eficientemente o fluxo para a produção de MG. As estirpes sugeridas foram construídas e caracterizadas em “batch” e em contínuo, tendo o rendimento e a produtividade aumentado para 25,30 $mg_{MG}.g_{PS}^{-1}$ e 3,4 $mg_{MG}.L^{-1}h^{-1}$, respetivamente.

Keywords: manosilglicerato, levedura, estabilizador de proteínas, abordagem *in silico*

Abstract

Mannosylglycerate (MG) is a compatible solute implicated in the response to osmotic or heat stresses in many marine microorganisms adapted to hot environments. MG shows a remarkable ability to protect model proteins, especially against heat denaturation; however, high production costs prevented the industrial exploitation of these features. This thesis has two main objectives: i) to assess the efficacy of MG as protein stabilizer in the intracellular milieu; and ii) to develop a bio-based process for production of MG at competitive cost. The first goal was achieved by using a yeast model of Parkinson's disease in which an aggregation-prone protein, eGFP-tagged α-synuclein, was expressed along with the biosynthetic activities that catalyze the formation of MG from GDP-mannose and 3-phosphoglycerate. There was a reduction of 3.3-fold in the number of cells containing fluorescent foci of α-synuclein, in comparison with a control strain without MG. It was also proven that inhibition of aggregation was due to direct MG-protein effects*, i.e.*, MG acted *in vivo* as a chemical chaperone. This opened a way for drug development against diseases related with protein misfolding. Towards the second objective, genes *PMI40* and *PSA1* of the GDP-mannose pathway were over-expressed in the industrial microorganism, *Saccharomyces cerevisiae*, to redirect metabolic flux towards that MG precursor. This strategy led to 2.2-fold increase in MG production (15.86 $mg_{MG}.g_{DW}^{-1}$) for cells cultivated in controlled batch mode. Further improvement was achieved by cultivation in chemostat mode at a dilution rate of 0.15 h^{-1}; a constant productivity of 1.79 $mg_{MG}.g_{DW}^{-1}h^{-1}$ was reached. Next, a holist approach was undertaken by using *in silico* tools to identify engineering strategies that would lead to efficient channeling of substrates to MG production. The proposed strains were constructed and characterized in batch fermentation and continuous mode and led to an improved MG production of 25.3 $mg_{MG}.g_{DW}^{-1}$ and 3.4 $mg_{MG}.L^{-1}h^{-1}$, respectively.

Keywords: mannosylglycerate; yeast; protein stabilizer; *in silico* tools.

Table of contents

List of Figures

List of Tables

List of abbreviations

1,3bPG	1,3-Bisphospho-D-glycerate
2PG	2-Phosphoglycerate
3PG	3-Phosphoglycerate
3PHP	3-phosphohydroxypyruvate
ADP	Adenosine diphosphate
BPCY	Biomass product coupled yield
BSA	Bovine Serum Albumin
CDPdiacylglycerol	Cytidine diphosphate-diacylglycerol
CTP	Cytidine triphosphate
DHAP	Dihydroxyacetone phosphate
DHR123	Dihydrorhodamine 123
DIP	Di-*myo*-inositol phosphate
DNA	Deoxyribonucleic acid
DW	Dry weight
EDTA	Ethylene diamine tetraacetic acid
eGFP	Enhanced Green Fluorescent Protein
ENO2	Gene coding enolase
F6P	Fructose-6-phosphate
FBA	Flux balance analysis
FT-fuels	Fischer–Tropsch Fuels
G3P	Glyceraldehyde 3-phosphate
GAL1	Gene coding galactokinase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDP	Guanosine diphosphate
GDPman	GDP-mannose
GPR	Gene-protein-reaction
GRAS	Generally Regarded As Safe
HPLC	High-performance liquid chromatography
Hsp	Heat shock protein
LMOMA	Linear minimization of metabolic adjustment
M1P	Mannose-1-phosphate
M6P	Mannose-6-phosphate
MDIP	Mannosyl-di-*myo*-inositol phosphate
MG	Mannosylglycerate
MGS	Mannosylglycerate synthase
MILP	Mixed integer linear programming
MiMBl	Minimization of metabolites balance
MOMA	Minimization of metabolic adjustment

MPGP	Mannosyl-3-phosphoglycerate phosphatase
MPGS	Mannosyl-3-phosphoglycerate synthase
mRNA	Messenger ribonucleic acid
msgD	Mannosyl-3-phosphoglycerate synthase/phosphatase gene
NAD^+	β-nicotinamide adenine dinucleotide
NADH	β-nicotinamide adenine dinucleotide, reduced form
$NADP^+$	β-nicotinamide adenine dinucleotide
NADPH	β-nicotinamide Adenine dinucleotide 2′-phosphate, reduced form
NMR	Nuclear Magnetic Resonance
OECD	Organization for Economic Co-operation and Development
PAGE	Polyacrylamide Gel Electrophoresis
PD	Parkinson´s disease
PEP	Phosphoenolpyruvate
pFBA	Parsimonious flux balance analysis
Pi	Inorganic phosphate
PMI40	Gene coding the mannose-6-phosphate isomerase
PMI40p	Mannose-6-phosphate isomerase
PPi	Pyrophosphate
PS	Peso seco
PSA1	Gene coding the GDP-mannose pyrophosphorylase
PSA1p	GDP-mannose pyrophosphorylase
PYK1	Gene coding the pyruvate kinase
PYK1p	Pyruvate kinase
PYR	Pyruvate
ROOM	Regulatory on/off minimization
ROS	Reactive Oxygen Species
rpm	Rotations per minute
rRNA	Ribosomal ribonucleic acid
SA	Simulated Anneling
SD	Synthetic dextrose medium
SDS	Sodium Dodecyl Sulfate
SEC53	Gene coding the phosphomannomutase
SEC53p	Phosphomannomutase
SER3	Gene coding the 3-phosphoglycerate dehydrogenase and alpha-ketoglutarate reductase
SER33	Gene coding the 3-phosphoglycerate dehydrogenase and alpha-ketoglutarate reductase (paralog from *SER3*)
SNCA	Gene coding α-Synuclein
ThioT	Thioflavin T
tRNA	Transfer ribonucleic acid

URA	Uracil
vvm	Gas volume flow per unit of liquid volume per minute
YPD	Yeast extract, peptone and dextrose
α-Syn	Alpha-synuclein

Motivation, objectives, and thesis outline

For centuries humans have unknowingly modified microbial metabolism to produce food and beverages. Nowadays, we hold the knowledge and the tools to rationally redirect microbial metabolism from poor natural production to high titers. Even more, with the development of molecular biology, high-throughput techniques and computational tools, it is possible to identity uncultivated organisms and to illuminate their metabolic capacities.

Encouraged by these developments we set to produce, at competitive costs, mannosylglycerate, a remarkable compound with ability to protect proteins, especially against high temperature. A rational *in silico* design approach was applied with the goal to find engineering strategies that would lead to efficient channeling of substrates towards the production of mannosylglycerate. *Saccharomyces cerevisiae*, a widely-used cell factory, was selected to harbor the genetic alterations. By combining *in silico* design and optimization of microbial growth we aim at increasing the production of mannosylglycerate.

Context and Motivation

Mannosylglycerate (MG) is an osmolyte highly confined to Bacteria and Archaea adapted to thrive at high temperatures. Generally, this solute is accumulated in response to osmotic stress, but a few organisms accumulate MG at supraoptimal growth temperatures (Santos *et al.*, 2011). The strong correlation between thermophily and MG accumulation led to the view that this compound could play an important role in thermo-adaptation. Indeed, MG has a remarkable ability to protect proteins against heat damage and also to prevent protein aggregation *in vitro* (Faria *et al.*, 2003; Santos *et al.*, 2007). However, the question about the ability of MG to stabilize macromolecules in the intracellular milieu remained unanswered.

These properties led to the filing of several patents by our group and others (Lamosa *et al.*, 2009; Santos *et al.*, 1998; Schwarz, 2003). Major potential fields of application are: the cosmetic industry, as moisturizer and skin protector against UV damage, storage of vaccines and other biomaterials, protein stabilizer in analytical and clinical kits, and in biosensors. In particular, MG proved to increase the life-span of retroviral vectors (Cruz *et al.*, 2006) and to improve the sensitivity of DNA microarrays (Mascellani *et al.*, 2007). The major bottleneck towards the industrial utilization of MG derives from the current high production costs. Although a procedure for the chemical synthesis of MG has been developed, the yield is very poor (Lamosa *et al.*, 2006b). On the other hand, natural producers, generally thermophilic organisms, have very low growth yields and lead to prohibitive production costs. Therefore, the development of microbial systems for the industrial production of MG is mandatory to fully exploit the biotechnological potential of this osmolyte.

Metabolic engineering plays a crucial role in the production of valuable compounds from microorganisms. This is achieved by improving native routes of the selected organisms or by establishing new metabolic pathways optimized to fit the host. Further optimization usually includes cofactor balance and the modification of regulatory networks (Lee *et al.*, 2009). The acquired knowledge about microbial metabolism and the development of molecular biology techniques to modify genomes increased the number of bio-based products and reduced the time required to complete a project from bench to industry. A few examples of strain design using this built-up knowledge are the production of artemisinin (Paddon *et al.*, 2013), farnesene (Wang *et al.*, 2011) or 1,3-propanediol (Nakamura & Whited, 2003).

Advances in Systems and Synthetic Biology improved significantly the accuracy and success of strategies based on predicted solutions (Lee *et al.*, 2012). These solutions are obtained by using metabolic reconstructions that contain all information about the genetic elements and metabolic information of an organism (Feist *et al.*, 2009). Moreover, genome-scale metabolic models are useful to predict the phenotypic behavior of an organism in response to medium or genetic alterations, to identify drug targets or to predict genetic engineering strategies aimed at producing a desired compound (Maia *et al.*, 2016). Even more, the publication of these models encouraged researchers to formulate different and better methodologies to fit the user's need.

To date, the heterologous biosynthesis of MG is restricted to *Saccharomyces cerevisiae, Escherichia coli* and some plants (Empadinhas *et al.*, 2004; Sampaio *et al.*, 2003; Scheller *et al.*, 2010). The gene coding for the bifunctional mannosyl-3-phosphoglycerate synthase/phosphate from *Dehalococcoides mccartyi* (former *Dehalococcoides ethenogenes*) was expressed in *S. cerevisiae* by Empadinhas and co-workers, which resulted in the accumulation of MG (Empadinhas *et al.*, 2004). In *E. coli,* Sampaio and co-workers opted by using a different and exclusive pathway to produce MG by expressing a gene encoding the mannosylglycerate synthase from *Rhodothermus marinus* (Sampaio *et al.*, 2003). However, the production of MG was very weak and could only be detected using radiolabeled precursors.

S. cerevisiae is a model organism for which a great amount of information regarding physiology and metabolism is available. Moreover, it is widely used in industrial bioprocesses for the production of organic acids, amino acids, heterologous proteins and bioethanol (Borodina & Nielsen, 2014; Lee *et al.*, 2009). This microbial cell factory is easy to grow, have simple nutrient requirements and present a robust behavior necessary to thrive in industrial conditions (Blattner *et al.*, 1997; Goffeau *et al.*, 1996; Mortimer, 2000). A sequenced genome allied with a vast amount of molecular tools to express and edit genes, and several genome-scale metabolic models make this organism an ideal chassis to successfully manipulate metabolism to produce a variety of products (Hill *et al.*, 1986; McCloskey *et al.*, 2013; Osterlund *et al.*, 2012). Moreover, *S. cerevisiae* is stated as Generally Regarded As Safe (GRAS). The extensive knowledge and data availability

makes this organism an excellent candidate for strain design strategies aimed at increasing the production of MG up to industrial levels.

Research Aims

The ultimate goal of this thesis is to develop an efficient producer of mannosylglycerate at competitive costs. A well-established industrial organism, *S. cerevisiae*, was selected as host. The first objective was to prove that MG has ability to stabilize proteins *in vivo*. To do so, we studied the effect of MG in a yeast strain engineered to produce human α-synuclein - a protein prone to aggregate and involved in Parkinson's disease. Moreover, we examined different cultivation modes, assessed MG accumulation and characterized metabolism in all tested conditions. Using *in silico* strain design methods we identified and then implemented efficient engineering strategies, exploiting available genome-scale metabolic models developed for *S. cerevisiae*. A major concern was the maximization of carbon fluxes towards the production of precursors for MG synthesis (GDP-mannose and 3-phosphoglycerate), without compromising cell growth.

Outline of the Thesis

This thesis is organized into five chapters:

- Chapter 1 is dedicated to the origin and evolution of bioeconomics and the impact of Systems Biology in microbial cell factory engineering. Next, we describe the developments in the discovery of stress-protectant solutes, focusing on MG distribution, biosynthesis, function, and potentiality, as this thesis aims to improve MG productivity.
- In Chapter 2, we investigate the capability of MG to stabilize aggregating-prone proteins *in vivo.* To this goal, *S. cerevisiae* cells engineered to produce α-synuclein and MG were characterized and compared with cells producing α-synuclein, but not MG. In the end, we propose that MG acts as a chemical chaperone *in vivo*, and the stabilization mechanism involves direct solute/protein interactions.
- Chapter 3 describes the optimization of MG production based on the use of different cultivation modes. For this study, *S. cerevisiae* was genetically manipulated to overexpress the GDP-mannose pathway and to synthetize MG.
- In Chapter 4, several modelling methods were used in conjunction with SA algorithm to formulate different strategies and improve MG production based on *i*MM904 genome-scale model. The *in silico* solutions led to the construction of three strains intended to increase the flux towards MG precursors. These strains and the reference strain (harboring the gene to synthetize MG) were physiologically characterized using controlled bioreactors operated in batch mode and the experimental data was compared with *in silico* data.

- Chapter 5 comprises an overall discussion embracing the results presented in this thesis, the final conclusions and future perspectives.

This thesis is based on the following publications:

I - C. Faria, C.D. Jorge, N. Borges, S. Tenreiro, T.F. Outeiro, H. Santos, Inhibition of formation of α-synuclein inclusions by mannosylglycerate in a yeast model of Parkinson's disease., Biochim. Biophys. Acta. 1830 (2013) 4065–72. doi:10.1016/j.bbagen.2013.04.015.

II - C. Faria, N. Borges, I. Rocha [(a)], H. Santos [(a)]. Mannosylglycerate production in *Saccharomyces cerevisiae* by over-expression of the GDP-mannose pathway and bioprocess optimization. (2017). Manuscript in preparation.

[(a)] these authors contributed equally

III - C. Faria, N. Borges, H. Santos, I. Rocha. *In silico* design of *Saccharomyces cerevisiae* strains for high production of mannosylglycerate. (2017). Manuscript in preparation.

During the doctoral research period, an additional article was published that is not part of this thesis:

IV- R. Gallardo, C. Faria, L.R. Rodrigues, M.A. Pereira, M.M. Alves, Anaerobic granular sludge as a biocatalyst for 1,3-propanediol production from glycerol in continuous bioreactors., Bioresour. Technol. 155 (2014) 28–33. doi:10.1016/j.biortech.2013.12.008.

CHAPTER 1

Introduction

The manipulation of microorganisms to produce novel compounds is only possible thanks to decades of research on bioprocess engineering, recombinant DNA technology and the development of molecular tools. The expansion of computer power and the launch of “omics” approaches integrated with sophisticated genome-scale models, transformed metabolic engineering into a current holistic strategy with a sharp rise in the number of success cases.

There is an increasing social awareness on the pressing need to replace petrol-based production of chemicals by clean and sustainable processes. This societal change, concomitant with advances in cell physiology, evolved to a path where cells become the new factories for the production of all kind of commodity compounds, proteins and drugs. Additionally, the ability to produce a specific compound is no longer restricted to their natural producers and can be transferred to more suitable organisms. Along this line of thinking, we aimed at developing a microbial cell factory to produce mannosylglycerate, a remarkable protein stabilizer naturally produced by microorganisms adapted to thrive in hot environments.

1.1 The promising role of bioeconomy

The transformation of fossil resources has guaranteed the production of essential commodities - electricity, fuel and chemical compounds - that are used on a daily basis. Decades of industrial optimization using crude oil, natural gas and coal as raw materials, made the conversion of petrol-based products a highly productive and inexpensive technology. Nevertheless, the price of compounds derived from crude oil is highly susceptible to market variations, especially in non-petrol producing countries. The products derived from petrol can be categorized as low-price products and high-price products. Transportation fuel is an example of a cheap petrol by-product, while acrylic acid is a high-price, valuable compound (Clark *et al.*, 2015).

Exploitation of fossil resources allowed humankind to prosper, but it also propelled the emission of polluting gases and increased Earth´s average surface temperature, a phenomenon called global warming (OECD, 2011a). Climate changes are threatening many forms of life, and severely affect human well-being. Also, dependence on non-renewable raw material suffocates fragile economies due to unpredictable fluctuations in price. As a consequence, governments have been forced to look ahead and find alternative sources of energy and materials with increased sustainability and efficiency (Christensen *et al.*, 2008).

Governments are moving their attention to renewable resources, such as solar energy, wind power, water streams, and invest in the development of biotechnological strategies to decrease our environmental footprint. As a result of a worldwide campaign, the prefix Bio- is highly appreciated by conscious and environmentally aware consumers, which are looking for safer and cleaner products. Many reasons support a bio-based economy as this approach generates greener chemicals with reduced toxic side-products and low carbon dioxide emissions. Indeed, the development of processes to obtain renewable sources of energy and bio-based products is now part of many governmental guidelines (OECD, 2011b).

As modern societies shift towards a bio-based economy, the quest to produce fuels and chemicals from plants, microorganisms or wastes is on (Naik *et al.*, 2010). Two major success stories stand out in this area: bioethanol production from sugarcane (Goldemberg, 2007); and biodiesel production from vegetable oils and animal fats (Ma & Milford, 1999). Nonetheless, biotechnology is not only a mean to obtain safer energy or useful chemicals. It also brings in: innovations in agriculture as crops are improved to resist pesticides and cope with abiotic stresses such as extreme temperatures, high CO_2, excess or lack of water (Varshney *et al.*, 2011); production of a range of pharmaceutical drugs from antibiotics to anti-cancer and anti-parasitic agents, antioxidants, antivirals and hormones (Lee *et al.*, 2009); incorporation of “natural” flavors and scents in foods (Berger, 2007; Carroll *et al.*, 2015), and the production of valuable proteins for medical therapy and disease prevention (e.g., insulin, vaccines), or to be used as biocatalysts in a large range of industries (Demain & Vaishnav, 2009).

According to an OECD (Organization for Economic Co-operation and Development) prediction, by the year 2030, 35% of chemicals and other industrial products, 80% of pharmaceuticals and diagnostic products and 50% of all agricultural output will be produced via biotechnology (OECD, 2009). Then, bioeconomy could represent 2.7% of gross domestic product (GDP) for countries belonging to OECD and even higher percentages are expected for non-OECD members (OECD, 2009). In Europe, between 1,600 and 2,200 million tons of biomass are cultivated each year, generating 2 trillions of euros and employing more than 17 million people. Bioenergy and biomaterials produced using biomass account for 36% of the total production (Ronzon *et al.*, 2015). Bioscience companies in the USA generated in the last decade more than 111,000 highly-paid jobs; in 2014, about 1.62 million people were working in a total of 73,000 companies. Importantly, this industry was not touched by the 2007 economic crisis and has been growing 17% a year (Nat Biotech, 2014).

The successful implementation of a bioprocess is hampered by costs and time taken from engineering the organism to the development of a rentable production. A solution to reduce costs might pass by implementing bio-based processes blended in the well-established petrochemical industry. An alliance can be formed by using pre-existing refineries to produce solvents from biomass and/or petrochemical feedstocks, and reduce the inherent price fluctuations of food crops and petrol (Gu & Jérôme, 2013).

1.2 Biotechnology for production of chemicals

Worldwide, the production of bio-compounds reached 20 million of metric tons and is worth annually tens of billions of US dollars. In Europe, this market represents 25% of the total market of compounds and it is estimated that bio-based production will grow up to 1 million of metric tons by 2020 (Consulting, 2010).

It is known that bio-based economy arises from the need to produce compounds that are environmentally acceptable, which led to the appearance of a new area called ‘Green Chemistry’. Although this terminology was first referenced in 1998 by Anastas and Warner (Anastas & Warner, 1998), Green Chemistry started in the 80s with the principle of using environmentally friendly raw materials and processes to manufacture chemical products. The first definition states that: *Green Chemistry efficiently utilizes (preferably renewable) raw materials, eliminates waste and avoids the use of toxic and/or hazardous reagents and solvents in the manufacture and application of chemical products* (Sheldon, 2000). Indeed, Green Chemistry downgrades the importance of the economical prospect and embraces sustainable development and innovation. Anastas and Warner (Anastas & Warner, 1998) defined 12 principles of Green Chemistry that can be summarized as follows: minimize waste at source through the development of efficient processes instead of depending on remediation; improve health, safety, and protect the environment by

decreasing the use of toxic compounds and; use of renewable biomass in a sustainable manner, avoiding petroleum and other scarce resources (Sheldon *et al.*, 2007).

1.2.1 Biomass as the new substrate

There is a pressing urgency to find energy and materials sources able to replace crude oil, as population and their fair expectations to improve wealth rise worldwide. The current view is to substitute non-renewable fossil resources by biomass feedstocks to produce chemicals and fuels. This strategy reduces dependency on petrol sources and decreases carbon dioxide emissions in the manufacture of bio-compounds. To increase profitability, co-production methods are implemented in biorefineries by coupling biofuel manufacture with the production of commodity compounds (Vlysidis *et al.*, 2011).

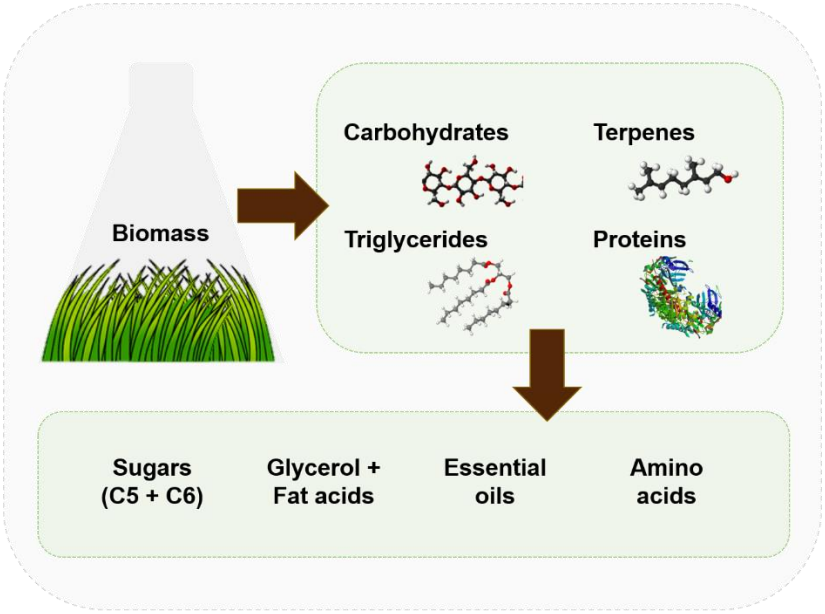

Figure 1.1 Biomass originates different kinds of components that can be transformed in precursors or end-products.

Generally, biomass is composed of carbohydrates (75%), lignin (20%), and triglycerides, terpenes and proteins (5%) (Sheldon, 2014). Carbohydrates present in biomass are starch, cellulose and hemicelluloses, which are the major components of plants (Figure 1.1). Arabinose and xylose are the most common C5-monomers, while glucose, galactose and mannose are representatives of abundant C6 sugars. Starch is a polymer of glucose that can be broken down into single molecules (Cherubini, 2010) Lignocellulosic biomass is composed of cellulose, hemicellulose and lignin. Cellulose is made of glucose molecules linked in a β(1-4) structure, which makes hard the task of breaking down into monomers. Hemicellulose is a mix of C5- and C6-molecules relatively easy to hydrolyze by using strong acids and heat. Lignin is a non-carbohydrate complex material that gives support and structure to plant cell walls via cross-linked phenol polymers. Vegetable oils and fats

fall into the triglyceride category and are composed of saturated and unsaturated fatty acids linked to a glycerol molecule. These are primarily used to produce biodiesel through esterification with methanol. Terpenes are a diverse class of organic compounds, generally exhibiting strong odors. Examples of well-known terpenes are limonene from citrus plants, farnesol, taxadiene, and carotenes. Proteins can be hydrolyzed into amino acids, which can be sold directly.

The first-generation of bio-based compounds was produced from crops commonly used in the food industry, such as maize, sugarcane and oilseeds. Nonetheless, as food market prices rose, a growing controversy surrounding the use of edible crops to produce chemicals and fuels started to echo. General public defends that land intended to grow food should not be relocated to manufacture chemicals (ActionAid, 2010).

To respond to this pertinent criticism, a second generation of bio-based compounds was envisaged that uses lignocellulose and non-edible seed oil as substrates. This is achieved by cultivating fast-growing and non-edible crops and by valorizing waste materials discarded by the food industry (Bautista *et al.*, 2007). Indeed, there is a huge waste of organic material generated in the harvest and processing of food that could be utilized as substrates in biorefineries. This way of producing fuels and chemicals truly respects the principles of green and sustainable chemistry. Even more, by eliminating waste it creates a closed loop between food and chemical production.

1.2.2 Conversion of biomass into commodity chemicals and fuel

Chemicals and fuels can be produced by chemical synthesis or by microbial fermentation or even by a combination of the two methods (Cherubini, 2010). Industry adapted crude extraction protocols to the new substrates - biomass - and has been able to produce several types of fuel and commodity chemicals. Some examples are summarized in Table 1.1.

Table 1.1 Examples of processes to produce chemicals and fuels from biomass.

Process	Intermediary compound	Function	Final product	Ref.
Gasification	Syngas (e.g., H_2, CO, CO_2, CH_4)	Biofuel	FT-fuels, dimethyl ester, ethanol, isobutene	(Maschio *et al.*, 1994),(Woolcock & Brown, 2013)
Gasification	Syngas (e.g., H_2, CO, CO_2, CH_4)	Chemicals	Methanol, ammonia, organic acids	(Maschio *et al.*, 1994),(Woolcock & Brown, 2013)

Pyrolysis	Pyrolytic oil Solid charcoal Light gases	Chemicals	Furfural, formic acid, acetic acid	(Liu *et al.*, 2014)
Hydrolysis	Monosaccharides (e.g., glucose, fructose, xylose)	Chemicals	Glycerol, glutamic acid, itaconic acid, malic acid, sorbitol	(Brethauer & Studer, 2015)
Transesterification	Methyl or ethyl esteres	Biofuel	Biodiesel, glycerol	(Sivasamy *et al.*, 2009)

Microorganisms have been used for centuries to produce fermented food and beverages. Back then, the knowledge about these processes was purely empirical and it was only in the XIX century that the existence of microbes was demonstrated. Nowadays, many microorganisms are recognized as being able to naturally produce commodity chemicals and fuels (Table 1.2). However, the natural production of these compounds is generally poor and unsatisfactory. To circumvent this problem, researchers were able to increase productivity by optimizing culture medium and environmental conditions. Another way to increase productivity involves the application of methods of random mutagenesis and directed evolution. This strategy is also called classical strain development and takes advantage of genetic mutations in an organism to enhance the productivity of a microbial cell factory (Patnaik, 2008).

Table 1.2 Commodity chemicals and fuels produced naturally by microorganisms using sustainable raw-materials and the correspondent industrial use for each product. Adapted from (Yin *et al*., 2015).

Organism	Substrate	Chemical	Use	Ref.
Aspergillus niger	Industrial waste	Citric acid	Food and pharmaceutical industries	(Singh Dhillon *et al.*, 2011)
Yarrowia lipolytica	Glycerol, ethanol, and vegetable oils	2-ketoglutaric acid	Dietary supplement and building-block chemical for cancer therapeutics	(Yu *et al.*, 2012)
Actinobacillus succinogenes	Xylose	Succinic acid	Additive and flavoring agent	(Bradfield & Nicol, 2016)
Rhizopus oryzae	Brewery wastewater	Fumaric acid	Food and beverage industries	(Das & Brar, 2014)
Rhizopus delemar	Corn straw hydrolyte	Malic acid	Food and beverage industries	(Li *et al.*, 2014)

Saccharomyces cerevisiae	Xylose, glucose, galactose	Ethanol	Fuel industry	(Olsson & Hahn-Hägerdal, 1993)
Lactobacillus helveticus	Whey permeate	Lactic acid	Pharmaceutical, chemical, and food industries	(Günther & Gottschalk, 1991)
Clostridium acetobutylicum	Maize mash	Butanol	Fuel industry	(Jones & Woods, 1986)
Anabaena sphaerica	CO_2 and light	Fatty acid methyl esters	Biofuel	(Anahas & Muralitharan, 2015)

1.3 Bioproduction using metabolic engineering

Metabolic engineering aims to improve the production of a compound with commercial value by inflicting modifications on target biochemical reactions and/or by improving cell fitness. These modifications involve recombinant DNA technology whether the strategy aims to alter endogenous reactions or to introduce novel pathways. In the early days of metabolic engineering, examples of directed genetic alterations towards the development of improved strains were scarce. Instead, scientists often relied on random mutagenesis protocols that offered no insight into the modifications imposed on cell metabolism (Stephanopoulos *et al.*, 1998). At this stage, metabolic engineering was powered by limited resources and focused on gene modification to improve a given pathway. Developments in PCR and Next-Generation DNA Sequencing boosted the area, leading to advanced tools that made organism manipulation fairly straightforward. These modifications include: gene deletion, expression of new genes synthesized to fit the host, overexpression of endogenous genes, modification of the transcript levels, or alterations on the regulatory network of the cell, to cite some examples.

A metabolic engineering strategy typically comprises three stages: i) selection of the host strain; ii) selection of genetic tools and iii) multiple rounds of metabolic pathway optimization. As stated above, the extraction of valuable compounds from natural sources is often characterized by low productivity. Changing the host into a cell factory compatible with the desired product, is the first step to increase productivity and transform a prohibited process into cost-effective production. However, selection of the host strain to accommodate a pathway is not an easy task. This organism must offer the "*chassis*" to produce the desired metabolite; it needs to grow easily up to very high cell density; and efficient genetic tools, such as expression vectors, chromosomal knockout protocols and insertion methods, must be available. The selected host should have

simple nutritional requisites and tolerate harsh environmental conditions. Additionally, it must be able to grow at various scales, ranging from flask to industrial-scale fermenters (Lee *et al.*, 2009).

The construction and optimization of metabolic pathways towards the production of a compound frequently comprises increasing specific metabolic fluxes and shutting down competing pathways to maximize product formation. The genetic modifications typically include over-expression of genes belonging to key routes, deletion of genes related with side products and the introduction of foreign genes (Liu *et al.*, 2015a). The overall goal is usually to increase the production of the desired compound without compromising cell growth.

A complete map of the metabolic pathways started to be drawn as organisms became engineerable. Allied to this knowledge, advances in analytical tools created a platform for systematic and more rigorous data. New strategies to balance cofactors and manipulate regulatory elements were added to the yet simple methodologies of gene expression and knockout. Today, metabolic engineering strategies go beyond targeting a small number of product-related pathways to an intervention that contemplates the grid of metabolites. In the end, the modifications aim to increase product formation, decrease process time and energy, annul by-product formation and develop strains resistant to environmental stresses (Kumar & Prasad, 2011).

Success cases of strain design are frequent in the literature. A few examples: the production of chemicals such as fumarate (Song *et al.*, 2013) and succinate (Zhang *et al.*, 2010) in *Escherichia coli;* or biofuels such as ethanol in *Saccharomyces cerevisiae* (Sonderegger *et al.*, 2004), butanol (Atsumi *et al.*, 2008) and isopropanol (Hanai *et al.*, 2007) by *E. coli;* pharmaceuticals with the production of terpenoids (Martin *et al.*, 2003) and artemisinin (Paddon & Keasling, 2014) by *E. coli*. Despite all efforts, some obstacles hinder the path that aims to substitute traditional production methods by effective bioprocesses. One way to overcome these difficulties is to build knowledge that will untangle the complexity of cell metabolism. Recent development in systems and synthetic biology areas will surely increase the success rate.

1.4 Systems biology applied to microbial engineering

Systems biology is a biology-based interdisciplinary field of study that focuses on obtaining a quantitative description of the biological entities using computational and mathematical modeling. This holistic approach aims to decode the inherent complexity of cells and predict how they react over time and variable conditions (Otero & Nielsen, 2010). Integrative biology aims to integrate knowledge at all levels of organization of living beings. Though this discipline has been active for the last 30 years, it was the development of computational tools and the ability to collect high-throughput data that boosted the area and allowed it to evolve to its current status (see Figure 1.2).

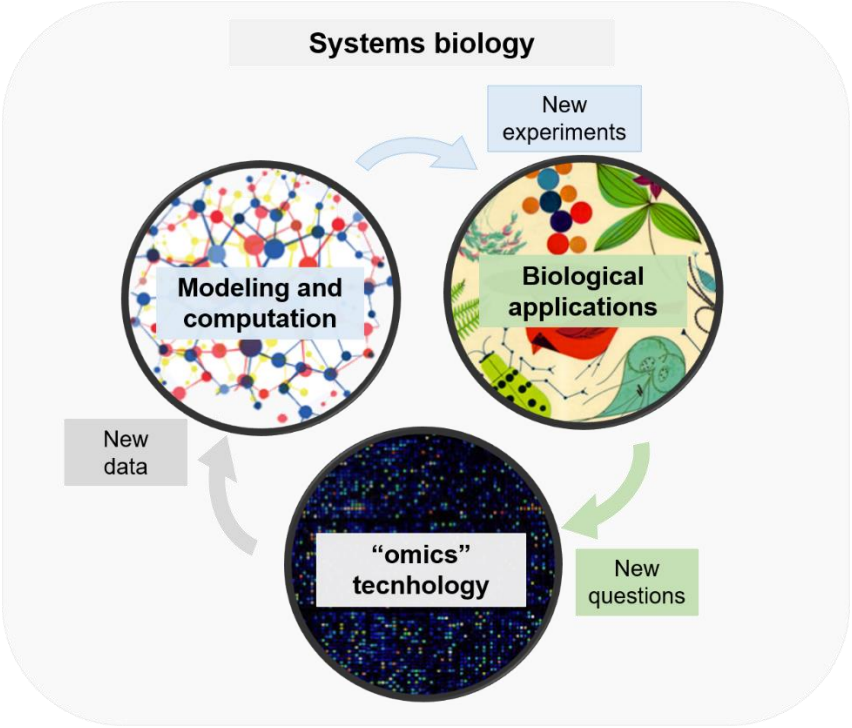

Figure 1.2 Systems biology as a system-level approach to understand biology through the use of the "omics" data and modeling.

The concept of systems biology is not new and has been applied in different branches of biology since the XIX century (Westerhoff & Palsson, 2004). However, advances in genome sequencing, transcription, protein and metabolite profiling in conjunction with the development of high-throughput methods were critical to the progress in this field by generating valuable information on the interactions among the network elements.

1.4.1 "Omics" tools in the context of metabolic engineering

Ideally, "Omics" technologies enable us to capture a global set of data from all single parts that compose a cell. In other words, we obtain information on which genes are expressed and the levels of expression, the pool of proteins and their modifications, the nature and concentration of metabolites and the fluxes through the complex metabolic network. These global approaches provide valuable information and guidelines for metabolic engineering strategies (Figure 1.3).

Genomes contain all the information regarding metabolism and physiology of an organism. All these elements are finely tuned into an extraordinarily complex network that responds to the environment in an attempt to survive and reproduce. This interconnected network of genes and regulatory elements present in genomes is the subject of study in the interdisciplinary field of genomics. The major goal of this area is to identify patterns that define and characterize cells and their surrounding environment (Lee *et al.*, 2012). As a result, the success of metabolic and cellular

engineering strategies increased significantly and led to microorganisms with enhanced yield and productivity of bio-products (Vemuri & Aristidou, 2005).

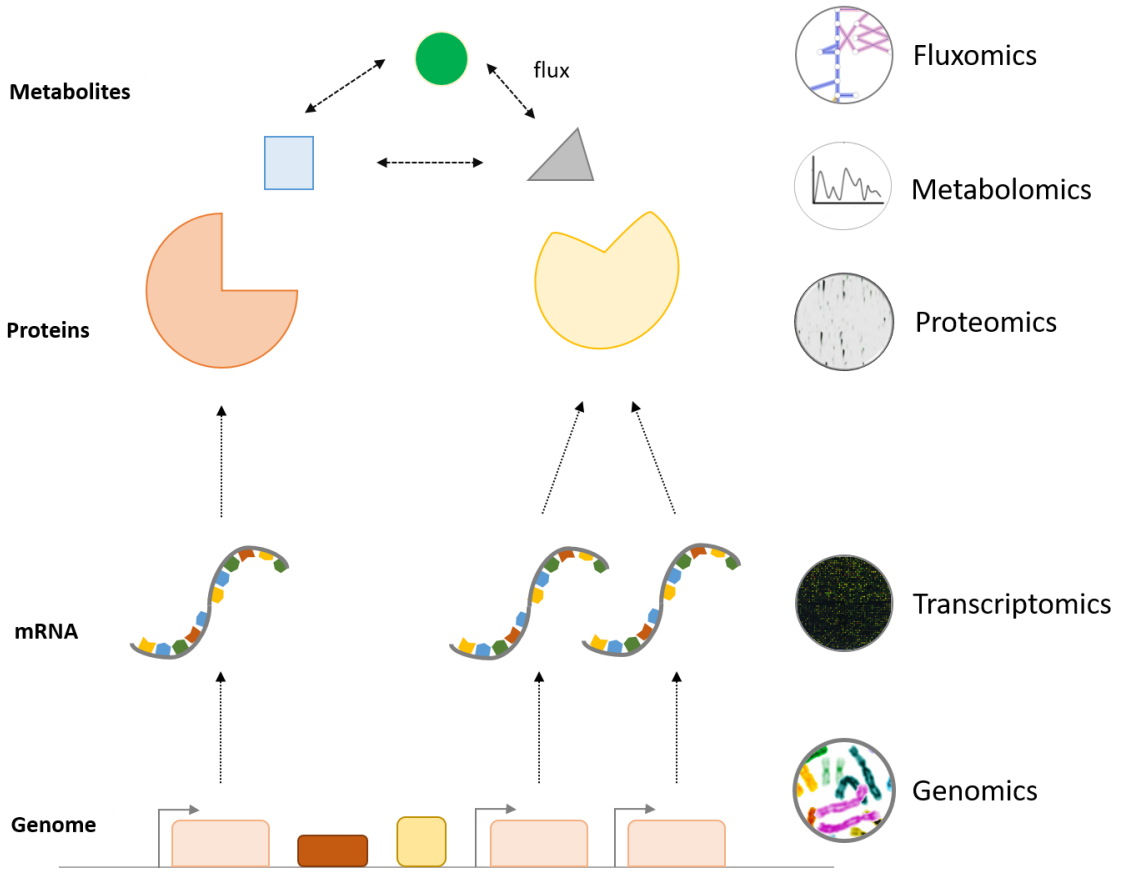

Figure 1.3 Schematic representation of the “omics” technologies and their correspondent objective of study.

Transcriptomics is the study of the complete set of mRNA transcripts generated in a cell and was firstly proposed in 1996 by Charles Auffray (Pietu *et al.*, 1996). This field emerged from the development of automated Sanger sequencing and microarrays technology (Lander, 1999; Schena *et al.*, 1995). From then on, it was possible to identify the changes in gene expression induced by environmental or genetic perturbations (Figure 1.3). The use of microarrays grew exponentially since this technology was able to determine correlations between environmental conditions/gene expression and the correlation between different genes (Vemuri & Aristidou, 2005). An important example was the inventory of the activated gene pool throughout growth of *S. cerevisiae* on glucose. The growth curve of *S. cerevisiae* on glucose presents a diauxic shift corresponding to a metabolic change from fermentation to respiration. DeRisi and co-workers analyzed this shift and were able to identify genes whose expression changed upon glucose depletion, which permitted the elaboration of a regulatory map for central carbon metabolism of *S. cerevisiae* growing on glucose (DeRisi *et al.*, 1997). However, microarray usage implies several restrictions as it demands pre-existing information about the genome, and presents methodology problems such as high background noise levels or signal saturation. Also, data from microarrays

are difficult to normalize and compare between experiments (Okoniewski & Miller, 2006). More recently, the development of high-throughput DNA sequencing methods (deep-sequencing technologies) lead to the invention of a new method that maps and quantifies transcriptomes (Wang *et al.*, 2009). This technique, called RNA-seq, can identify and quantify the majority of RNA species in the cell with an error in the range 1-5%, depending on the type of RNA analyzed and equipment used.

Proteins are the functional units of the cell, assembled in response to gene expression. Cellular protein pools are investigated by proteomics, which includes the determination of protein abundance, protein function, interaction, and pathways in which these catalysts are involved. Developments in 2D electrophoresis, mass spectrometry, fluorescence-based interaction assays and protein arrays allowed to gather a significant part of the proteome dataset that is used to understand the links between genome and protein abundance (Han & Lee, 2003). In the context of microbiology and bioprocessing, proteome profiling provides insights into cell metabolism and hints at strategies to enhance productivity and improve cell performance.

Moreover, knowledge of the pool of metabolites in cells attracts considerable attention from the scientific community, as metabolome constitutes the last piece that complements the genetic and protein information. Metabolite levels are the result of a finely tuned regulatory network that functions at the cellular level and the study of the complete set of metabolites is called metabolomics. Although we cannot identify and quantify all compounds present in a cell, we are able to measure concentrations of many small molecules such as carbohydrates, organic acids, alcohols, amino acids, lipids, among others. This is possible due to progress made in analytical methods such as mass spectrometry (MS), nuclear magnetic resonance (NMR), infrared absorption spectrometry (IR), ultraviolet spectroscopy (UV), high-performance liquid chromatography (HPLC), gas chromatography (GC) or near-infrared spectroscopy (NIR). Furthermore, ^{13}C labeling has provided information beyond end-product production by permitting the quantification of intermediary isotopomers from central and secondary metabolism. With this information, it is possible to infer the flux of production/consumption of many metabolites. Currently, 3755 metabolites of *E. coli* and 2027 metabolites of *S. cerevisiae* are annotated in ECMDB (Guo *et al.*, 2013) and YMDB (Jewison *et al.*, 2012), respectively. Usually, metabolites are identified through a sequence of steps that starts with quenching, which consists in stopping metabolism immediately after cell harvest and separating cells from the medium by centrifugation or fast filtration. Afterwards, metabolites are extracted and separated according to their chemical properties. Then, cell extracts can be analyzed using GC or LC coupled with a mass spectrometer or with NMR.

Spotting the limiting reactions on the network can be achieved through metabolomic analysis. For example, substrate accumulation is an indication that one or more enzymes are working at a slower rate than the preceding reactions. The identification of these bottlenecks allows the direct identification of these reactions as possible metabolic engineering targets. One such example was

described by Yamamoto and co-workers while attempting to over-produce alanine in *Corynebacterium glutamicum* (Yamamoto *et al.*, 2012). After observing the accumulation of intermediary compounds belonging to the alanine pathway, Yamamoto and co-workers designed a strategy to optimize the expression of several genes and balance co-factors, resulting in a 6.4-fold increase in productivity (Yamamoto *et al.*, 2012).

Metabolomic approaches provide information on the nature and concentration of a large set of metabolites present in an organism, at a specific moment, but do not infer on its origin. Hence, no information can be gathered to assess metabolic fluxes through the reactions occurring in the cell. Substrate limitation/accumulation is not the only factor conditioning the flux through a reaction, as there are other relevant parameters such as co-factor availability, regulation and enzyme concentration. Metabolic flux analysis (MFA) is used to determine flux distribution based on experimental data (Christensen & Nielsen, 2000; Wiechert & de Graaf, 1996; Wittmann, 2007). These data are collected by measuring labeled amino acids from cellular proteins using GC–LC/MS and/ or 2D-NMR. To label the amino acids, cells are grown on labeled ^{13}C substrates and harvested once proteins are fully-labeled, which normally occurs when cells are growing in steady-state for at least five residence times. With this information, it is possible to predict the metabolic capability of strains, energy and redox balancing or even investigate regulatory circuits that are condition-dependent (Lee *et al.*, 2011). The estimation of fluxes in a network is not trivial, however methodologies such as constraint-based flux analysis can use optimization-based simulation techniques to simulate the flux distributions in a genome-scale model in different conditions. Alternatively, it is also possible to estimate the fluxes in central metabolism using isotope-labelled substrates such as in 13C-based flux analysis. This methodology uses relative isotopomer abundance data to compute the flux distribution that best fits the available experimental data (Zamboni *et al.*, 2009). Briefly, ^{13}C-based flux analysis uses isotope-labeled substrates to determine the intracellular fluxes during the course of time. The fluxes are estimated by fitting iteratively the simulated fluxes with stoichiometric models. Constraint-based flux analysis uses optimization-based simulation techniques to detect the metabolic differences of two tested conditions. To implement this approach, a genome-scale model is needed.

The development of the “omics” technologies coupled to high-throughput assays led to the accumulation of an enormous amount of data whose analysis demand huge computer power. New programs and tools have been released that not only are able to process these data but can also build models and predict solutions for a given problem. This is the case of MET-IDEA that extracts semi-quantitative mass spectrometry datasets into organized data matrices (Broeckling *et al.*, 2006), or MZmine that processes mass spectrometry raw data and includes features as peak area parameterization of a large set of data (Katajamaa *et al.*, 2006).

1.4.2 Genome-scale models and *in silico* methods to improve metabolic engineering strategies

The sequencing of complete genomes provides information on the coding and non-coding elements of an organism. After the release of the first genome sequence in 1995 (Fleischmann *et al.*, 1995), Edwards and Palsson started to construct the first genome-scale metabolic model, which was published in 1999 (Edwards & Palsson, 1999). Nowadays, genome sequencing is an easy and affordable task. In 2016, the number of sequenced genomes listed in NCBI was 2708 for Eukaryotes and 59804 for Prokaryotes. Consequently, the number of genome-scale metabolic models available increased quickly and reached 134 in the year 2014 (Monk *et al.*, 2014).

These models can be used to predict the phenotypic behavior of a wild-type organism in response to different environmental stimuli. Even more, they allow the simulation of selected knockouts/overexpression of genes, outputting parameters such as specific growth rate and flux changes on the network (Milne *et al.*, 2009; Patil *et al.*, 2004). It is also possible to optimize the network by redirecting fluxes to produce, for instance, a desired compound. The first task in the construction of a genome-scale model passes by annotating each gene with a metabolic function to an enzymatic reaction. This work would be a daunting task if it was not for the automated reconstruction tools currently available (see Table 1.3). These automated tools are connected to important databases as Genbank (Benson *et al.*, 2013) or GENE (Maglott *et al.*, 2004), and are able to download information regarding an organism into their frameworks. For example, Merlin is able to annotate the genome of an organism, construct a map of reactions, identify genes encoding transporters and attribute a reaction to these carriers. Moreover, it performs compartmentalization of the model, predicts organelle localization of the proteins and attributes a gene-protein-reaction (GPR) association (Dias *et al.*, 2015).

Table 1.3 Computational tools for reconstruction of genome-scale models.

Tool	Brief description	Ref.
GLAMM	Reconstruction of metabolic networks from genome data; visualization of the network and construction of pathways	(Bates *et al.*, 2011)
Merlin	Automated association of gene-protein-reaction; reconstruction of metabolic networks and identification of transporters, compartmentalization and organelle localization. Adds biomass function	(Dias *et al.*, 2015)
Model SEED	Gene annotation from internal databases containing metabolic data	(Overbeek *et al.*, 2005)
Pathpred	Web-based server to predict enzyme-catalyzed reaction pathways from a compound using the	(Moriya *et al.*, 2010)

	KEGG and chemical structure alignments of substrate-product pairs.	
Pathway tools	Creates a type of model-organism database called a Pathway/Genome Database; Integrates genes, proteins, metabolic network and regulatory network of an organism.	(Paley & Karp, 2006), (Karp *et al.*, 2010)
RAVEN toolbox	Constructs semi-automated genome-scale models; Contains methods for visualizing simulation results and omics data	(Agren *et al.*, 2013)
SuBliMinal toolbox	Generates draft reconstructions, determines metabolite protonation state, mass and charge balancing reactions, suggesting intracellular compartmentalization; Adds transport reactions and a biomass function.	(Swainston *et al.*, 2011)

Multiple protocols can be found in the literature that aim to describe step by step all stages regarding the reconstruction of a model, including manual curation and validation (Feist *et al.*, 2009; Rocha *et al.*, 2008a; Thiele & Palsson, 2010). Briefly, information regarding genes, enzymes and transporters are collected. Next, the metabolic reactions of that microorganism are identified. Supported by Brenda, BKN-react and MetaCyc, the stoichiometry of the network is checked, along with compartmentalization and localization of the reactions. Once these steps are complete, the model needs experimental data regarding metabolism or, if not available, published data for a taxonomic-related organism is used. These data include biomass composition, growth-associated energy requirements, substrate consumption, end-product production and maximal growth rate in chemically defined medium (Dias *et al.*, 2015). Model validation is the last step and consists of comparing *in silico* growth in several environmental conditions with experimental data. The model is considered complete when it is able to predict biomass, metabolic yields and knockout effects in agreement with *in vivo* results.

Flux balance analysis (FBA) is a well-established approach to simulate metabolism and is commonly used to predict flux distributions in metabolic networks (Edwards & Palsson, 1999; Orth *et al.*, 2010). In the first stage, FBA represents the metabolic reactions in the form of a numerical matrix of the stoichiometric coefficients of each reaction, which directly arise from the model reconstruction process. This stoichiometry establishes flow-constrains for each reaction creating a constrain-based approach in which mass is balanced. This means that for all compounds, total consumption is equal to the total amount produced; for this, we have to assume that the system is in a steady-state mode. Also, every reaction is limited by an upper and lower bound for either consumption or production. Thus, by balancing mass and creating bounds we restrict the space of solutions regarding the distribution of flux belonging to each reaction. FBA operates with an

objective function which is, in the majority of cases, the maximization of biomass synthesis (growth), supported by the assumption that cells' main goal is to grow and divide, and to do so, their metabolic map of reactions have to lean towards the production of cellular components (Orth *et al.*, 2010).

1.4.3 Integration of genomic-scale models with the 'omics´ data

The precision of metabolic targeting can be enhanced by integrating datasets from transcriptomics, proteomics and fluxomics into the metabolic network. This integration favors the simulation of phenotypes for specific conditions (Saha *et al.*, 2014). There are two main categories of methods to integrate omics data with metabolic: a) the switch approach – it activates or deactivates the reaction flux based on the expression levels detected in the transcriptome profile concerning a certain condition (e.g., GIMME) (Becker & Palsson, 2008), and b) the valve approach – that inputs the relative gene/proteins levels from transcription or proteomic data on the reaction flux (e.g., E-flux) (Hyduke *et al.*, 2013). The main integration tools are summarized in Table 1.4.

Table 1.4 Examples of computational tools used to integrate transcriptomics/metabolomics data into genome-scale models. Adapted from (Liu *et al*., 2015a).

Tools	Brief description	Ref.
AdaM	Bi-level optimization framework to integrate transcriptomics data with genome-scale models	(Töpfer *et al.*, 2012)
E-flux	Predicts changes in metabolic flux capacity using gene expression data	(Colijn *et al.*, 2009)
EXAMO	Integrates gene-expression measurements with genome-scale models of metabolism	(Rossell *et al.*, 2013)
FCGs	Correlates gene expression transitions with their corresponding flux values	(Kim *et al.*, 2013)
GIM^{3}E	Develops condition-specific models based on an objective function, transcriptomics and cellular metabolomics data	(Schmidt *et al.*, 2013)
GIMME	Uses transcriptomic data and metabolic objectives to redesign a genome-scale model to a specific context	(Becker & Palsson, 2008)

GX-FBA	Combines gene expression data with flux balance analysis	(Navid & Almaas, 2012)
MADE	Uses changes in gene or protein expression to build a condition-specific metabolic model	(Jensen & Papin, 2011)
mCADRE	Produces human tissue-specific models based on gene expression data and metabolic network topology	(Wang *et al.*, 2012)
TEAM	Predicts temporal metabolic flux distributions using time-series gene expression data	(Collins *et al.*, 2012)
TIGER	Integrates metabolism, expression and regulation data to build genome-scale models	(Jensen *et al.*, 2011)

1.4.4 Computational methods for strain simulation and optimization

Genome-scale metabolic models possess the biochemical information of an organism and are used to simulate metabolic phenotypes of wild-type strains. Moreover, these models are able to predict the effects of genetic modifications in the behavior of an organism. Several methods have been developed to predict phenotypic alterations in the metabolic network (Table 1.5). As discussed above, FBA was the first method developed for simulating the growth rate of a wild-type or knock-out mutant. However, the assumption of biomass maximization usually used in FBA simulations can result in unrealistic predictions for engineered strains (Shlomi *et al.*, 2005a). This phenomenon can be explained by the fact that biological networks were formed in a slow process as result of natural selection, or in a more updated term – as a consequence of an evolutionary adaptation. As these processes occur, the cells with the fittest phenotypes will reproduce and be transferred to the next generation. Since cells with robust networks will be more adapted to a changing environment, natural selection promotes the appearance of cells that can grow and divide after an environmental or even genetic disturbance. In order to mimic the capacity of metabolic networks to counteract disturbances, Segrè and co-workers developed a method to simulate the impact of knockouts by minimizing the distance between fluxes of the wild-type and mutant strains (Segrè *et al.*, 2002).Their approach, called Minimization Of Metabolic Adjustment (MOMA), simulates gene knockouts, by setting the wild-type as reference and having as objective function the minimization of the Euclidean distance between the flux distribution of the mutant and wild-type strains. MOMA simulations are better than FBA at predicting gene essentiality because its assumption of minimal changes in the metabolic network describes more accurately the short-term phenotypes measured in the laboratory experiments. Following MOMA's publication, a modified version was released called Linear MOMA that solves some issues of the original

formulation by minimizing the Manhattan distance between the mutant and wild-type flux *distribution* (Becker *et al.*, 2007). All these methods are integrated in a plaflorm called OptFlux (Rocha *et al.*, 2010). Multiple strain engineering success cases have been published using MOMA, including the improvement of lycopene biosynthesis (Alper *et al.*, 2005) and L-valine (Park *et al.*, 2007) in engineered *E. coli* strains. Moreover, an 85% increase in cubebol titer was obtained in *S. cerevisiae* by applying a solution obtained by using OptGene as strain design framework and MOMA as simulation method (Asadollahi *et al.*, 2009).

Table 1.5 Simulation methods for genome-scale models.

Method	Description	Ref.
FBA	Calculates the flow of metabolites through the network assuming a cellular objective, usually biomass formation. Used to predict flux and growth in wild-type or knock-out strains.	(Orth *et al.*, 2010)
pFBA	Two-level target formulation that aims to predict flux distributions with FBA by minimizing the number of active reactions or the total sum of fluxes.	(Ponce de León *et al.*, 2008)
MOMA	Uses wild-type fluxes as a reference to calculate the flux distribution of a knockout mutant. This distribution is computed by minimizing the Euclidean distance between the reference and the mutant fluxes.	(Segrè *et al.*, 2002)
LMOMA	Intended to refine MOMA by using the Manhattan distance to simulate the network fluxes.	(Becker *et al.*, 2007)
MiMBl	Assumes that cells try to minimize the metabolite turnovers, meaning that metabolite production and consumption will be minimized in relation to the wild-type fluxes.	(Brochado *et al.*, 2012)
PSEUDO	The formulation of this method determines the metabolic fluxes from a nearly optimal growth and calculates the network fluxes of the mutants by deviating minimally from this region.	(Wintermute *et al.*, 2013)
ROOM	Predicts metabolic fluxes in response to gene knockouts by minimizing the number of flux changes using a comparison with the wild-type.	(Shlomi *et al.*, 2005a)
Under/Over-expression	Incorporates transcriptional / translational information into the flux constraints over related reactions and determines the overall effect of these changes.	(Gonçalves *et al.*, 2012)

Strain optimization is another feature deriving from genome-scale metabolic models construction and consists in automatically finding alterations in the network that will shift carbon fluxes towards

the production of the desired compound. These alterations include: gene deletion, gene over- or under-expression, heterologous insertions, or cofactor alterations (Maia *et al.*, 2016).

Nowadays, there is a large variety of computational methods that find strain engineering strategies to meet the purpose of designing novel cell factories using existing pathways or foreign ones (Table 1.6). The first strain design method was published in 2003 and was named as OptKnock (Burgard *et al.*, 2003). This method is formulated as a bi-level optimization problem (i.e., cellular objective and chemical production) and is able to identify multiple gene deletions that simultaneously combine cell growth with chemical production. OptKnock predicts sets of knock-outs that promote the production a target metabolite while maintaining a minimum growth rate. Although growth coupling is not guaranteed, this algorithm produces solutions where the production of the target compound becomes a necessary by-product of cell growth.

Since the first OptKnock article, several methods were published dealing with problems related to the initial formulation, as is the case of RobustKnock, which increased robustness of OptKnock solutions by optimizing the FBA formulation (Tepper & Shlomi, 2010), or ReacKnock in which the bi-level optimization problem was simplified to a single level (Xu *et al.*, 2013), and OptORF that adds regulatory information to the optimization, among others (Kim & Reed, 2010). However, the number of gene deletions set by the user progressively increases the time required to solve the problem. This issue limits the number of genetic modifications that can be included in the strain designer solution obtained with this type of method.

To tackle this drawback the first evolutionary optimization method emerged and was called OptGene (Patil *et al.*, 2005). Evolutionary methods enable solving large gene modification problems in less computational time and can be combined with simulation methods such as FBA or MOMA. Evolutionary methods were applied successfully in the production of succinic acid in *S. cerevisiae*, with 43-fold improved yield (Otero *et al.*, 2013), or in the production of vanillin, which increased 2-fold in *S. cerevisiae* (Brochado *et al.*, 2010) in comparison with previous reports (Hansen *et al.*, 2009), to cite some examples.

Table 1.6 Examples of computational methods for strain engineering using genome-scale metabolic models.

Method	**Description**	**Ref.**
OptForce	This algorithm defines the wild-type fluxes constrains either by calculating them or by importing experimentally measured fluxes and sets these fluxes as the interval to find target genes that will increase the productivity of the mutant strain.	(Ranganathan *et al.*, 2010)
CosMos	This method has a simpler formulation from OptForce by relying on continuous modifications to predict solutions that include deletions, downregulations, and	(Cotten & Reed, 2013)

	upregulations of fluxes that increase the desired compound	
FSEOF	Iterative linear optimization method that allows the identification of gene amplifications targeted to enhance metabolism towards production of a given compound	(Choi *et al.*, 2010)
EMILio	Identifies reactions with individually optimized fluxes that follow the objective of improving growth and production of the desired compound and predicts the optimal flux ranges that maximize that production	(Yang *et al.*, 2011b)
GDLS	First method to reduce significantly the computational burden by using iterative local search steps, enabling the use of complex models with complex questions	(Lun *et al.*, 2009)
OptGene	Evolutionary algorithm that optimizes strain design by discovering gene targets using available simulation methods present in Table 2.5.	(Patil *et al.*, 2005)
OptKnock	Bi-level optimization method that uses the FBA formulation in combination with the maximization of fluxes towards the production of a desired compound. The optimal solution found by OptKnock is a combination of knockout genes that links the target with cellular growth.	(Burgard *et al.*, 2003)
OptORF	Similar formulation than used in OptKnock with the addition of transcriptional regulation constrains. Target solution includes gene knockouts and over-expressions.	(Kim & Reed, 2010)
ReacKnock	Similar to the OptKnock method reformulated to change the bi-level optimization problem to a single optimization problem by using the Karush-Kuhn-Tucker method.	(Xu *et al.*, 2013)
RobustKnock	This method tackles OptKnock lack of robustness derived from the FBA simulation and guarantees robust solutions.	(Tepper & Shlomi, 2010)
SA/SEA	Simulated annealing and evolutionary algorithms to search near optimal sets of solutions that maximize growth coupling with production of the compound of interest.	(Rocha *et al.*, 2008b, Rocha *et al.*, 2010)(Rocha et al., 2008b)(Rocha et al.,

		2008b)(Rocha et al., 2008b)

The possibility of evaluating solutions formulated computationally to be used in strain design is a breakthrough scientific achievement and encourages the development of synthetic approaches. This excitement ascends from the perception that the time to finalize a project of strain design towards the production of a target compound is realistically and significantly smaller, in comparison with the traditional methods of trial and error.

1.5 *Saccharomyces cerevisiae* as a cell factory

Saccharomyces cerevisiae is one of the most important organisms used in industry and in strain design strategies to produce metabolic compounds of interest. In the scientific field, *S. cerevisiae*, also known as baker's yeast, is a model organism used all around the globe to understand fundamental cell processes that can often be replicated in higher Eukaryotes, including Humans (Botstein *et al.*, 1997). In industry, *S. cerevisiae* is used for millennia (early 6th millennium B.C.) in the production of beer, bread and wine (Mortimer, 2000). Nowadays, the utility spectrum has grown exponentially as this yeast is used for the production of bioethanol, nutraceuticals, chemicals and pharmaceuticals (Borodina & Nielsen, 2014).

S. cerevisiae was the first Eukaryotic organism to be fully sequenced (Goffeau *et al.*, 1996). It is easily manipulated and has a rich set of available engineering tools from genome editing to expression plasmids. Moreover, yeast cells are easy to grow in the laboratory, display fast growth rates and achieve high densities even with chemically defined media, a good characteristic to study genetics and proteomics. The development of a production process using cells often includes the expression of heterologous genes and construction of an optimized route to produce the desired compound. *S. cerevisiae* is a genetically stable organism, stated as generally regarded as safe and considered a robust organism to be used in industrial processes for its capability to tolerate low pH, preventing contaminations and reducing the production costs with acidic products. Additionally, this yeast shows a resilient behavior towards fermentation inhibitors produced during growth (Goffeau *et al.*, 1996).

An active and supportive community of researchers allowed the creation of general strain banks (Euroscarf) with single-knockout strains (Winzeler *et al.*, 1999), a collection of strains with green fluorescent protein (GFP) gene fused with a yeast gene (Huh *et al.*, 2003), and a strain collection engineered for protein over production (Jones *et al.*, 2008). *S. cerevisiae* is the best-studied Eukaryote and for this reason, there is an enormous amount of published data that was collected and integrated into online databases specific for each area of study. For instance, the SGD stores

information regarding genes (Cherry *et al.*, 2012), the YEASTRACT holds manually curated information regarding transcriptional regulation (Teixeira *et al.*, 2014), and YMDB, which is another manually curated database, stores information on the metabolome (Jewison *et al.*, 2012), among others. In addition, there is a strong development of metabolic models including 13 genome-scale models (Osterlund *et al.*, 2012).

1.6 Product optimization in *S. cerevisiae* by using fermentation design

A variety of methods and equipment/labware is available to cultivate microorganisms. However, and for its stability, bioreactors are the ideal equipment to cultivate microbial cells. Many reasons support this statement: a) substrates are consumed without any deviation; b) offers no dead zones or clumps of media; c) feeding can be immediately distributed throughout the reactor; d) the sparger provides a transfer of air into the liquid medium and e) optimal growth conditions are maintained (e.g., pH, O_2, agitation, temperature), and measured online with feedback control. This controlled environment guarantees the collection of reliable information regarding cell physiology and metabolism and is a powerful tool to decipher cell biology. The development of an optimized bioprocess can significantly increase the productivity and is the first step towards industrialization (Nielsen *et al.*, 2003). There are mainly three modes of cultivation that are applied in industrial processes: a) batch; b) fed-batch; and c) continuous mode (Figure 1.4). These three configurations are briefly discussed below for the specific case of *S. cerevisiae.*

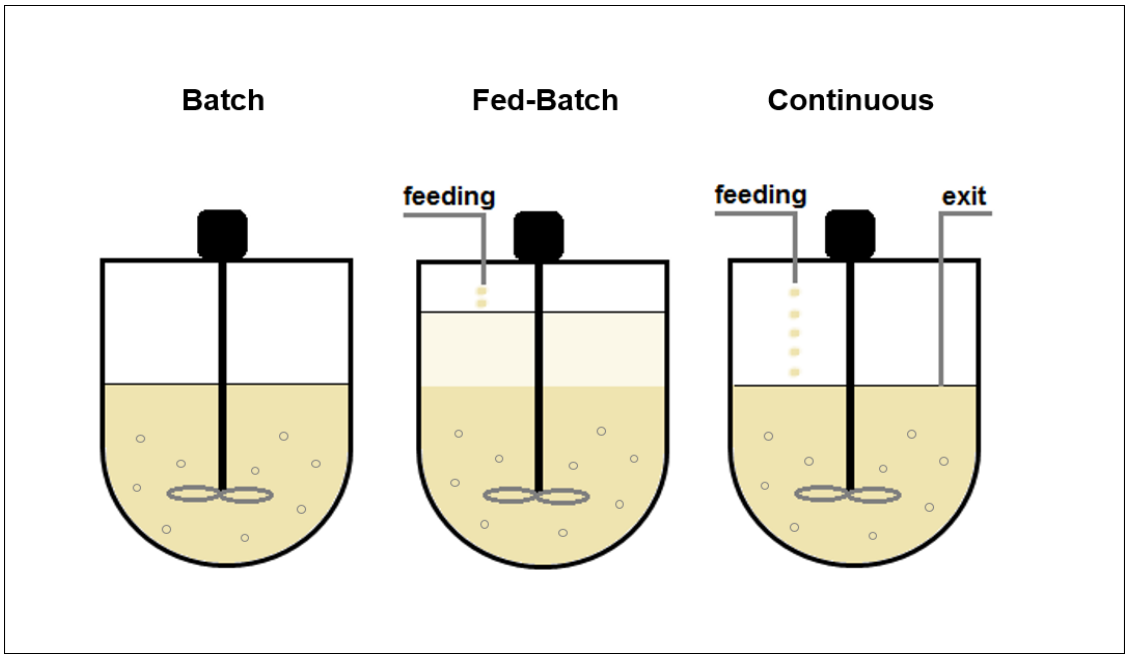

Figure 1.4 Simplified scheme of the main cultivation methods to establish a bioprocess in *S. cerevisiae*. Cells can be grown in batch mode characterized by the absence of feeding or exit of medium; fed-batch cultivation starts after a batch, with the addition of concentrated medium and no removal of culture broth; and continuous mode or chemostat is initiated with fresh medium after a batch growth, removed at the same dilution to maintain a constant working volume.

The batch mode is the easiest method of cultivation. Temperature, agitation, pH and dissolved oxygen are controlled, and a broth medium containing a carbon source plus all essential nutrients is provided to cells at the inoculation time. The bioreactor volume is kept constant and an exponential growth at a maximal rate is obtained. In an aerobic cultivation of *S. cerevisiae*, glucose is consumed in a respiratory/fermentative metabolism to produce ethanol (major end-product), acetate and glycerol, which are produced in smaller proportions. These end-products are consumed in a second-phase of growth after a period of metabolic adaptation. Batch mode has the advantage of allowing an everyday use for different fermentations and is easily sterilized with low risk of contamination (Nielsen *et al.*, 2003). Additionally, total conversion of substrate is possible and is often an efficient process. However, operating bioreactors requires skilled labor and the time-off fermentation which comprises sterilization, growth of pre-culture and cleaning conditions the productivity.

Fed-batch cultivation starts after cells were grown in a batch culture. This first step aims to guarantee a stabilized cell metabolism with a workable quantity of biomass. Feeding containing concentrated medium is added to the bioreactor near the phase when the carbon source is almost exhausted. The bioreactor working volume increases over time, consequence of the feeding and the absence of medium removal. Different feeding strategies can be applied in accordance with the process physiology. In the case of *S. cerevisiae,* the feeding strategy must have into account the type of metabolism favoring product-production. To maintain yeast cells on respiratory mode, glucose must be added below the threshold of the Crabtree effect, which is approximately 0.04 $g.L^{-1}$ (Pham *et al.*, 1998). In contrast, respiratory/fermentative metabolism is achieved with an exponential feeding rate maintained until oxygen reaches limiting conditions. From then on, feeding is kept at a constant rate, resulting in a low and constant concentration of glucose in the medium and a declining specific growth rate. Fed-batch cultivation combines the advantages of batch and continuous operation and offers great control and an optimization window to a desired product.

The continuous mode starts with a batch cultivation as operated in fed-batch mode. Feeding normally has the same composition of the batch culture medium and enters the bioreactor at the same rate as a level tube extracts the culture broth, to maintain a constant working volume. After several hours of cultivation, cells enter a metabolic steady-state. The dilution rate determines the specific growth rate of the culture until it reaches the μ_{max} of the culture (determined in batch cultivation), where $D = \mu_{max}$. Above this dilution rate, cellular growth can no longer accompany the dilution rate and cells start to be washed out from the vessel. This condition is observed when $D > \mu_{max}$. Chemostat cultivation is a powerful technique to unveil the metabolism of cells under different concentrations of glucose (Larsson *et al.*, 1993). Even more, it can pinpoint metabolic subtleties detected between the steady-states of different dilutions. The continuous mode, using *S. cerevisiae,* is industrially used to produce biomass, antibiotics and proteins (Nielsen *et al.*,

2003). Typically, these cultivations are made with a D = 0.1 h^{-1} or higher to obtain a productivity advantage in relation to batch cultivation. Continuous cultivation offers steady and continuous removal of product which is an enormous advantage in comparison to batch cultivation. However, this process can be hampered by contaminations and the introduction of deleterious mutations that affect production of the desired compound.

The combination of system biology expertise, synthetic biology tools and the optimization of down-stream and up-stream processes is opening a world of opportunities regarding the development of cell factories to produce exquisite compounds that once were never thought possible.

1.7 Stress-protectant solutes and their biological role

Astrobiologists have been actively searching for the presence of life beyond Earth for the last decades with no success so far, suggesting that life is a rare event. Even Earth, a planet that is at its most habitable, has extreme environments close to what Earth-forming conditions must have been. These extreme environments were thought to be sterile until microbial cells were found to thrive in these habitats (Brock, 1967).

Scientists were immediately intrigued by how these organisms could cope with high temperatures, radiation, pressure, desiccation, salinity, high or low pH. Some of these questions are now answered and adaptations include alterations on membrane composition to adjust fluidity (for high temperature), potent efflux pumps to remove toxic compounds (heavy-metal rich environments), or structural modifications regarding the rigidity of proteins to guarantee functionality (Rothschild & Mancinelli, 2001). Another physiological difference is the accumulation of small organic compounds to cope with osmolarity, temperature or other harsh conditions. These compounds can accumulate up to molar concentration without stressing cellular components and for that reason are called “compatible solutes” (Brown, 1976). Compatible solutes are widely spread throughout the different taxonomic groups from bacteria to archaea, fungi, algae, plants and animals (Santos & da Costa, 2002). These molecules are produced via different pathways and can be organized by their chemical structure as amino acids, sugars and derivatives, phosphodiesters and polyols (Figure 1.5).

Trehalose is a ubiquitous solute that can be found in lower orders of plants, lichens and algae, or in higher plants as the resurrection plant – *Selaginella lepidophylla,* and the model plant - *Arabidopsis thaliana*. Trehalose can also be found in different bacteria from the genera *Streptomyces*, *Corynebacterium*, *Mycobacterium*, *Propionibacterium*, *Deinococcus*, *Thermus* and also in *Escherichia coli,* among many others. In the domain Archaea, the genera *Sulfolobus* and *Thermoproteus* accumulate high levels of trehalose, but this solute is found in several other archaeons. In animals, it can be found in insects, eggs and round worms. It is a predominant solute in yeast and fungi where it accumulates in regular cells, spores, fruiting bodies and vegetative

cells (Elbein, 2003). Actually, trehalose can represent from 28-30% of a spore dry weight (Feofilova, 1992). Various functions are attributed to this solute: a source of energy, signal and regulator molecule to control pathways (yeast and plants), protector of proteins and cellular membranes against stress conditions including desiccation, dehydration, heat, cold and oxidation.

In Bacteria, the most frequent compatible solutes are: ectoine, glycine betaine, glycerol, proline and glutamate (da Costa *et al.*, 1998). Particularly ectoines, which are one of the most abundant osmolytes in Nature, can be found in organisms adapted to low-water-activity environments and function by minimizing the denaturation of proteins and other biological entities by interfering with the physical properties of water (Graf *et al.*, 2008). Also, glycine betaine acts as an osmo- and thermoprotectant in response to osmotic up-shock or increased temperature above optimal growth (Caldas *et al.*, 1999). This osmolyte is transported from the medium to the intracellular space in species such as *Escherichia coli, Staphylococcus aureus* or *Pseudomonas aeruginosa*, or it is synthesized from choline in two reaction steps. Only a few species are known to have a *de novo* synthesis pathway as in the chemoheterotrophic bacteria *Actinopolyspora halophila* (da Costa *et al.*, 1998).

On the other hand, solutes accumulated by thermophiles and hyperthermophiles are generally different from those found in mesophilic organisms and are often restricted to these two group of organisms (Santos & da Costa, 2002). These compounds can be divided into two categories: α-hexose derivatives and phosphodiesters (Lamosa *et al.*, 2007). Regarding the latter group, di-*myo*-inositol-phosphate (DIP) is the most distributed solute, and is restricted to microorganisms growing optimally above 60 ºC, including one of the most extreme hyperthermophiles known, *Pyrolobus fumarii* (Blöchl *et al.*, 1997; Gonçalves *et al.*, 2008). DIP can also be found mannosylated, yielding mannosyl-DIP and di-mannosyl-DIP, in members of the genera *Thermotoga* and *Aquifex* (Rodrigues *et al.*, 2009). Also, belonging to the phosphodiesters group there is diglycerol phosphate which is accumulated by members of the genus *Archaeoglobus,* and glycero-phospho-inositol, present in the genera *Archaeoglobus* and *Aquifex* (Gonçalves *et al.*, 2003; Lamosa *et al.*, 2006b).

Among the α-hexose derivatives, mannosylglycerate (MG) is the most common solute and can be found both in Bacteria and Archaea (Bouveng *et al.*, 1955). Some strains of *Rhodothermus marinus* also accumulate a MG derivative called mannosylglyceramide (Silva *et al.*, 1999). Mannosyl-glucosylglycerate is found in *Petrotoga* spp. and glucosyl-glucosylglycerate in *Persephonela marina,* which also accumulates glucosylglycerate (Fernandes *et al.*, 2010; Jorge *et al.*, 2007). However, there are exceptions that do not fit in these categories, as the case of β-glutamate accumulated by the halotolerant organisms *Methanothermococcus thermolithotrophicus, Methanocaldococcus jannaschii*, and *Methanotorris igneus* (Roberts, 2005; Robertson *et al.*, 1990) and also in hyperthermophilic bacteria (Fernandes *et al.*, 2010; Lamosa *et al.*, 2006b; Martins *et al.*, 1996) or aspartate present in the genera *Palaecoccus* and

Thermococcus (Lamosa *et al.*, 1998; Neves *et al.*, 2005), to cite some examples. As a general rule, ionic solutes are predominantly found in organisms adapted to thrive in hot environments, while mesophiles accumulate preferentially neutral solutes.

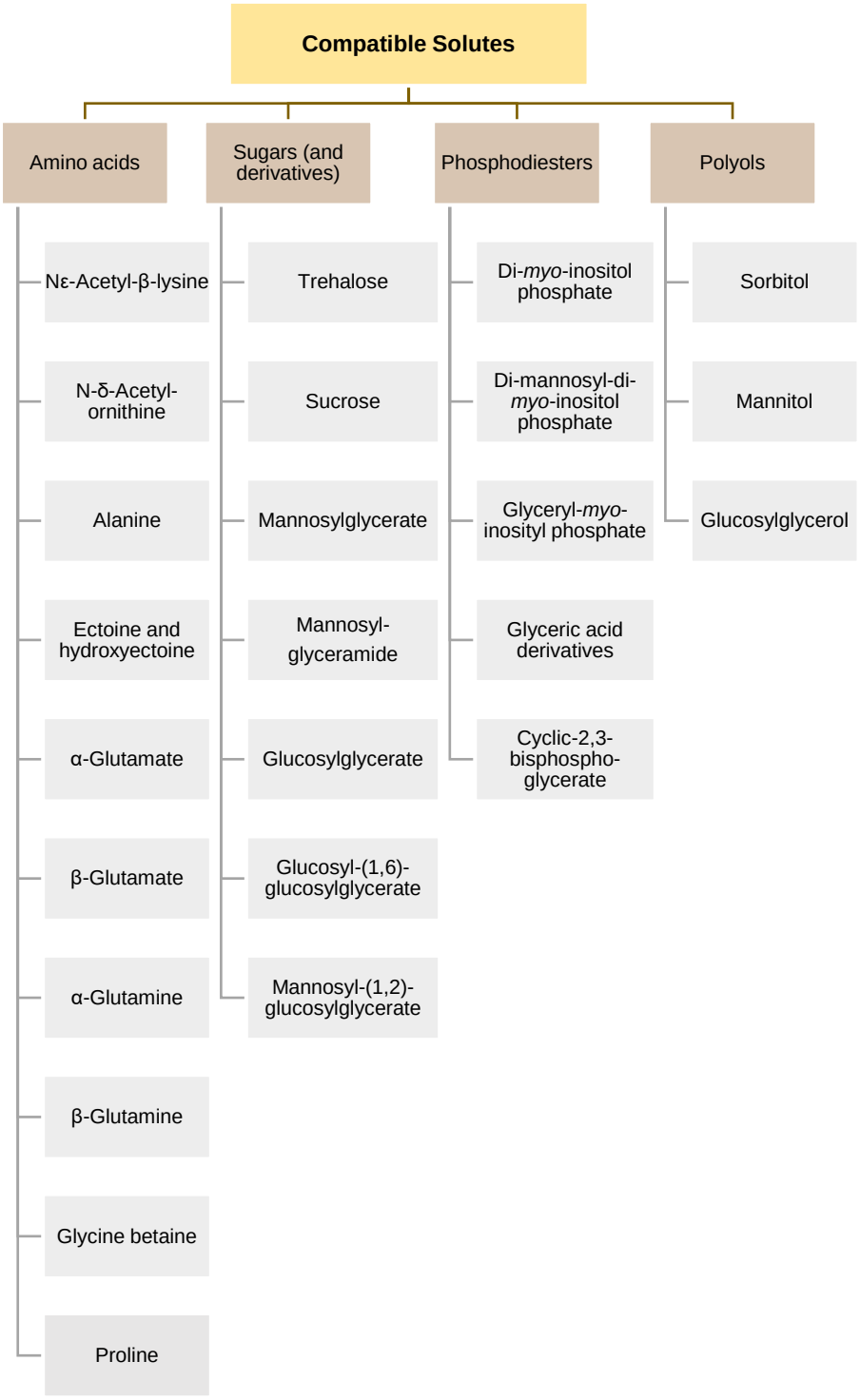

Figure 1.5 Organization of compatible solutes in groups according to their chemical nature.

1.8 Mannosylglycerate

Most microorganisms possess mechanisms to effectively respond to increasing concentrations of salt in the environment, by the counterbalance of ions or solutes. However, these stress-protectant or compatible solutes surpass their function as “osmotic balancers” and evolved to help cells in the protection of their molecules and cellular structures under different stressful conditions. This is the case of mannosylglycerate (MG), a compound found in the three domains of the Tree of Life (Liu *et al.*, 2015b): Bacteria, Archaea and Eukarya.

1.8.1 Structure and distribution of mannosylglycerate

Mannosylglycerate or 2-*O*-α-D-mannopyranosyl-D-glycerate is a compound belonging to the glycopyranosyl–glycerols family and is composed of a glycerate molecule linked to a mannose sugar (see Figure 1.6) (Bouveng *et al.*, 1955; Claude *et al.*, 2009; Silva *et al.*, 1999). MG was firstly named in 1939 as digeneaside, identified by Colin and Augie in a seaweed called *Polysiphonia fastigiata* (Colin & Augier, 1939). MG structure was first elucidated by Bouveng and co-workers in 1955 (Bouveng *et al.*, 1955) and later established by NMR (Ascêncio *et al.*, 2006; Silva *et al.*, 1999). Other species also belonging to the order Ceramiales were found to accumulate MG (Kremer, 1980) and more recently it was detected in Rhodophyceae of the orders Gelidiales, Gigartinales, Porphyriales, Rhodymeniales, Gracilariales and Stylonematales (Eggert *et al.*, 2007; Karsten *et al.*, 2007; Yang *et al.*, 2011a). Although MG can be found in 46 genera of algae, the physiological role of this solute in these cells is yet unclear, raising the suspicion that MG may serve as carbon reserve (Borges *et al.*, 2014).

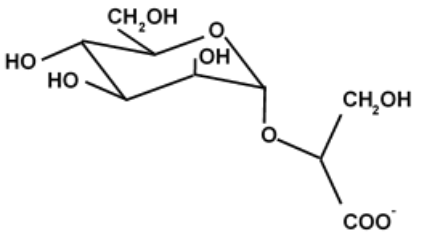

Figure 1.6 Schematic representation of mannosylglycerate, also known as digeneaside.

MG was assumed to be restricted to algae until it was first identified in the bacteria *Thermus thermophilus* and *Rhodothermus marinus* in 1995 (Martins & Santos, 1995; Nunes *et al.*, 1995). In *T. themophilus,* MG is accumulated for adaptation at low osmolarity, but otherwise trehalose is the major osmolyte (Alarico *et al.*, 2007; Silva *et al.*, 2003). The same function is observed in *R. marinus* for low-salinity conditions, in which MG concentration increases concomitantly. However, at higher salinities, trehalose is substituted by a MG derivative – mannosylglyceramide (Silva *et al.*, 1999). In addition, MG was also observed to accumulate in response to heat stress (Empadinhas *et al.*, 2001; Martins & Santos, 1995; Neves *et al.*, 2005). MG is also present in organisms extremely resistant to gamma-radiation, belonging to the genus *Rubrobacter* (Carreto *et al.*, 1996). However, no effect on MG accumulation is observed when an osmotic stress is applied to *Rubrobacter xylanophilus* that grows optimally at 60 ºC (Empadinhas *et al.*, 2007). In

Archaea, MG has been found in the hyperthermophiles of the genera *Pyrococcus*, *Thermococcus*, *Palaeococcus*, *Archaeoglobus*, *Aeropyrum* and *Stetteria* (Martins & Santos, 1995; Neves *et al.*, 2005; Santos & da Costa, 2002). Its accumulation is a response to increasing levels of NaCl or, for the specific case of *Palaeococcus ferrophilus,* MG production is associated with thermal stress (Neves *et al.*, 2005).

1.8.2 Biosynthesis of mannosylglycerate

The biosynthetic pathway of MG was first revealed in *R. marinus* by the discovery of a glucosyltransferase named mannosylglycerate synthase (MGS) that catalyzes, in a single step, the conversion of GDP-mannose and D-glycerate into MG. Strangely, it was later found that *R. marinus* has an alternative route to produce MG involving two-steps and two enzymes. This alternative pathway uses 3-phosphoglycerate (3PG) and GDP-mannose which are condensed into mannosyl-3-phosphoglycerate by the mannosyl-3-phosphoglycerate synthase (MPGS), which is subsequently dephosphorylated by the mannosyl-3-phosphoglycerate phosphatase (MPGP), yielding mannosylglycerate (Figure 1.7).

Only three organisms are known to produce MG via the single-step pathway, i.e., the thermophilic bacterium *Rhodothermus marinus* (Martins *et al.*, 1999), the mesophilic red alga *Caloglossa leprieurii* (Borges *et al.*, 2014), and the plant *Selaginella moellendorffii* (Nobre *et al.*, 2013). The two-step pathway for the synthesis of MG is much more common and can be found in the thermophilic bacteria *Rhodothermus marinus*, *Thermus thermophilus, Rubrobacter xylanophilus* and the hyperthermophilic archaea *Pyrococcus* spp.*, Palaeococcus ferrophilus and Thermococcus litoralis* (Empadinhas *et al.*, 2001, 2003; Empadinhas & da Costa, 2011; Martins *et al.*, 1999; Neves *et al.*, 2005). The genes coding for the MPGS and MPGP are usually separated in the genome but placed close together, except for the bacterium *Dehalococcoides mccartyi*, (former *D. ethenogenes*) where the *mpgS* and *mpgP* genes are fused in a single coding sequence. This bifunctional enzyme was expressed in *Saccharomyces cerevisiae* and proved to be functional by allowing the accumulation of MG in this Eukaryote (Empadinhas *et al.*, 2004). However, whether MG accumulates in *D. mccartyi* remains unknown. Also interesting is the presence of the gene *mpgS* in *Magnaporthe grisea* which was functionality verified in *S. cerevisiae*. However, this organism lacks the gene *mpgP,* indicating that *M. grisea* is only capable of producing mannosyl-3-phosphoglycerate (Empadinhas, 2004).

Since the disclosure of the genes involved in the biosynthesis of MG it is possible to search for homologs within the publicly available genomes. Regarding the single-step pathway, the homologs predicted for the enzyme MGS are: red algae *Griffithsia japonica, Griffithsia okiensis, Chondrus crispus, Kappaphycus alvarezii, Eucheuma denticulatum*, and *Gracilaria changii*, green algae *Penium margaritaceum, Klebsormidium flaccidum, Spirogyra pratensis,* and *Chaetosphaeridium globosum* and also in the moss *Physcomitrella patens* subsp. *patens* (Borges *et al.*, 2014).

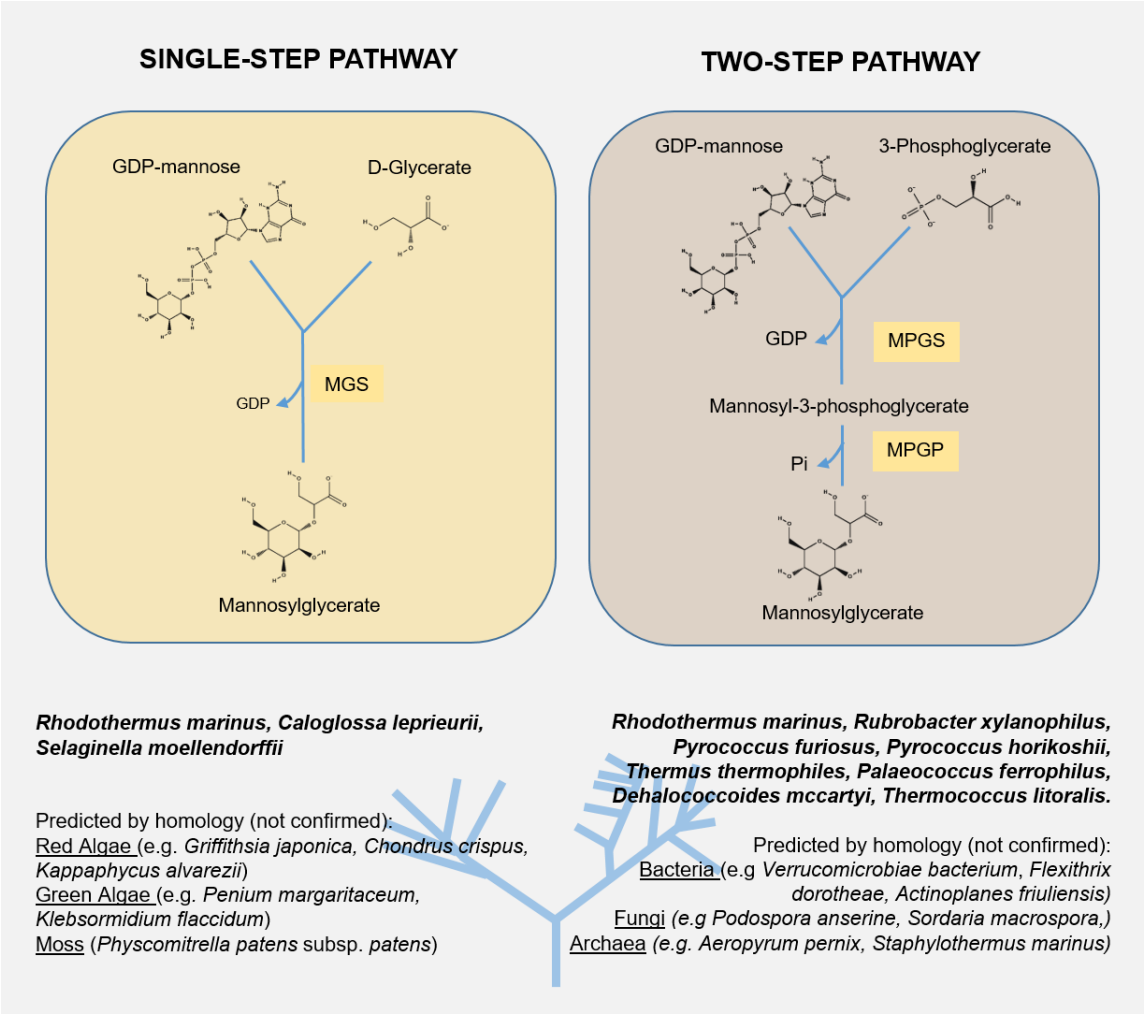

Figure 1.7 Biosynthetic pathways for the production of mannosylglycerate and their occurrence in the Tree of Life. Organisms with confirmed MG enzymes are in bold. Homology predictions were obtained using *Rhodothermus marinus* and *Pyrococcus horikoshii* as template for the single-step and two-step pathway, respectively. Abbreviations: mannosylglycerate synthase (MGS); mannosyl-3-phosphoglycerate synthase (MPGS), mannosyl-3-phosphoglycerate phosphatase (MPGP).

Enzymes for the two-step pathway are much more common in the Tree of Life. We found hits in Archaea (Euryarchaeota: *Pyrococcus, Thermococcus, Palaeococcus, Methanothermus* and *Archaeoglobus*; Crenarchaeota: *Aeropyrum* and *Staphylothermus; Thaumarchaeota: Candidatus Nitrososphaera*), Bacteria (*Rhodothermus, Flexithrix, Verrucomicrobiae, Dehalococcoides, Dehalogenimolas, Thermus, Marinithermus, Catelliglobosispora* and *Actinoplanes*), and the Eukarya (Fungi: *Sordaria, Neurospora, Podospora, Chaetomium, Thielavia, Myceliophthora, Magnaporthe, Gaeumannomyces, Togninia, Eutypa, Colletotrichum, Pyrenophora, Phaeosphaeria* and *Leptosphaeria)* (Borges *et al.*, 2014).

Until recently, there was no information on whether cells could recycle the MG that is produced in response to abiotic stresses. Nobre and co-workers were able to identify MG hydrolases in the

organisms *T. thermophilus*, *R. radiotolerans* and *S. moellendorffii* that are capable of hydrolyzing MG into mannose and D-glycerate (Nobre *et al.*, 2013).

1.8.3 Biological function of mannosylglycerate

The presence of MG in organisms adapted to hot environments linked MG to survival in extreme environments, especially in response to osmotic stress. That is the case of *R. marinus* and *T. thermophilus*, in which MG is produced in response to fluctuations of salt (Nunes *et al.*, 1995; Silva *et al.*, 1999). In addition, these two organisms and the archaea *P. ferrophilus* also use MG as a protector against thermal stress (Empadinhas *et al.*, 2001; Martins & Santos, 1995; Neves *et al.*, 2005). Conversely, the physiological role of MG is difficult to understand when the level of this solute remains unchanged upon application of osmotic or thermal stresses. This is the case of organisms belonging to red algae and the gamma-radiation resistant bacterium *R. xylanophilus* (Empadinhas & da Costa, 2011). In red algae, mannitol, digalactosylglycerol and sorbitol are the most frequently found compatible solutes responsible for cells survival in response to osmotic stress (Karsten & West, 1993). In face of these results, some authors speculate that MG may be used by these organisms as a carbon reserve.

Recently, in an attempt to study the physiological role of MG in *Pyrococcus furiosus* the genes encoding the key biosynthetic enzyme were deleted. The resulting mutant was submitted to heat and osmotic stress and its growth performance compared with the parent strain (Esteves *et al.*, 2014). The changes are small, and solute replacement makes difficult the interpretation of results, but there is evidence that MG is especially suited to protect these cells against osmotic stress conditions. On the other hand, MG is as effective as DIP in the protection under heat stress conditions.

1.8.4 Potential applications of Mannosylglycerate

Osmolytes are known to protect organisms from low temperature, desiccation, urea accumulation and high salinity. Particularly, MG shows a remarkable ability to protect proteins, especially against heat denaturation and to prevent protein aggregation *in vitro* (Faria *et al.*, 2003; Santos *et al.*, 2007). For this reason, we recognize that MG could have applications in several areas of biotechnological interest:

a) Engineering of transgenic organisms with stress-resistant properties

The cultivation of improved crops able to tolerate harsher habitats and resist to abrupt climate changes is currently on demand. The expression of genes to allow the synthesis of osmolytes has been shown to improve the fitness of organisms against several stresses. That is the case of a halo- and thermo- tolerant transgenic *Arabidopsis thaliana* modified to produce glycine betaine (Alia *et al.*, 1998); or a transgenic *Nicotiana tabacum* (tobacco plant) containing genes to allow the

production of glycine betaine, which confers tolerance to colder temperatures and salt stress (Holmström *et al.*, 2000). Also, the transgenic plants *A. thaliana,* tobacco plant and grasses already showed improved tolerance to drought and heat stress when engineered with genes to produce MG (Scheller *et al.*, 2010).

b) Stabilizer in the pharmaceutical, life science and high technology companies

MG has exquisite properties in the protection of enzymes and could be used in the preservation of thermosensitive products, such as, aggregating-prone proteins, enzymes and vaccines. Applications in diagnostic kits are also envisaged. It has been proven that MG improves the life-span of retrovirus vectors and the quality of DNA microarray experiments (Cruz *et al.*, 2006; Mascellani *et al.*, 2007). Indeed*, in vitro* experiments revealed that the incubation of the model hexapeptide STVIIE (forms self-fibrils) with MG, causes a strong inhibition of fibril formation; moreover, MG is also able to disassemble pre-formed fibrils of STVIIE (Santos *et al.*, 2007). Faria and co-workers set to investigate the effect of MG, through thermodynamic analysis, in the thermal folding of the model protein ribonuclease A. In this study, MG revealed to be an efficient thermostabiliser of ribonuclease A by inducing an increase of 6 ºC/Molar in the melting temperature of the protein. The mechanism of action suggested by the authors consists on the stabilization of the protein structure by forcing a more rigid structure (Faria *et al.*, 2003). Upon denaturation of ribonuclease A, MG inhibited the formation of aggregates, demonstrating that this solute can be used as a protein stabilizer. In another study using a *Staphylococcus aureus* recombinant nuclease A (SNase), MG increased the melting temperature of the protein in 7 ºC and increased in 2-fold the unfolding heat capacity (Faria *et al.*, 2004). An in-depth study was designed to understand the mechanism of action. Using NMR, the authors concluded that the protein backbone was significantly constrained in the presence of MG. In addition, the internal motions of the SNase were gradually restricted with increasing concentrations of MG, linking protein stabilization with protein rigidification (Pais *et al.*, 2009).

c) Use in cosmetic lines as a skin protector and moisturizer

Having into consideration the extraordinary properties of MG, it is feasible to assume the potentiality of this solute as a component of UV damage protecting creams. Osmolytes are able to protect cells from different stresses which extended to the protection of skin cells opens an opportunity to use this compound in cosmetic products with anti-aging and revitalization properties. Several patents demonstrate these properties for MG (da Costa *et al.*, 2003; Lamosa *et al.*, 2006a; Santos *et al.*, 1996, 1998). Ectoine is already used as a skin protector against UVA-induced damage (Buenger & Driller, 2004) and to ameliorate symptoms of atopic dermatitis (Marini *et al.*, 2014).

d) Chemical chaperone to treat neurodegenerative diseases

Some neurodegenerative diseases are characterized by an extensive inflammatory response to bulks of aggregated proteins which promotes cell death in the affected area. The most important

and prevailing diseases presenting this pathology are: Alzheimer's, Parkinson's and Huntington's disease. Several osmolytes showed to be promising candidates to serve as therapeutic drugs (reviewed by (Jorge *et al.*, 2016)). In a mammalian model of Alzheimer's disease, MG was able to reduce protein aggregation and consequently improve cell viability by detoxifying cells from Aβ42 oligomers/fibrils (Ryu *et al.*, 2008). However, this study was elaborated by adding different concentrations of MG to the extracellular space of cells. As some diseases, present protein dysfunction in the cytoplasm, we set to evaluate the anti-aggregating potential of MG in the intercellular milleu. To do so, we have engineered MG-production in a Parkinson's disease yeast model (see Chapter 2). MG performance in preventing aggregation of α-synuclein reaffirms the potentially of this compound as a therapeutic drug.

The major bottleneck towards the industrial utilization of MG is the high production costs. Although a procedure for the chemical synthesis of MG has been developed (Lamosa *et al.*, 2006b), the yield is poor. Moreover, there is a strong bias of consumers against synthetic products. Currently, commercial MG is produced (gram amounts) from fermentation and "milking" of *R. marinus*, a natural producer, but the efficiency of the process is unsatisfactory. In this context, the development of an efficient industrial producer of MG is mandatory.

CHAPTER 2

Inhibition of formation of α-synuclein inclusions by mannosylglycerate in a yeast model of Parkinson´s disease

Protein aggregation in the brain is a central hallmark in many neurodegenerative diseases. In Parkinson's disease, α-synuclein (α-Syn) is the major component of the intraneuronal inclusions found in the brains of patients. Current therapeutics is merely symptomatic, and there is a pressing need for developing novel therapies. Previously we showed that mannosylglycerate (MG), a compatible solute typical of marine microorganisms thriving in hot environments, is highly effective in protecting a variety of model proteins against thermal denaturation and aggregation *in vitro*. *Saccharomyces cerevisiae* cells expressing eGFP-tagged α-Syn, were further engineered to synthesize MG. The number of cells with fluorescent foci was assessed by fluorescence microscopy. Fluorescence spectroscopy and transmission electron microscopy were used to monitor fibril formation *in vitro*. We observed a 3.3-fold reduction in the number of cells with α-Syn foci and mild attenuation of α-Syn-induced toxicity. Accordingly, sucrose gradient analysis confirmed a clear reduction in the size-range of α-Syn species in the cells. MG did not affect the expression levels of α-Syn or its degradation rate. Moreover, MG did not induce molecular chaperones (Hsp104, Hsp70 and Hsp40), suggesting the implication of other mechanisms for α-Syn stabilization. MG also inhibited α-Syn fibrillation *in vitro.* MG acts as a chemical chaperone and the stabilization mechanism involves direct solute/protein interactions. This is the first demonstration of the anti-aggregating ability of MG in the intracellular milieu. The work shows that MG is a good candidate to inspire the development of new drugs for protein-misfolding diseases.

This chapter is published in:

C. Faria [(a)], C.D. Jorge [(a)], N. Borges, S. Tenreiro, T.F. Outeiro, H. Santos, Inhibition of formation of α-synuclein inclusions by mannosylglycerate in a yeast model of Parkinson's disease., Biochim. Biophys. Acta. 1830 (2013) 4065–72. doi:10.1016/j.bbagen.2013.04.015.

[(a)] these authors contributed equally

2.1 Introduction

Protein misfolding and aggregation is a key characteristic in many neurodegenerative diseases afflicting modern society. In particular, severely debilitating disorders such as Alzheimer´s, Huntington´s, Parkinson´s and prion diseases are associated with protein misfolding and aggregation in the brain. Given the devastating social impact of these diseases there is a pressing need to find drugs that prevent protein misfolding and arrest or reverse the aggregation process. Parkinson´s disease (PD) is the most prevalent neurodegenerative movement disorder, affecting 1-2% of the population over the age of 65. This currently incurable disease is characterized by the presence of neuronal inclusions known as Lewy bodies and Lewy neuritis (Trojanowski & Lee, 2003). The major component of these inclusions are fibrillar forms of α-synuclein (α-Syn), an abundant pre-synaptic protein of unknown function. *In vitro* studies showed that, under certain conditions, the natively unfolded α-Syn can self-aggregate to form oligomers, protofibrils and other intermediates that eventually give rise to mature fibrils. A considerable research effort has been directed to characterize the pathological species that trigger neuron dysfunction and degeneration, as this information would uncover targets for new therapies. However, progress has been delayed by the complexity of the fibrillation process. While more precise knowledge of the pathological pathway is lacking, preservation of the native conformation of α-Syn and/or inhibition of aggregation seem to be pertinent targets for drug development (Danzer & McLean, 2011). Small-molecules such as natural osmolytes have been shown to stabilize the functional forms of proteins, promote correct folding and inhibit aggregation, hence they are called “chemical chaperones” (Ignatova & Gierasch, 2006; Khan *et al.*, 2010; Papp & Csermely, 2006; Santos *et al.*, 2007). These molecules accumulate in the cytoplasm in response to an increase in the external osmotic pressure and their main role is to preserve cell turgor. Interestingly, they also act as stabilizers of proteins and other macromolecules. Typical osmolytes of mesophilic organisms (trehalose, glycerol, ectoine), are effective at relatively high concentrations (over 100 mM), and their mode of action involves alterations of the solvent properties (Bolen & Baskakov, 2001; Timasheff, 1992). On the other hand, the so-called pharmacological chaperones can promote proper protein folding at much lower concentration since their action involves specific binding to native proteins (Morello *et al.*, 2000). As the affinity constants are usually very high, pharmacological chaperones are effective in the μM range and are potential therapeutic agents (Bernier *et al.*, 2004). The role of natural osmolytes goes beyond the mere adjustment of the intracellular osmotic pressure and some of them are involved in the response to other stresses, namely heat stress (Santos *et al.*, 2007, 2011). In particular, marine organisms adapted to grow at temperatures near 100ºC synthesize exclusive solutes that usually bear a negative charge and are rarely found as part of stress adaptation in moderate environments. These curious ionic compounds (mannosylglycerate, di-*myo*-inositol-phosphate, diglycerol phosphate, etc), accumulate in hyper/thermophiles, presumably to protect proteins and other macromolecules against the deleterious effects of heat. They act at lower concentrations and have shown much better performance than neutral osmolytes (trehalose, glycerol) in preventing heat-induced denaturation or aggregation of model

proteins (Faria *et al.*, 2008; Santos *et al.*, 2011). Microorganisms adapted to hot environments evolved to accumulate the most effective protectors against heat damage, *i.e*., ionic solutes. Therefore, we deemed important to examine the anti-aggregating properties of these compounds in the overcrowded cytoplasm of living cells using α-Syn as a model aggregation-prone protein. Among the solutes closely associated with thermophily, mannosylglycerate (MG) stands out for its great ability to stabilize a variety of proteins and to inhibit fibril formation *in vitro* (Faria *et al.*, 2008; Santos *et al.*, 2007). For that reason, MG was selected to study the effect on the formation of α-Syn inclusions in a yeast model of PD (Outeiro & Lindquist, 2003). In this work, the *S. cerevisiae* strain VSY72 (Sancenon *et al.*, 2012), expressing eGFP-tagged α-Syn, was further engineered to synthesize MG. We demonstrated that intracellular MG directly inhibited α-Syn aggregation without affecting the expression level and the degradation rate of α-Syn, or the expression level of representative molecular chaperones. Additionally, the anti-aggregating properties of MG on α-Syn fibrillation were studied *in vitro*. The results strongly indicate that MG acts as a chemical chaperone in the intracellular milieu of yeast cells.

2.2 Materials and methods

2.2.1 Yeast strains and genetic procedures

Yeast VSY72 strain (a derivative of W303-1A *MAT*a, *ade2-1*, *can1-100*, *ura3-1*, *leu2-3,112*, *his3-11,15*, *trp1-1*), harboring two copies of human wild-type α-Syn encoding gene (*SNCA*) tagged with eGFP under the control of a galactose inducible promoter (*GAL1*) (Sancenon *et al.*, 2012), was used as the host strain for genetic manipulations and was kindly provided by Dr. Paul Muchowski (Gladstone Institute for Neurological Disease, USA). The VSY72 strain was transformed with the plasmid p425::mgsD, containing the mannosyl-3-phosphoglyceratesynthase/phosphatase gene from *Dehalococcoides mccartyi* (former *Dehalococcoides ethenogenes*) under the control of constitutive enolase 2 (*ENO2*) promoter (Empadinhas *et al.*, 2004). This plasmid was kindly provided by M. S. da Costa (Coimbra, Portugal). Standard lithium acetate heat shock procedure was used for the transformation (Guthrie & Fink, 1991). Selection was carried out at 30ºC on solid synthetic complete (SC) medium without uracil, tryptophan, and leucine (Guthrie & Fink, 1991). This procedure led to the isolation of the MG-producer mutant. The empty plasmid pRS425 was introduced in the strain to yield the “Control strain”. Strains Y4791 and Y4792 harbor two copies of WTα-Syn or A53Tα-Syn, respectively, and were first described in reference (Outeiro & Lindquist, 2003). They were selected for showing toxicity associated with the expression of α-Syn. These strains were transformed with p425::mgsD or pRS425 (MG-Producer and Control, respectively). Unless stated otherwise the results in this report refer to strain VSY72.

2.2.2 Yeast cell culture

The MG-producer mutant and the Control strain were grown at 30ºC and 140 rpm in 50 ml of SC medium supplemented with 2% glucose (glucose medium), until mid-exponential phase (OD_{600} of 2). Cells were harvested (7,000 × g, 10 min, 4ºC) and re-suspended in water. The resulting cell suspensions were used to inoculate 500 ml of SC medium supplemented with 2% galactose (galactose medium), to a final OD_{600} of 0.2. After 10 h of growth, samples (350 ml) were removed, washed three times with water, and the pellets saved for the determination of organic solutes and protein levels.

2.2.3 Quantification of cells displaying α-Syn fluorescent foci

Cells were examined by fluorescence microscopy using a Zeiss AxioImager microscope (Zeiss, Oberkochen, Germany) with a 63×/1.4 objective (Zeiss Plan-APO oil) and a GFP filter. For each culture sample, a total of 300 cells were inspected for the presence of fluorescent foci. The percentage of cells with α-Syn inclusions was determined by dividing the number of cells containing at least one fluorescent focus by the total number of cells counted, and then multiplied by 100. Seven independent experiments were performed.

2.2.4 Extraction and quantification of organic solutes

Cell pellets were suspended in water and aliquots used for dry weight determination(Santos *et al.*, 2006). The remaining cell suspension was treated twice with boiling 80% ethanol; the supernatants were pooled and the ethanol removed by evaporation. Non-polar compounds were subsequently extracted with chloroform. Finally, the aqueous fractions were lyophilized, dissolved in deuterated water, and MG and trehalose levels were quantified by NMR (Santos *et al.*, 2006).

2.2.5 Western blot analysis

For each culture sample, the number of cells per ml was determined with a Neubauer chamber. Samples, containing 1.8 × 108 cells, were re-suspended in Laemmli buffer (Laemmli, 1970) and boiled for 10 min. After electrophoresis on SDS-PAGE (10%), proteins were transferred onto nitrocellulose membranes (Millipore, USA). These membranes were treated with a blocking solution (TBS-T buffer: 0.1% (v/v) Tween-20, 150 mM NaCl, 50 mM Tris-HCl, pH 7.5 plus 5% (w/v) BSA) for 1 h at room temperature, and incubated sequentially with the following primary antibodies: mouse anti-β-actin (1:600; Abcam); mouse anti-α-Syn mAb (1:5000; BD Transduction Laboratories); mouse anti-GAPDH (1:5000; Ambion, UK); mouse anti-Hsp70 (1:1000; Stressgen); rabbit anti-Hsp104 (1:1000; Stressgen), and mouse anti-Hsp40 (1:2000; Biosciences Inc). After overnight incubations at 4ºC, membranes were washed three times with TBS-T buffer and probed with the appropriate secondary antibodies (anti-mouse or anti-rabbit IgG-horseradish peroxidase, GE healthcare), for 1 h at room temperature. Signals were revealed with the ECL Plus detection

kit (Millipore, USA) and quantified using the Quantity One software (Bio-Rad, Hercules, CA). The signals were normalized with respect to that of β-actin or GAPDH.

2.2.6 Promoter shut-off studies

Cells were pre-grown in SC medium containing glucose and lacking uracil, leucine and tryptophan for 12 hours and shifted to SC medium containing galactose and lacking uracil, leucine and tryptophan for 10 h to induce α-Syn expression. Cells were then washed twice with water and shifted to SC medium supplemented with glucose and lacking uracil, leucine and tryptophan to shut off the promoter. Cells were harvested at 0 and 6 h, centrifuged (3000 rpm, 30ºC, 4 min) and washed with water. Then, cells were re-suspended in 100 µl of Tris-HCl buffer pH 7 containing protease and phosphatase inhibitors and stored at -20ºC. Cells were lysed with glass beads (3 cycles of 30 sec in the beadbeater and 2 min in ice) and centrifuged to remove cell debris (700 × g, 3 min, 4ºC). The supernatant was sonicated for 10 s at 10 mA and protein quantification was performed using the BCA Protein Assay Kit (Pierce, USA). Protein sample buffer (200 mM Tris-HCl pH 6.8, 6% 2-mercaptoethanol, 8% SDS, 40% glycerol, 0.4% bromophenol blue) was added to each sample and heated for 10 min at 100ºC before acrylamide gel loading. Total protein samples were run in 12% SDS-PAGE and immunoblotting was described above.

2.2.7 Assessment of α-Syn aggregation in yeast by sucrose gradient

Yeast cells were harvested by centrifugation, washed with sterile water and resuspended in spheroplasting solution (Tris pH 7.5 20 mM, MgCl2 0.5 mM, BME 50 mM, sorbitol 1.2 M and zymolyase 0.5 mg/ml) and incubated at 30ºC for 30 min. Samples were centrifuged at 800 rcf (3000 rpm) for 5 min at room temperature and the supernatant was completely removed. The cells were re-suspended in 250 µl of lyses buffer (Tris-HCl pH 7.5 100 mM, NaCl 50 mM, SDS 0.4%, Triton X-100 0.2%) with inhibitors for proteases and placed for 20 min in ice. Then cells were mechanically disrupted by forcing the solution to pass through a 25G syringe 6 times. Total protein was quantified and 1 mg was applied on a 5 to 30% sucrose gradient and centrifuged at 4°C for 16 h with a swinging bucket rotor (SW-55Ti rotor, Beckman Instruments, Co., Palo Alto, CA) in a Beckman XL8 90 S/N ultracentrifuge at 45,000 rpm. Fractions were collected, precipitated for 4 h at 4°C in trichloroacetic acid, washed in acetone and suspended in protein sample buffer (0.5 M Tris-HCl, pH 6.8, glycerol, 10% (w/v) SDS, 0.1% (w/v) bromophenol blue). Proteins were resolved by SDS-PAGE.

2.2.8 Reactive oxygen species (ROS) assay

The level of ROS was determined using the dihydrorhodamine 123 (DHR 123) staining (Madeo *et al.*, 1999). The Control strain and the MG-producer mutant were grown overnight in glucose medium and centrifuged (7,000 × g, 10 min, 4ºC). Subsequently, the cell pellets were re-

suspended in galactose medium to induce α-Syn expression and allowed to grow for 9 h. To determine the basal level of ROS, the Control strain pellet was also re-suspended in fresh glucose medium and allowed to grow for further 9 h. Aliquots of the three cultures (around 3 × 106 cells) were collected and incubated with DHR 123 dye (final concentration of 5 μg/ml) for 1 h. In the cell, DHR123 is oxidized by ROS into rhodamine 123, which emits fluorescence at 535 nm when excited at 485 nm. The fluorescence intensity of the samples was measured with a Varian Cary Eclipse spectrofluorimeter (Varian, USA).

2.2.9 Spotting experiments

MG-producers and Control strains were grown at 30ºC in glucose medium, until midexponential phase (OD_{600} of 2). Cells were transferred to galactose medium to a final OD_{600} of 0.1. After 10 h, cells were removed and spotted on plates containing YPD medium (1% yeast extract, 2% peptone and 2% glucose) and incubated at 30ºC for 2 days. Shown are 3-fold serial dilutions starting with equal number of cells.

2.2.10 In vitro assays of α-Syn fibril formation

Human wild-type untagged α-Syn was expressed and purified as previously described (Outeiro *et al.*, 2009). Lyophilized α-Syn was dissolved in 20 mM Tris-HCl (pH 6.5) and aggregated materials were removed with Amicon filters (cut-off = 100 kDa). The protein concentration was determined with the BCA Protein Assay kit (Pierce). MG was purified from *Rhodothermus marinus* biomass as previously reported (Silva *et al.*, 1999). The influence of this compound on α-Syn fibril formation was assessed in reaction mixtures (250 μl), containing: 20 mM Tris-HCl (pH 6.5), 200 μM α-Syn and 100 mM MG. Each reaction mixture was incubated at 37ºC under constant agitation (650 rpm), using 0.5 ml siliconized eppendorf tubes. The fibrillation process was monitored with thioflavin T (Sigma-Aldrich, St. Louis, MO) fluorescence: 5 μl aliquots were removed at various time points from the incubated samples and added to 1 ml of 10 μM thioflavin T (ThioT) in 50 mM Na-glycine (pH 8.2). The fluorescence intensity of the samples was measured at an excitation wavelength of 450 nm and at an emission wavelength of 485 nm with a Varian Cary Eclipse spectrofluorimeter, using 1-cm light-path quartz cuvettes with both excitation and emission bandwidths of 10 nm. Experiments were run at least in duplicate and averaged.

2.2.11 Transmission Electron Microscopy (TEM)

Samples of α-Syn fibrils formed with and without MG were placed onto glow discarded carbon grids and allowed to stand for 1 min before removing the excess solution. The grid was washed once with distilled water, and once with 1% uranyl acetate before the sample was stained with fresh 1% uranyl acetate for an additional 2 min. The preparations were examined in a JEOL-1200

EX electron microscope. The grids were thoroughly examined to obtain an overall evaluation of the structures present in the sample.

2.3 Results

2.3.1 Mannosylglycerate reduces α-Syn inclusion formation in yeast

The inhibitory effect of MG on the aggregation of several model proteins has been demonstrated *in vitro*, yet it remains unclear whether this feature prevails in the overcrowded milieu of living cells (Faria *et al.*, 2008). A yeast model of PD was used to investigate the effect of MG on α-Syn inclusion formation. The yeast strain VSY72, expressing α-Syn fused to eGFP under the control of a galactose inducible promoter (Sancenon *et al.*, 2012), was used as the host strain to construct two mutants: the MG-producer harbors the pRS425 containing the gene encoding the synthase and phosphatase activities implicated in MG synthesis; the Control strain carries the empty plasmid. The intracellular content of MG was determined by proton NMR analysis of ethanol/chloroform extracts of the Control strain and the MG-producer mutant. The extracts were derived from cells collected after 10 h of induction with galactose. MG was not detected in the Control strain, while the MG-producer contained 80.3 ± 11.6 μmol MG/g dry weight (Figure 2.1). Using 2.38 ml/g dry weight for the intracellular volume of *S. cerevisiae* (Theobald *et al.*, 1997), the intracellular concentration of MG is approximately 33 mM. Trehalose, a compatible solute accumulated naturally by *S. cerevisiae* (Wiemken, 1990), was also detected (Figure 2.1).

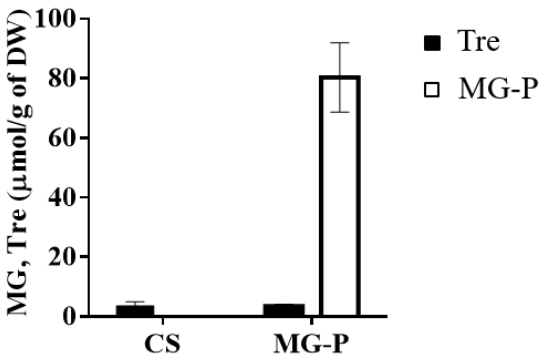

Figure 2.1 Mannosylglycerate and trehalose accumulation in yeast cells (strain VSY72). The Control (CS) and MG-producer (MG-P) strains were grown in glucose medium until late-exponential phase and then both cultures were switched to galactose medium for 10 h. Intracellular levels of trehalose (Tre, black bars) and MG (white bars) accumulated by the Control and MG-producer strains were determined by NMR in cell extracts. Data are shown as mean ± S.D from seven independent experiments.

It was confirmed that trehalose accumulated to the same low level in the two strains (circa 3.6 μmol/g dry weight or 1.5 mM). Therefore, the reduction of the α-Syn-eGFP fluorescent foci (see below) is associated with MG accumulation. As described, over-expression of α-Syn-eGFP in

yeast cells leads to the formation of cytosolic fluorescent foci that stained with thioflavin S (Outeiro & Lindquist, 2003). Moreover, it is well established that the eGFP tag does not interfere with α-Syn inclusion formation in yeast cells (Outeiro & Lindquist, 2003; Zabrocki *et al.*, 2005). Here, we found that the highest percentage of cells displaying α SyneGFP foci in the Control strain was observed at 10 h of growth in galactose medium. We confirmed that the Control strain and the MG-producer displayed identical growth profiles (Figure 2.2).

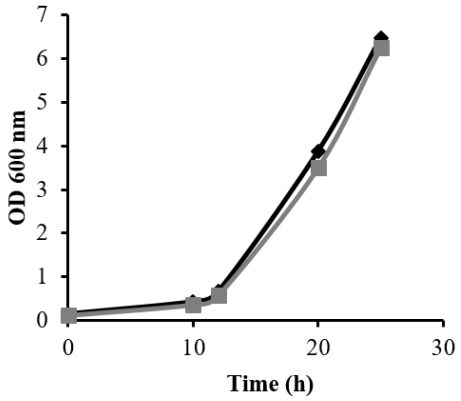

Figure 2.2 Mannosylglycerate production does not affect yeast growth. Growth curve of the Control strain (black diamonds) and the MG-producer strain (grey squares) in galactose medium. Cells used as inoculums were late-exponential phase cells cultivated in glucose medium that at time zero were transferred to galactose medium to induce α-Syn expression. Values are representative from three independent experiments.

Western blot analysis showed no significant difference in the expressing levels of α-Syn between the Control strain and the MG-producer mutant (Figure 2.3A and B) and MG accumulation caused a significant decrease in the percentage of cells with α-Syn-eGFP foci, from 39.8 ± 4.6 % in the Control strain to 11.7±3.2 % in the MG-producer (Figure 2.3C and D). To interpret this result we set to evaluate the α-Syn oligomeric species size formed in both strains. For that, we used a sucrose gradient to separate the oligomeric species by molecular weight. In agreement with what was observed by fluorescence microscopy, in the Control strain α-Syn forms oligomeric species of higher molecular weight than in the MG-producer strain (Figure 2.3E and F). Indeed, the most prevalent α-Syn oligomeric species are of smaller mass when yeast cells accumulate MG, in comparison with the Control strain.

2.3.2 MG reduces the accumulation of reactive oxygen species and slightly alleviates α-Syn induced toxicity

It is known that overproduction of α-Syn leads to the formation of reactive oxygen species (ROS), such as hydrogen peroxide and superoxide in a variety of cell systems, including yeast (Flower *et al.*, 2005). Hence, we examined the effect of MG accumulation on the generation of ROS using DHR 123 staining. The Control and MG-producer strains were grown as described above, first in glucose medium and then in galactose medium to induce high expression levels of α-Syn. The

basal levels of ROS were determined in cells without α-Syn expression, *i.e.,* in cells of the Control strain grown in glucose medium. As expected, the ROS level in the Control strain expressing α-Syn was higher (more than 3-fold), than that in control cells without induction of α-Syn expression. There was a 15% reduction of the ROS level in MG-producing cells (Figure 2.4A). As the expression of α-Syn in strain VSY72 did not induce significant levels of toxicity (Figure 2.4B), we used the strains Y4791 and Y4792, known to exhibit α-Syn-induced toxicity (Outeiro & Lindquist, 2003) to assess whether MG was able to improve cell viability. The two strains were transformed with the p425::mgsD, to induce the accumulation of MG, and with the empty vector to serve as control. Cell viability was evaluated by spotting assays to show that the accumulation of MG decreased α-Syn cytotoxicity (Figure 2.4C and D). Importantly, when MG was produced, we observed a 1.4 and 1.6 -fold reduction in the percentage of cells with α-Syn-eGFP foci in Y4791 and Y4792 strains, respectively (Figure 2.5).

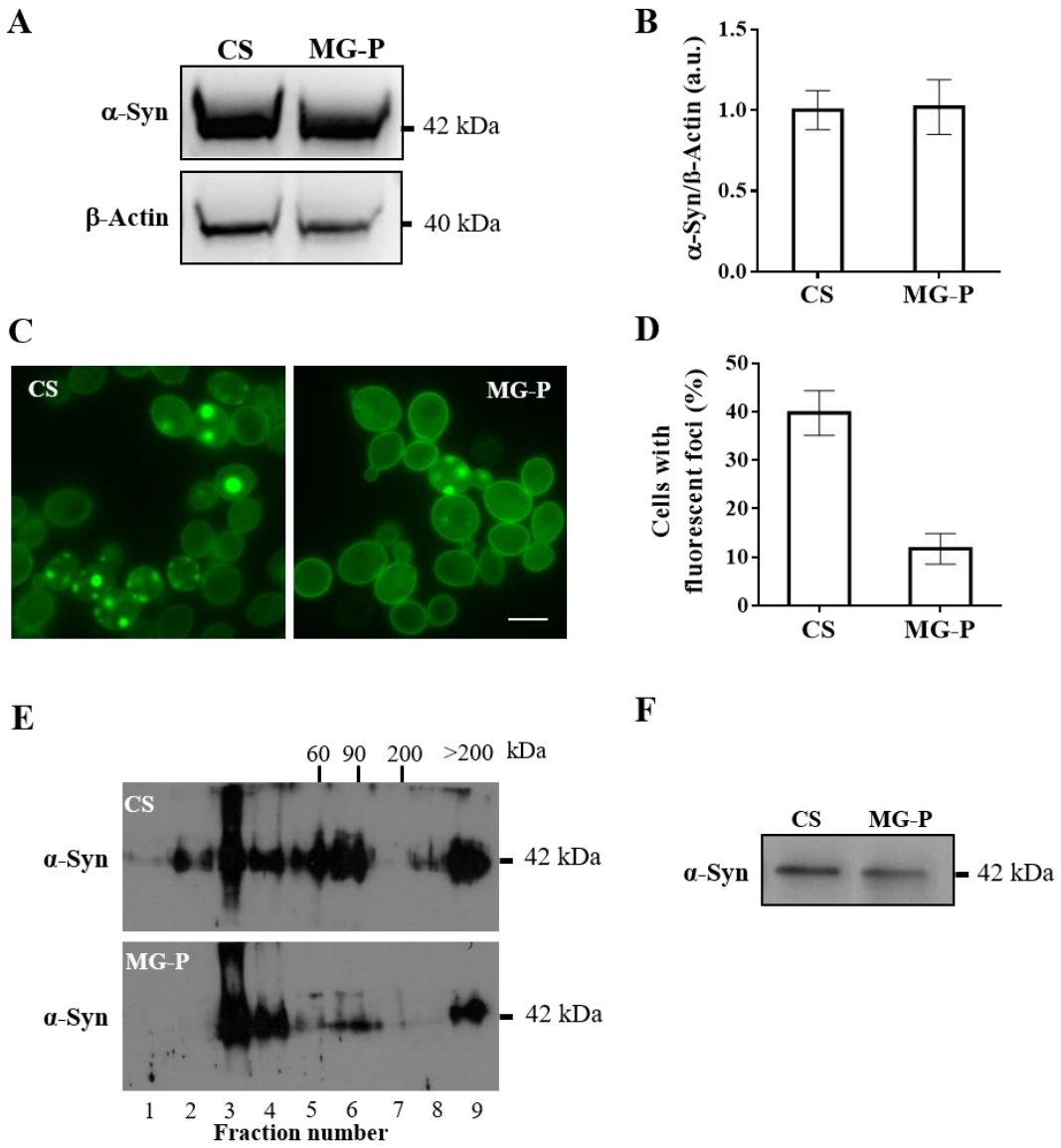

Figure 2.3 Mannosylglycerate prevents the formation of α-Syn fluorescent foci in yeast (strain VSY72). The Control (CS) and MG-producer (MG-P) yeast strains were grown as described in Figure 2.1. (A) Western blot

analysis of total cell lysates from the control and MG-producer strains revealing the endogenous levels of α-Syn and β-Actin. The mass of α-Syn-eGFP is about 42 kDa (B) Densitometric analysis of the immunodetection of α-Syn relative to the intensity obtained with a specific antibody for β-actin used as loading control, of at least three independent experiments represented in (A). (C) Representative images of yeast cells from the control and MG-producer strains exhibiting α-Syn-eGFP fluorescent foci. Scale bar: 5 μm. (D) Percentage of yeast cells (Control and MG-producer strains) containing α-Syn fluorescent foci. For each experiment a total of 300 cells were counted. Data are shown as mean ± SD from seven independent experiments. (E) The α-Syn oligomeric species formed by control and MG-producer strains, resolved on sucrose gradient. The collected fractions were applied to a SDS-PAGE followed by immunoblot with an antibody against α-Syn. (F) Western blot analysis showing that the same amount of total protein (approx. 1 mg) of the Control and MG-producer strains was applied on the sucrose gradient. Results are from one representative experiment from at least three independent experiments

2.3.3 Hsp40, Hsp70 and Hsp104 levels are not affected by MG accumulation

Small molecules such as geldanamycin, and its analogue 17-AAG, prevent α-Syn aggregation through up-regulation of molecular chaperones (McLean *et al.*, 2002; Neckers, 2002; Putcha *et al.*, 2010; Riedel *et al.*, 2010). To verify whether the effect of MG resulted from the induction of molecular chaperones, immunoblot analysis was performed. The assays revealed similar expression levels of Hsp104, Hsp70 and Hsp40 in the Control and MG-producer strains. Thus, there is no evidence for MG affecting the expression of molecular chaperones (Figure 2.6A and B).

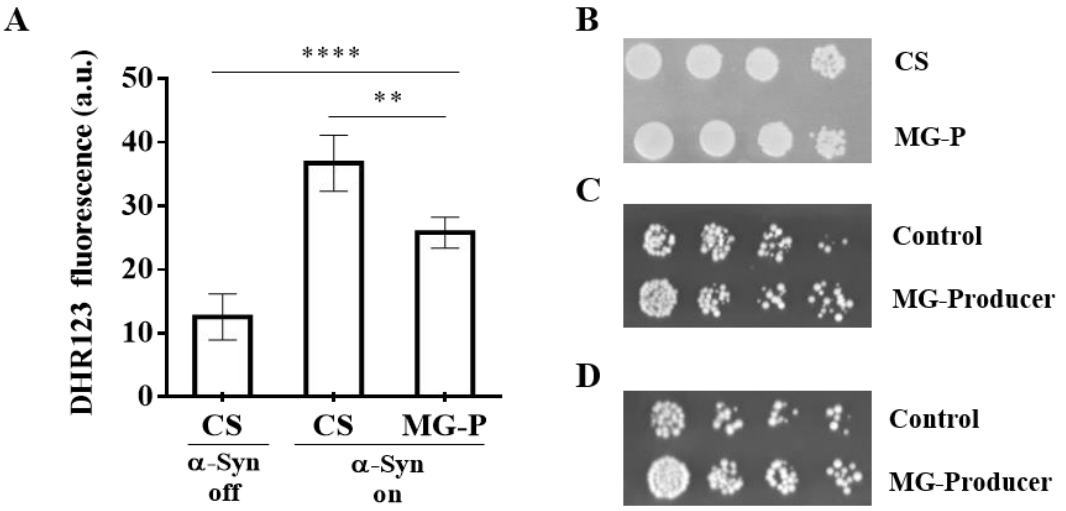

Figure 2.4 Mannosylglycerate accumulation reduces ROS levels in strain VSY72 and increases cell viability of the two strains displaying high α-Syn-associated toxicity. (A) ROS levels measured in strain VSY72. The Control (CS) and MG-producer (MG-P) strains were grown overnight in glucose medium, and then cultures were switched to galactose medium and growth continued for 9 h. The basal levels of ROS were determined in cells from the Control strain that were grown overnight in glucose medium and then re-suspended in fresh glucose medium for further 9 h. Intracellular ROS levels were determined in these strains by fluorescence using the DHR 123 staining method, as described in the "Material and methods". Data from three independent experiments are shown as mean ± S.D. (Significance of the data was determined by one-way ANOVA with Tukey's Multiple Comparison Test; **p-value<0.01; ****p-value<0.0001). (B) Cell viability assessment of the CS and MG-P strains. Cells were grown in glucose medium until mid-exponential (OD600 of 2) to accumulate MG and transferred to galactose medium to induce α-Syn expression. After 10 h, cells were removed and spotted on plates containing YPD medium and incubated at 30oC for 2 days. Shown are 3-fold serial dilutions starting with equal number of cells. (C) Cell viability assessment of the strain Y4791 that harbors two copies of wild-type α-Syn transformed with p425::mgsD (MG-producer) and with the empty vector pRS425 (Control).

(D) Cell viability assessment of strain Y4792 that harbors two copies of A53T α-Syn transformed with p425::mgsD (MG-producer) and with the empty vector pRS425 (Control).

2.3.4 MG does not interfere with α-Syn degradation

It has been reported that small compounds, namely trehalose, indirectly inhibit the aggregation of proteins implicated in neurodegenerative diseases by induction of degradation pathways (Aguib *et al.*, 2009; Davies *et al.*, 2006; Riedel *et al.*, 2010; Sarkar *et al.*, 2007). In view of these findings, it was important to verify whether the observed reduction of α-Syn inclusion formation in the MG-producer was associated with a reduction in the expression level of α-Syn and/or acceleration of protein degradation pathways. Therefore, we set out to determine these whether any of these mechanisms were involved in the effect on α-Syn inclusion formation.

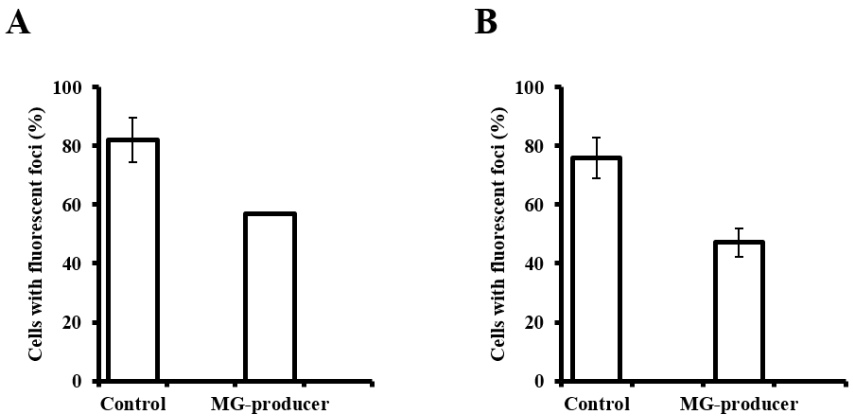

Figure 2.5 Mannosylglycerate prevents the formation of α-Syn fluorescent foci in yeast. (A) Percentage of yeast cells (Control and MG-producer strains) containing Wild-type α-Syn fluorescent foci (strain Y4791). (B) Percentage of yeast cells (Control and MG-producer strains) containing A53T α-Syn fluorescent foci (strain Y4792). For each experiment 200 cells were counted. Data are shown as mean±SD from four independent experiments

For this, the Control and the MG-producer strains were grown overnight in glucose medium to promote the synthesis of MG; then, cultures were shifted to galactose medium to induce the expression of α-Syn. At time 10 h, cells were inoculated in glucose medium to shut off the *GAL1* promoter and repress the synthesis of α-Syn. The level of α-Syn was monitored over time by immunoblot analysis. A significant decrease in the level of α-Syn was observed at 6 h after promoter repression, however no differences were found on α-Syn clearance between the two strains (Figure 2.6C and D). In summary, the results show that MG inhibits α-Syn inclusion formation, without affecting the protein expression levels or the rate of degradation.

2.3.5 Effect of MG on the kinetics of α-Syn fibril formation in vitro

The results reported above show that MG reduces α-Syn inclusion formation in the cytosol of yeast cells. In view of these findings, we enquired whether this was a direct effect on the aggregation

process, and conducted *in vitro* experiments to assess the effect of MG on α-Syn fibrillation. The kinetics of amyloid fibril formation was studied by thioflavin T fluorescence and fibrils were visualized by transmission electron microscopy. α-Syn (200 μM) was incubated in the absence and presence of 100 mM MG. Potassium chloride (100 mM) and glycerol (250 mM) were used as controls for ionic strength and viscosity, respectively. In all cases, the fibrillation of α-Syn followed typical sigmoidal curves that are consistent with a nucleation-dependent polymerization model comprising a lag phase, an exponential growth phase and a final plateau (Figure 2.7A). The lag and exponential phases of α-Syn fibrillation were not affected by the presence of MG or glycerol, while these phases were extended with KCl. The solute MG led to a clear decrease in the final equilibrium level as compared to no-solute, KCl or glycerol (Figure 2.7A). Transmission electron microscopy confirmed the presence of α-Syn fibrils and enabled us to infer whether the inhibitory effect of MG upon protein aggregation influences the amount of fibrils and their morphology (Figure 2.7B). After 96 h incubation of pure α-Syn at 37ºC, under constant agitation, the fibrils formed in the presence of MG were short in size, appeared in low amounts and in individual structures. In contrast, clusters of long fibrils (longer than 200 nm), were observed under the other conditions examined (Figure 2.7B). Therefore, it is clear that MG inhibits the formation of α-Syn fibrils in vitro, in agreement with the observations in living yeast cells.

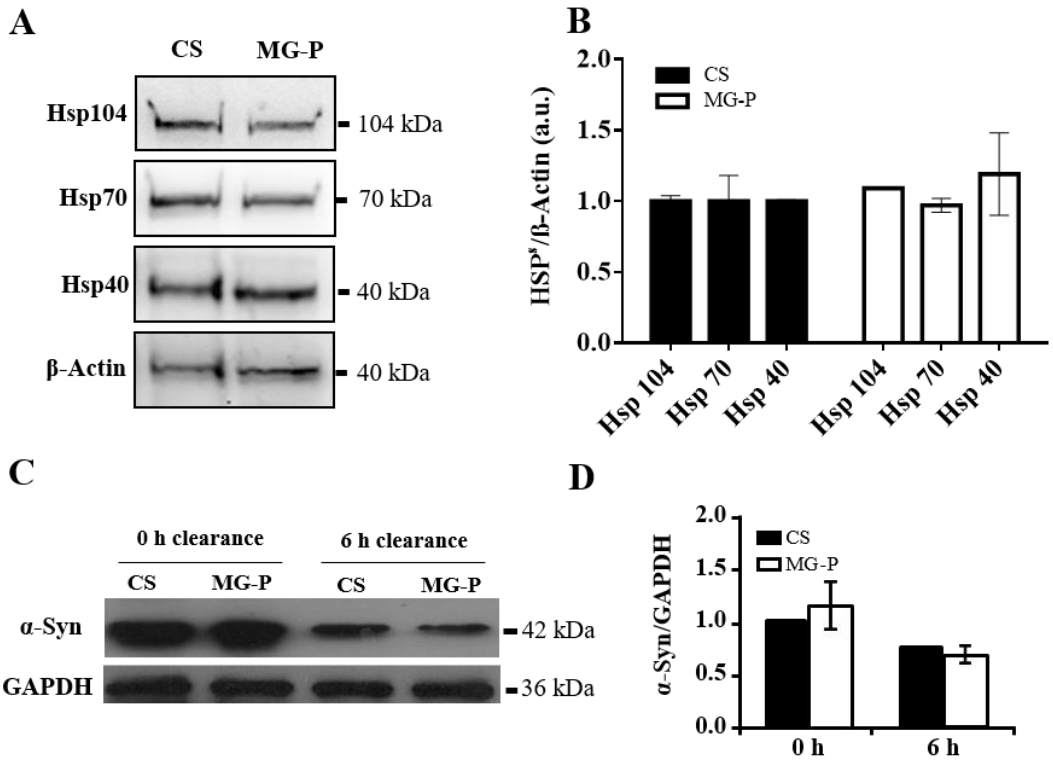

Figure 2.6 Mannosylglycerate does not induce expression of molecular chaperones or α-Syn degradation mechanism in yeast. The Control (CS) and MG-producer (MG-P) strains were grown as described in Figure 2.1. (A) Representative western blots of total cell lysates from both strain showing the endogenous levels of Hsp104, Hsp70, Hsp40 and β-actin. (B) Densitometry analysis of the immunodetection of the indicated Hsp relative to the intensity obtained with a specific antibody for β-actin, used as loading control at least three independent experiments represented in (A). (C) Western blot analysis of protein total extracts of yeast cells after 6 h of α-Syn clearance; GAPDH was used as a loading control. Results shown are from one

representative experiment from at least four independent experiments. (D) Densitometry analysis of the immunodetection of α-Syn relative to the intensity obtained with a specific antibody for GAPDH.

2.4 Discussion

In this study, we explored the potential of a simple cellular model to assess the efficacy of MG as an anti-aggregating agent of α-Syn. A yeast strain expressing α-Syn was further manipulated to accumulate MG. It was shown that MG significantly decreased α-Syn inclusion formation and this reduction did not arise from an alteration in the expression level of α-Syn or upregulation of protein degradation pathways. Likewise, MG did not affect the expression of the molecular chaperones Hsp104, Hsp70 and Hsp40, which are generally activated as part of the cellular defense mechanism against protein misfolding/aggregation (Lo Bianco *et al.*, 2008; McLean *et al.*, 2004). Additionally, the reduction in the ROS levels induced by MG synthesis further supports the conclusion that this compound inhibits α-Syn inclusion formation. Indeed, it is known that the expression of this aggregation-prone protein in yeast, as well as in higher eukaryotes, causes the formation of ROS (Flower *et al.*, 2005; Witt & Flower, 2006), and our data showed over 3-fold increase in ROS upon expression of α-Syn in the Control strain. Therefore, the lower level of oxidative stress observed in the MG-producer correlates with the decreased extent of α-Syn inclusion formation.

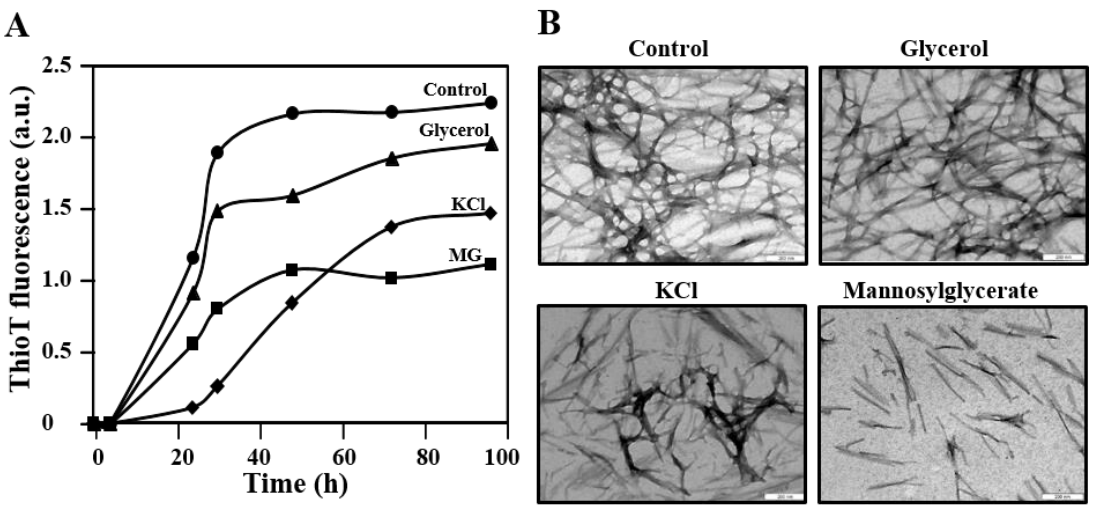

Figure 2.7 Mannosylglycerate prevents α-Syn fibril formation in vitro. (A) Fibrillation of α-Syn monitored with fluorescence spectroscopy using ThioT. The reaction mixtures containing 200 μM α-Syn and 20 mM Tris-HCl (pH 6.5) buffer were incubated at 37ºC for the indicated times in the absence (control, circles) and presence of 100 mM MG (squares), 100 mM KCl (diamonds) or 250 mM glycerol (triangles). Each point represents the mean of three independent experiments. (B) TEM micrographs of α-Syn fibrils grown during 96 h in the absence (Control) and presence of 100 mM MG, 100 mM KCl or 250 mM glycerol. White bars indicate a length of 200 nm.

Moreover, using model strains in which α-Syn expression results in high levels of toxicity (Y4791 and Y4792) (Outeiro & Lindquist, 2003), we showed that MG accumulation also led to mitigation

of cytotoxicity (Figure 2.4C and D). These results provide further evidence in support of a correlation between α-Syn inclusions and cytotoxicity (Outeiro & Lindquist, 2003). All these findings, together with the observation that MG was able to reduce considerably the α-Syn oligomerization *in vivo* and the fibrillation of pure α-Syn, lead to the conclusion that MG directly interferes with the process of inclusion formation of α-Syn in this yeast model of PD. To our knowledge this is the first demonstration that this ionic solute, closely associated with stress adaptation in hyper/thermophiles, acts as a potent chemical chaperone *in vivo*, preventing misfolding/aggregation of α-Syn. The ability of canonical osmolytes (amino acids, polyols, methylamines), to assist protein folding and reduce aggregation *in vitro* has been illustrated in numerous studies (Borwankar *et al.*, 2011; Ignatova & Gierasch, 2006; Kanapathipillai *et al.*, 2008; Schein, 1990; Uversky *et al.*, 2001), but demonstrations of these protecting properties in living cells are scarce. In microbial cell models the accumulation of the target osmolyte is generally achieved through uptake of exogenously-provided osmolytes while cells are challenged with an osmotic shock. This approach has been used to load *E. coli* cells with proline and glycine betaine, but has the drawback of inducing a concomitant osmotic shock response that can obscure the interpretation of results (Borwankar *et al.*, 2011; Ignatova & Gierasch, 2006). Obviously, this kind of strategy is not applicable to exclusive compounds such as MG, which accumulates via *de novo* synthesis and for which no uptake systems have been identified (Santos *et al.*, 2011). Therefore, the yeast model constructed in this work had to include the heterologous expression of the specific MG biosynthetic genes. The chemical chaperone activity of trehalose, the most common osmolyte in mesophilic organisms, has been proposed to explain the reduction of aggregation of mutant protein kinase C in neuronal cells (Seki *et al.*, 2010). Aggregation of this protein causes spinocerebellar ataxia type 14 (Chen *et al.*, 2003). However, most studies have shown that trehalose reduces protein aggregation indirectly, via induction of autophagy, thereby enhancing the clearance of protein aggregates (Aguib *et al.*, 2009; Sarkar *et al.*, 2007). This mode of action has been proposed for trehalose in several neuronal cell models overexpressing mutants of huntingtin, α-Syn and prion proteins (Aguib *et al.*, 2009; Sarkar *et al.*, 2007). On the other hand, induction of proteosome activity by trehalose has been reported in a mouse model of oculopharyngeal muscular dystrophy, a disease linked to aggregation of poly(A)- binding protein nuclear 1 (Davies *et al.*, 2006). The chaperone activity of trehalose has been amply demonstrated *in vitro* with a variety of proteins (Arora *et al.*, 2004; Singer & Lindquist, 1998). Moreover, the thermoprotecting effect of trehalose in yeast was demonstrated in an elegant study (Singer & Lindquist, 1998). However, these stabilizing properties require high concentrations of this disaccharide (over 0.5 M). At the low intracellular concentrations attained in animal cell models of diseases (up to 20 mM), trehalose reduces protein aggregation indirectly, via activation of the degradation processes and the chaperone activity is negligible (Sarkar *et al.*, 2007). Moreover, a genome-wide screen performed to identify genes that when deleted enhance α-Syn toxicity in yeast model of PD similar to that used in this study, has identified *TSL1* gene (encoding the α-subunit of trehalose 6-phosphate synthase) as having a role in alleviating α-Syn cytotoxicity (Willingham *et al.*, 2003). Trimethylamine-*N*-oxide and glycerol were reported to interfere directly

with aggregation of prion protein in neuronal cells, but unfortunately the authors did not verify whether these compounds affected proteolytic degradation systems and/or molecular chaperone levels (Tatzelt *et al.*, 1996). On the other hand, data on the anti-aggregating effect of proline in an *E. coli* model (over 400 mM intracellular concentration), convincingly point to a chemical chaperone mode of action of this osmolyte (Borwankar *et al.*, 2011). Understanding the molecular mechanisms underlying the stabilization of protein structure by osmolytes has attracted considerable attention not only because of the clinical relevance of finding new therapies for protein misfolding/aggregation diseases, but also because of the biotechnological importance of protein stabilization. Timasheff and coworkers showed that, upon unfolding, stabilizing osmolytes are preferentially excluded from the protein surface and the protecting effect arises from the greater destabilization of the denatured state over the native state (Timasheff, 1992). More recently, Bolen and co-workers proposed the transfer free energy model, and concluded that the unfavorable interaction of osmolytes with the protein backbone is the major driving force in protein stabilization, with the final stabilizing outcome depending on the balance of osmolyte/backbone and amino acid sidechain/solvent interactions (Bolen & Baskakov, 2001; Liu & Bolen, 1995). However, many details remain poorly understood and the picture is even less clear for ionic organic solutes such as MG. Recently, a direct link between MG-induced stabilization and protein rigidification was established for the model protein staphylococcal nuclease. It was shown that MG exerts a generalized reduction of very fast backbone motions (ps-ns timescale), but not of side chain motions; importantly, there is specific restriction of slow motions of the β-sheet residues, denoting some degree of specificity with respect to secondary structural elements of the protein (Pais *et al.*, 2009, 2012). It is generally accepted that the formation of oligomers with β-sheet is a hallmark in α-Syn amyloidogenesis (Cerdà-Costa *et al.*, 2007; Pellarin & Caflisch, 2006). Therefore, it is tempting to speculate that MG might hinder the process of α-Syn fibrillation through rigidification and stabilization of β-sheet oligomers. Further studies with other proteins will be needed to substantiate this hypothesis. The results reported here show that MG is a potent inhibitor of protein aggregation in the crowded environment inside yeast cells; the observed reduction of 30% in the number of cells exhibiting fluorescent foci is well above that reported for trehalose in cellular models of neurodegenerative diseases (Sarkar *et al.*, 2007; Seki *et al.*, 2010). This chaperone activity of MG is even more notable if one considers its relatively low concentration of this osmolyte in the cell (33 mM). Thus, the high specific activity of MG is a promising feature for the development of new MG-inspired drugs to combat protein-misfolding diseases. In conclusion, we demonstrated the efficacy of MG in preventing protein aggregation in the context of living yeast cells. The yeast model can be easily modified to test the effect of MG in the protection of other aggregation-prone proteins, thereby determining how general the *in vivo* MG-stabilizing ability is. This work also reinforces the view that MG plays a physiological role in the stabilization of proteins in the natural host organisms, which thrive optimally in hot environments where the deleterious effects of heat on macromolecule structures have to be offset efficiently.

Chapter 3

Mannosylglycerate production in *Saccharomyces cerevisiae* by over-expression of the GDP-mannose pathway and bioprocess optimization

Mannosylglycerate (MG) is one of the most widespread compatible solutes among marine microorganisms adapted to hot environments. This ionic solute holds excellent ability to protect proteins against thermal denaturation, hence a large number of biotechnological and clinical applications have been put forward. However, the current prohibitive production costs impose severe constraints towards large-scale applications. Most native producers synthesize MG from GDP-mannose and 3-phosphoglycerate via a two-step pathway in which mannosyl-3-phosphoglycerate is the intermediate metabolite. In an early exploratory work, this pathway was expressed in *Saccharamyces cerevisiae* by Empadinhas *et al.* (2004), but the level of accumulation was unsatisfactory. Therefore, we decided to invest further effort and convert *S. cerevisiae* into an efficient cell factory for MG production. To this end, the pathway for synthesis of GDP-mannose, one of the MG biosynthetic precursors, was overexpressed in *S. cerevisiae* along with the MG biosynthetic pathway. Moreover, MG production was evaluated under different cultivation modes, *i.e*., flask bottle, batch, and continuous mode with different dilution rates. Using the MG-producing yeast *S. cerevisiae* (MG01) as background host, the genes encoding mannose-6-phosphate isomerase and GDP-mannose pyrophosphorylase were cloned. The resulting engineered strain (MG02) showed around a 2-fold increase in the activity of these enzymes in comparison to strain MG01. In batch mode, strain MG02 accumulated 15.86 $mg_{MG}.g_{DW}^{-1}$, representing a 2.2-fold increase relative to the reference strain (MG01). In continuous culture, at a dilution rate of 0.15 h^{-1}, there was a 1.4-fold improvement in productivity (from 1.29 to 1.79 $mg_{MG}.g_{DW}^{-1}.h^{-1}$).

C. Faria, N. Borges, I. Rocha[(a)], H. Santos[(a)]. Mannosylglycerate production in *Saccharomyces cerevisiae* by over-expression of the GDP-mannose pathway and bioprocess optimization. (2017). Manuscript in preparation.

[(a)] these authors contributed equally

3.1 Introduction

Enzymes and other proteins are used in a myriad of applications such as clinical and analytical tests, research, therapeutics, vaccines, food processing, textile industry, cleaning, or biofuel production. In every case, the preservation of the native structure under working conditions is a prerequisite for efficacy. While some proteins are remarkably robust and can stand harsh conditions or repeated work cycles, others require the presence of extrinsic stabilizers, or chemical chaperones, to prevent unfolding and/or assist refolding. Osmolytes, like glycerol and trehalose, which accumulate inside the cell to counterbalance the external osmotic pressure, are well known protein protectors (da Costa *et al.*, 1998; Elbein, 2003).

The discovery of marine hyperthermophilic organisms in the early 1980´s uncovered an untapped source of new osmolytes (Santos & da Costa, 2002). Microorganisms adapted to those extreme environments accumulate exquisite organic solutes, usually bearing a negative charge, which are absent from mesophilic prokaryotes. Importantly, these ionic solutes accumulate not only in response to elevated osmotic pressure, but also in response to supraoptimal temperature, suggesting a potential role in thermoprotection of cell components *in vivo* (Lamosa *et al.*, 2007).

Mannosylglycerate and di-*myo*-inositol-phosphate are the most widespread solutes among marine hyper/thermophiles (da Costa *et al.*, 1998; Lamosa *et al.*, 2007; Santos *et al.*, 1998; Santos & da Costa, 2002). *In vitro* studies have shown that these compounds are better protein stabilizers than trehalose, glycerol or any other known neutral solutes (Borges *et al.*, 2002; Lamosa *et al.*, 2000; Longo *et al.*, 2009; Ramos *et al.*, 1997; Santos *et al.*, 2011; Shima *et al.*, 1998). For example, 0.5 M of MG resulted in a 10.5ºC increase in the melting temperature of the model protein, staphylococcal nuclease, while the same concentration of glycerol or mannosylglyceramide produced increments of < 1ºC and 2.7 ºC, respectively (Faria *et al.*, 2008). Importantly, MG was able to inhibit formation of α-synuclein inclusions in the cytoplasm of yeast cells, a hint on its potential application in drug development against protein misfolding diseases (Faria *et al.*, 2013, 2008).

All native producers, Archaea and Bacteria, examined thus far synthesize MG via a two-step pathway: the enzyme mannosyl-3-phosphoglycerate synthase catalyzes the reaction of GDP-mannose and 3-phosphoglycerate into mannosyl-3-phosphoglycerate, which is subsequently dephosphorylated by mannosyl-3-phosphoglycerate phosphatase to form MG (Martins *et al*. 1999; Borges *et al.*, 2014). Empadinhas *et al*. (2004) noticed that the genome of the mesophilic bacterium *Dehalococcoides mccartyi* (formerly *D. ethenogenes*), comprised a gene encoding two domains with high homology to known mannosyl-3-phosphoglycerate synthase and mannosyl-3-phosphoglycerate phosphatase. The same authors confirmed the functionality of this gene, designated *mgsD,* by cloning and expression in *Saccharomyces cerevisiae*, which thereby accumulated MG intracellularly up to 7.5 mg per g cell dry weight (Empadinhas *et al.*, 2004).

Capitalizing on these encouraging results, we deemed it interesting to convert *S. cerevisiae* into a cell factory for MG production. Currently, MG is produced by Bitop AG (Witten, Germany) through fermentation and “milking” of a natural producer. However, this process has high production costs,

hampering the utilization of this compound at an industrial scale. On the other hand, *S. cerevisiae* is a well-known industrial workhorse used for the production of many chemicals, such as sesquiterpenes, ethanol, artemisinic acid, succinic acid and vanillin (Asadollahi *et al.*, 2009; Brochado *et al.*, 2010; Otero *et al.*, 2013; Ro *et al.*, 2006; van Zyl *et al.*, 2007).

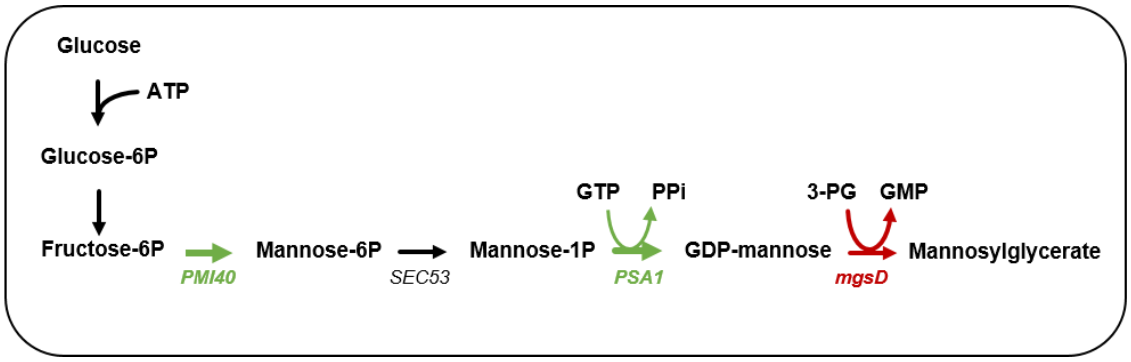

Figure 3.1 Biosynthesis of mannosylglycerate (MG) in *Saccharomyces cerevisiae* using glucose as carbon source. MG is produced from the reaction of GDP-mannose and 3-phosphoglycerate (3PG) with the release of GMP. To produce MG, a gene from *Dehalococcoides mccartyi* coding for MG synthase/phosphatase *(mgsD*) was cloned in a plasmid and transformed in *S. cerevisiae* to yield strain MG01. A second plasmid containing the genes *PMI40* and *PSA1* from *S. cerevisiae* was constructed and transformed in MG01 yielding MG02.

GDP-mannose and 3-phosphoglycerate are the two precursors for the synthesis of MG. We are aware that *S. cerevisiae's* anabolic pathways imply a high demand for GDP-mannose. Indeed, this activated sugar is the precursor for protein mannosylation and synthesis of oligomannans, important components of the yeast cell wall (Hashimoto *et al.*, 1997; Janik *et al.*, 2003). Therefore, our engineering strategy was intended to enhance the flux towards GDP-mannose synthesis by overexpressing the genes encoding phosphomannomutase and GDP-mannose pyrophosphorylase (Figure 3.1). The production of MG by the resulting engineered strain was evaluated under different cultivation modes, including flask bottles, controlled batch and continuous fermentations at different dilution rates.

3.2 Methods

3.2.1 DNA manipulation

S. cerevisiae genomic DNA for colony PCR and gene amplification was prepared as described by Lõoke *et al.*, 2011 . Primers were purchased from Metabion (Germany). Gene amplification and colony PCR were performed using Phusion polymerase and DreamTaq DNA polymerase (both enzymes from Thermoscientific, Waltham, USA), respectively. PCR reactions were performed in a Thermal Cycler from Bio-Rad (Hercules, USA) using the protocols recommended by the manufacturers for each polymerase. Plasmid extraction, PCR product purification, and DNA gel

extraction were carried out using Zymo Research kits (Irvine, USA). All restriction enzymes were obtained in fast-digestion format from Thermoscientific (Waltham, USA). The plasmid pDES containing the gene *mgsD* that codes for mannosyl-3-phosphoglycerate synthase/phosphatase from *D. mccartyi* is under the control of the *ENO2* promoter from Enolase II and was kindly provided by the group of Prof. Milton da Costa (University of Coimbra, Portugal) (Empadinhas *et al.*, 2004). *S. cerevisiae* transformation was performed using the lithium acetate method as described by Gietz & Woods, 2002 . *Escherichia coli* DH5α was used for plasmid isolation and maintenance using the competence and transformation procedures developed by Dower *et al.*, 1988 .

3.2.2 Construction of engineered strains

The plasmid pSP-GM was kindly provided by Prof. Jens Nielsen's group (Chalmers University of Technology, Sweden) (Partow *et al.*, 2010). This plasmid was used to express the mannose-6-phosphate isomerase (*PMI40*) and GDP-mannose pyrophosphorylase (*PSA1*) encoding genes from *S. cerevisiae*. The gene *PMI40* was amplified using genomic DNA isolated from *S. cerevisiae* strain CENPK2-1C with the primers pair 5'-CCGCGGCCGCAAAAAAATGTCCAACAAGCTGTTCAGG-3´ and 5´-CCGAGCTCCTAATTTGGTTCCACAAAGGC-3´. This PCR product was digested with *Sac*I/*Not*I and ligated into pSP-GM between promoter *TEF1* and terminator *ADH1* using a T4 ligase (Thermoscientific, Waltham, USA), yielding pSP01. Next, the gene *PSA1* was amplified using again *S. cerevisiae* strain CENPK2-1C genomic DNA with primers pair 5´-GGCCCGGGAAAAAAATGAAAGGTTTAATTTTAGTCGG-3´ and 5´-CCAAGCTTTCACATAATAATAGCTTCCTTTGG-3´. This PCR product was digested with *Hind*III/*Xma*I and ligated into pSP01 using a T4 ligase between promoter *PGK* and terminator *CYC1*, resulting in plasmid pSP02. Colonies harboring pSP01 and pSP02 were identified by performing a colony PCR with the primers used in the amplification. Correct constructions of plasmids were confirmed by restriction analysis and DNA sequencing (GATC Biotech, Germany). The strain *S. cerevisiae* CENPK2-1C, obtained from EUROCARF (Germany), was used as the background strain in this work. The strain MG01 was obtained by transforming CENPK2-1C cells with plasmid pDES (conferring yeast cells the ability to produce MG) and strain MG02 was obtained by transforming MG01 with plasmid pSP02 (resulting in strains with the ability to produce MG and with over-production of GDP-mannose).

3.2.3 Strains maintenance and cultivation media

Selection and maintenance of plasmids in *E. coli* was performed in LB medium containing 10 $g.L^{-1}$ of peptone, 10 $g.L^{-1}$ of NaCl, 5 $g.L^{-1}$ of yeast extract, and supplemented with 100 $mg.L^{-1}$ of ampicillin. The solid LB medium also included 20 $g.L^{-1}$ of agar. All cultivations of *E. coli* were made at 37ºC and 200 rpm of agitation.

Cells from *S. cerevisiae* strain CENPK2-1C (MATa ura3-52 his3-Δ1 leu2-3,112 trp1-289, MAL2-8c SUC2) were cultivated in YPD medium, containing 10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ of peptone, and 20 $g.L^{-1}$ of glucose, at 30ºC and 160 rpm. Recombinant strains were selected on Synthetic Dextrose (SD) medium containing 6.7 $g.L^{-1}$ of yeast nitrogen base without amino acids from Difco (New Jersey, USA), 20 $g.L^{-1}$ of glucose and 20 $g.L^{-1}$ of agar. When necessary, histidine (20 $mg.L^{-1}$), tryptophan (20 $mg.L^{-1}$), uracil (20 $mg.L^{-1}$) and leucine (30 $mg.L^{-1}$), were added to cover auxotrophies. All cultivations of *S. cerevisiae* were performed at 30 ºC with 160 rpm agitation. Strains were grown overnight and stored at -80 ºC in selective media plus 30% (w/v) of glycerol.

3.2.4 Fermentation conditions

The SD medium used for shake flasks, batch in controlled bioreactors and continuous cultivation contained 6.7 $g.L^{-1}$ of yeast nitrogen base without amino acids from Difco (New Jersey, USA) supplemented with 20 $g.L^{-1}$ of glucose. For strain MG01, SD medium was supplemented with histidine (20 $mg.L^{-1}$), tryptophan (20 $mg.L^{-1}$), and uracil (20 $mg.L^{-1}$), while for strain MG02 it was complemented only with histidine and tryptophan, in the concentrations described above (Guthrie & Fink, 1991).

In shake flask cultivations, MG01 and MG02 were pre-grown in 50 mL of SD medium at 30ºC and 160 rpm. Then, cells were inoculated with an initial OD_{600} of 0.1 in 500 mL shake flasks containing 100 mL of SD medium. Growth was followed using a spectrophotometer V-560 from Jasco (Japan); cells were harvested at the end of the exponential phase (24 h upon inoculation).

Batch cultivations were performed as following: firstly, cells were grown overnight in shake flasks from a single colony. Each fermenter was inoculated with an initial OD_{600} of 0.1. The batch fermentations were performed in a BIOSTAT-B Plus system (Sartorius, Germany) coupled with a 2-L culture vessel. The operating conditions are the following: the fermentation volume was 0.8 L, the temperature was maintained at 30ºC, the airflow was 1 vvm, the pH was kept at 5.5 controlled by addition of 2 M NaOH, and the dissolved O_2 was kept above 30% of saturation by feedback control of the stirring speed from 200 rpm until 600 rpm. Concentrations of O_2 and CO_2 in the exhaust gas were monitored by Bluesens gas analyzers (Germany). Samples were collected at the end of the exponential phase, when CO_2 production dropped drastically.

Continuous cultivations were also carried out in a Sartorius BIOSTAT-B Plus system coupled with a 2-L culture vessel with a constant working volume of 0.8 L, with the same controlled variables and set-points as before. The same medium described above was used to feed the bioreactor at dilution rates of 0.05, 0.1, 0.15 and 0.2 h^{-1}. The volume was kept at 0.8 L by controlling the level of broth inside the vessel. To guarantee that the culture reached a steady-state mode, at least 5 volumes of medium were allowed to pass through the culture whenever the dilution rate was changed. Samples were collected when O_2 consumption and CO_2 production were constant.

3.2.5 Quantification of MG and fermentation end products

For each sampling, 50 mL of culture broth were removed from the culture and centrifuged at 3600 × *g* for 10 min to separate cells from the medium. Next, the supernatant was filtered through a membrane filter with a pore size of 0.22 µm into HPLC vials and stored at -20 ºC until further analyzed for glucose, glycerol, acetate, and ethanol. MG is an intracellular compound and was extracted from the pellet centrifuged above by adding 5 mL of water, 5 mL of methanol and 10 mL of chloroform. This mixture was vigorously shaken with the help of a vortex and then centrifuged at 12,000 × *g* for 30 min at 4 ºC. The top part, corresponding to the aqueous phase, was carefully transferred to a new tube and the process of centrifugation was repeated after the addition of 5 mL of water. Once the aqueous phase was separated, samples were evaporated using a Savant™ SPD131DDA SpeedVac (Thermoscientific). Next, the evaporated samples were dissolved in 1 mL of water and transferred to an HPLC vial. Analysis was performed in an HPLC apparatus from Jasco (Japan) model LC-NetII/ADC equipped with UV-2075 Plus and RI-2031 Plus detectors, also from Jasco. The samples were analyzed using an Aminex HPX-87H column from Bio-Rad, which was kept at 50 ºC, and a solution of 0.01 M of H_2SO_4 was used as the mobile phase with a flow rate of 0.3 $mL.min^{-1}$. Quantitative analysis of glucose, MG, glycerol, acetate and ethanol was performed by comparison with a mixture of standards with known concentrations of each metabolite. Calibration curves were prepared using the peak areas of the RI detector for glucose, glycerol, acetate and ethanol and of the UV absorbance for MG. The cell dry weight was determined by filtering 5 mL of culture broth through a 0.22 µm pore Millipore filter and washing once with 5 mL of water. Filters were then dried for 10 min at 150 W in a microwave oven and weighted using an analytical balance.

3.2.6 Quantification of enzymatic activities

For the determination of mannose-6-phosphate isomerase and GDP-mannose pyrophosphorylase activities, strains MG01 and MG02 were grown in shake flasks until mid-exponential phase in SD medium supplemented with 20 $g.L^{-1}$ of glucose at 30°C and 160 rpm. Then, cells were harvested by centrifugation and re-suspended in water with a cocktail of protein inhibitors (Roche, USA). Next, cells were broken three times using a Fastprep 24 system (MP, USA). Supernatants were obtained by centrifugation at 13,500 × g for 5 min and used to determine enzymatic activities. Total protein was determined using the Bradford method (Bradford, 1976). The methods used to determine the enzymatic activities have been described by Bergmeyer & Gawehn, 1974 . Briefly, the mannose-6-phosphate isomerase activity was determined using a spectrophotometric method. The assay mixture (final volume of 1 mL) contained: 50 mM Tris-HCl (pH 7), 5 mM mannose-6-phosphate, 8 mM $MgCl_2$, 2.8 mM NADP, 0.27 U glucose-6-phosphate dehydrogenase, 0.27 U phosphoglucose isomerase, and 0.25 mg of cell extract. Reactions were started by the addition of mannose-6-phosphate and absorbance was measured at 340 nm for 2 min. The GDP-mannose pyrophosphorylase activity was determined using a spectrophotometric method. The assay

mixture (final volume of 1 mL) contained: 50 mM of Tris-HCl (pH 7), 0.5 mM NaF, 10 mM $MgCl_2$, 0.1 mM ADP, 1 mM glucose, 1 mM NADP, 2 mM PPi, 5.5 mM GDP-mannose, 1 U each of nucleoside kinase, hexokinase, and glucose-6-P dehydrogenase, and 0.25 mg of cell extract. The reaction was started by the addition of GDP-mannose, and the formation of NADPH was monitored by recording the absorbance at 340 nm for 2 min.

3.3 Results and discussion

3.3.1 Strains construction and characterization

S. cerevisiae strain CENPK2-1C was genetically manipulated to overexpress the genes involved in the synthesis of GDP-mannose and MG (Figure 3.1). Two engineered strains were constructed: i) the first strain (named MG01) harbors the plasmid containing the heterologous gene encoding the MG bifunctional enzyme (plasmid pDES); and, ii) the second strain (named MG02) harbors the plasmid pDES plus the plasmid pSP02 containing the mannose-6-phosphate isomerase (*PMI40*) and GDP-mannose pyrophosphorylase (*PSA1*), both genes from *S. cerevisiae*. The gene *PMI40* was cloned under the control of *PGK1* promoter and terminator *CYC1*, while the gene *PSA1* under the promoter *TEF1* and the terminator *ADH1*. These promoters have previously shown high expression levels during growth on glucose (Partow *et al.*, 2010). To determine the expression levels of *PSA1* and *PMI40* genes, mannose-6-phosphate isomerase and GDP-mannose pyrophosphorylase activities were measured in the MG02 strain and compared with the background strain (*S. cerevisae* strain CENPK2-1C). The MG02 strain showed an increase of 2-fold in the mannose-6-phosphate isomerase activity and 1.4-fold in the GDP-mannose pyrophosphorylase activity in comparison with the background strain (Table 3.1).

Table 3.1 Mannose-6-phosphate isomerase (*PMI40*) and GDP-mannose pyrophosphorylase (*PSA1*) activities in the background and MG02 strains.

	CENPK2-1C (U/mg of protein)	MG02 (U/mg of protein)	Increase (n-fold)
Mannose-6-phosphate isomerase	0.120 ± 0.001	0.244 ± 0.010	2.0
GDP-mannose pyrophosphorylase	0.045 ± 0.007	0.065 ± 0.014	1.4

3.3.2 Physiological characterization of engineered strains

Cultivation conditions exert a great impact in the performance of producing strains. In particular, *S. cerevisiae* is known to accumulate ethanol as a major by-product in aerobic conditions under certain circumstances, and these different metabolic states may affect the production capabilities

of the strains. When there is a low glucose uptake rate, *S. cerevisiae* is reduced to a respiratory metabolism only, with no ethanol formation while, when the glucose concentration and/or uptake rate surpasses a critical threshold value, the metabolism becomes a combination of respiration and alcoholic fermentation (Sonnleitner & Käppeli, 1986). In this work, the engineered strains were grown under conditions that determined a respiratory vs respiro-fermentative metabolism, and the fitness and MG accumulation parameters were determined.

3.3.3 Physiological characterization of engineered strains cultivated in shake flasks and batch bioreactors

The engineered strains were initially characterized in shake flasks using SD medium with 20 $g.L^{-1}$ of glucose. Both strains present a similar growth rate (0.24 h^{-1} for MG01 vs 0.25 h^{-1} for MG02) but differ on biomass production: 0.134 ± 0.006 $g_{DW}.g_{glc}^{-1}$ for MG01 and 0.096 ± 0.003 $g_{DW}.g_{glc}^{-1}$ for MG02 (Figure 3.2). The intracellular amount of MG was determined after glucose depletion. The solute was extracted with methanol/chloroform solution and quantified by HPLC analysis. Under these conditions, MG02 showed a 1.5-fold increase in MG production compared to the reference strain MG01 (Table 3.2). *S. cerevisiae* GDP-mannose pool had been successfully increased before by the over-expression of the gene *MPG1* (*PSA1*) in a multi-copy plasmid (Janik *et al.*, 2003). In the present work, it was possible to increase MG production by over-expressing *PSA1* and *PMI40,* which re-directed the glycolytic flux towards the formation of GDP-mannose and consequently to MG. *S. cerevisiae* end-products were also determined by HPLC in the culture broth after glucose depletion (Table 3.2). In shake flasks, MG01 produces higher amounts of ethanol (1.5-fold), glycerol (1.4-fold), and acetate (3.6-fold), than MG02.

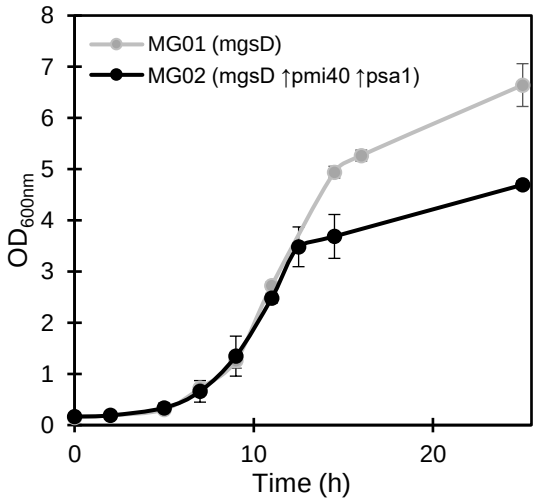

Figure 3.2 Growth curve profile of MG01 and MG02 in shake flasks. Cells were cultivated in SC medium with 20 $g.L^{-1}$ of glucose. MG01 harbors the plasmid pDES with the gene *mgsD* from *D. mccartyi* and MG02 harbors the same plasmid along with pSP-GM, which contains the *S. cerevisiae* genes *PMI40* and *PSA1*. Data represents the mean ± SD of four independent experiments.

In order to check whether MG accumulation could be favored by a controlled environment, strains MG01 and MG02 were grown in a 2-L bioreactor. In batch cultivation, MG02 produces 15.86 $mg.g_{DW}^{-1}$ of MG, which represents an increase of 2.2-fold in comparison to MG01 (Table 3.2). Also, higher yields of MG are obtained in batch cultivation in comparison with shake flask. In batch cultivation, using a bioreactor, it is possible to maintain optimal conditions regarding pH, concentration of oxygen and temperature. Under these conditions, *S. cerevisiae* is able to produce higher amounts of MG. Moreover, MG01 and MG02 produce the same amount of ethanol, unlike the behavior observed in shake flasks. Acetate is still produced in lower amounts in MG02, whether cells are cultivated in shake flasks or controlled batch (Table 3.2).

Table 3.2 MG and end-products yields for the engineered strains MG01 (mgsD) and MG02 (mgsD ↑pmi40 ↑psa1) cultivated in shake flask and batch fermenters. Data represent the mean ± SD of at least three independent experiments.

	$Y_{X/MG}$ [a]	$Y_{S/MG}$ [b]	$Y_{S/EtOH}$ [c]	$Y_{S/Acet}$ [d]	$Y_{S/Gly}$ [e]
Shake flasks					
MG01	8.082 ± 0.403	0.908 ± 0.045	0.531 ± 0.177	0.018 ± 0.011	0.013 ± 0.001
MG02	12.048 ± 1.646	1.015 ± 0.105	0.362 ± 0.102	0.005 ± 0.004	0.009 ± 0.002
Batch					
MG01	7.080 ± 1.256	0.937 ± 0.072	0.339 ± 0.032	0.058 ± 0.045	0.027 ± 0.007
MG02	15.862 ± 0.465	1.365 ± 0.084	0.385 ± 0.109	0.039 ± 0.007	0.076 ± 0.031

[a] Yield of mannosylglycerate on biomass, $mg_{MG}.g_{DW}^{-1}$
[b] Yield of mannosylglycerate on substrate, $mg_{MG}.g_{glc}^{-1}$
[c] Yield of ethanol on substrate, $g_{EtOH}.g_{glc}^{-1}$
[d] Yield of acetate on substrate, $g_{Acet}.g_{glc}^{-1}$
[e] Yield of glycerol on substrate, $g_{Gly}.g_{glc}^{-1}$

3.3.4 Physiological characterization of engineered strains cultivated in chemostat at different dilution rates

Although the relative contribution of respiratory and respiro-fermentative regimens in *S. cerevisiae* depends on the glucose availability, it is not easy to achieve a pure respiratory regimen in batch cultivations. Therefore, a clear distinction of the two metabolic states is only possible in fed-batch and chemostat cultures. In an aerobic environment and in the presence of large amounts of glucose, *S. cerevisiae* has a fermentative metabolism in addition to a respiratory metabolism due to a limited respiratory capacity (Sonnleitner & Käppeli, 1986). In addition, there is a molecular response that represses the TCA cycle in the presence of large amounts of glucose, increasing the metabolic flow towards the production of end-products with some reducing power, a phenomenon called Crabtree effect (Crabtree, 1929). This behavior happens independently of the presence of enough oxygen to support respiration and is seen as an adaptive response to a

competitive environment, as fermentation increases *S. cerevisiae's* growth rate. Nevertheless, it is possible to ensure a respiratory metabolism if glucose uptake rate is kept below the critical threshold that activates fermentation.

In an attempt to investigate the impact of the metabolic regimen in the biomass and MG accumulation yields, both MG01 and MG02 were cultivated in the continuous mode at different dilution rates: 1) a dilution rate of 0.1 h^{-1}, where glucose-limited conditions could be maintained, expressed by a steady-state glucose concentration below the HPLC detection level; and 2) a dilution rate of 0.2 h^{-1} for MG01 and 0.15 h^{-1} to MG02 where the glucose concentration in the medium upon steady-state was approximately 1 $g.L^{-1}$.

Table 3.3 Physiological parameters and MG yields for the engineered strains MG01 (mgsD) and MG02 (mgsD ↑pmi40 ↑psa1) in chemostat cultivation at different dilution rates. High-glucose experiments were performed at D = 0.2 and 0.15 h^{-1} for MG01 and MG02, respectively and glucose-limited experiments were performed at D = 0.1 h^{-1}.

	High-glucose		Glucose-limited	
	MG01	MG02	MG01	MG02
$Y_{X/MG}$ [b]	6.09 ± 1.49	11.93 ± 2.62	7.58 ± 1.21	11.71 ± 0.77
$Y_{S/MG}$ [a]	1.08 ± 0.04	1.46 ± 0.10	0.98 ± 0.02	1.82 ± 0.35
$Y_{S/EtOH}$ [a]	0.33 ± 0.06	0.33 ± 0.14	0.17 ± 0.03	0.33 ± 0.31
$Y_{S/Acet}$ [a]	0.05 ± 0.01	0.03 ± 0.01	0.05 ± 0.01	0.04 ± 0.02
$Y_{S/Gly}$ [a]	0.016 ± 0.003	0.024 ± 0.017	0.001 ± 0.001	0.013 ± 0.001
$Y_{X/S}$ [b]	0.15 ± 0.01	0.12 ± 0.02	0.13 ± 0.02	0.15 ± 0.02
$P_{X/MG}$ [c]	1.22 ± 0.29	1.79 ± 0.39	0.76 ± 0.12	1.17 ± 0.08
$P_{X/S}$ [c]	0.030 ± 0.002	0.019 ± 0.003	0.013 ± 0.002	0.015 ± 0.002
$P_{X/EtOH}$ [c]	0.41 ± 0.05	0.40 ± 0.12	0.14 ± 0.04	0.24 ± 0.17
$P_{X/Acet}$ [c]	0.02 ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.02 ± 0.01

[a] Yields on substrate for ethanol $Y_{S/EtOH}$, acetate $Y_{S/Acet}$ and glycerol $Y_{S/Gly}$ are represented as $g.g_{glc}^{-1}$ and MG yield $Y_{S/MG}$ is represented as $mg.g_{glc}^{-1}$.
[b] Yields on biomass for substrate $Y_{X/S}$ is represented as $g_{DW}.g_{glc}^{-1}$ and for MG $Y_{X/MG}$ as $mg_{MG}.g_{DW}^{-1}$.
[c] Productivities in chemostat cultivation are represented for ethanol $P_{X/EtOH}$, acetate $P_{X/Acet}$ and glycerol $P_{X/Gly}$ in $g.g_{DW}^{-1}.h^{-1}$ and for MG is represented as $mg.g_{DW}^{-1}.h^{-1}$.

On the chemostats at a higher dilution rate, MG01 showed a higher yield on biomass ($Y_{x/s}$) than MG02 (Table 3.3). It should be emphasized that MG01 supported a higher dilution rate than MG02, since there was washout over 0.15 h^{-1} for this strain, given the lower maximum specific growth rate. It can thus be concluded that either the higher production of MG or directly the over-expression of *PSA1* and *PMI40* cause a significant decrease of the biomass production. Under higher dilution rates, MG02 accumulated 11.93 $mg_{MG}.g_{DW}^{-1}$, representing a 2-fold increase in

relation to MG01 (Table 3.3). By using this set up, MG is produced at a rate of 1.79 $mg.g_{DW}^{-1}.h^{-1}$ for MG02, which is 46% higher compared to MG01 (Table 3.3). Biomass productivity is higher for MG01 - 0.030 $g.g_{DW}^{-1}.h^{-1}$ - than in MG02 - 0.019 $g.g_{DW}^{-1}.h^{-1}$ (Table 3.3). Ethanol and acetate production rates remained similar for both strains. These results are according to fundamental studies with *S. cerevisiae* in steady-state continuous mode. These studies have shown that cells cultivated at higher dilution rates close to the μmax produce higher amounts of ethanol, glycerol and acetate, than at lower dilution rates (Leuenberger, 1972).

To test if MG production would improve with lower dilution rates where respiration is favored and the production of ethanol is reduced (or even zero), both strains were cultivated at a dilution rate of 0.1 h^{-1}. Under these conditions, MG yield in MG02 is 1.5-fold higher than in MG01 (Table 3.3). Regarding MG productivities, the MG production rate is 0.76 $mg.g_{DW}^{-1}.h^{-1}$ for MG01 and 1.17 $mg.g_{DW}^{-1}.h^{-1}$ for MG02, which represents a 1.5-fold increase. Although there is a distinct difference between the production of MG in MG01 and MG02, the productivity of both strains in glucose-limited chemostat is lower when compared to high-glucose chemostat. Since MG yields on biomass are similar for both strains across the different dilution rates and MG yields on substrate even exhibit an opposite trend (being larger for the lower dilution rate for MG02), it can be concluded that the observed increase in MG productivity at high dilution rates is only due to the faster process caused directly by the dilution.

Biomass production in MG02 increased from dilution 0.15 h^{-1} (high-glucose) to 0.1 h^{-1} (glucose-limited) which is coherent with previews publications (Table 3.3) (Van Hoek *et al.*, 1998). When yeast cells are cultivated at high rates, the flux from glycolysis is re-directed to the production of by-products (e.g. ethanol) and respiration is kept at lower levels. However, when the dilution is reduced, respiration becomes the major source of energy through TCA cycle and consequently, the production of biomass increases (Van Hoek *et al.*, 1998). Strangely, this behavior is not observed for strain MG01, since biomass production decreases from high-glucose chemostats to glucose-limited. In both cases, however, the overlapping confidence intervals do not allow to extract clear conclusions.

It would be expected that for lower dilution rates the production of ethanol, glycerol and acetate decreased in relation to the high-glucose chemostats. In fact, at a dilution rate of 0.1 h^{-1} wild-type *S. cerevisiae* cultures are often in a pure respiratory regimen and by-product accumulation is thus absent (Postma *et al.*, 1989). It thus seems that MG production strongly affects the respiratory capacity of *S. cerevisae*. Also, only MG01 shows a reduction of ethanol and glycerol production, while acetate production remained the same (Table 3.3). For MG02, a reduction on the ethanol yield was likely to be observed, as attained in MG01. Interestingly, steady-state MG02 cells cultivated under a dilution rate of 0.1 h^{-1} showed a profile of ethanol production that is largely influenced by the concentration of ethanol produced during growth in the batch mode. Chemostat experiments were initiated with different ethanol concentrations, which apparently produced different results for the ethanol concentration at steady-state. However, the rate of ethanol decrease from the beginning of the chemostat until steady-state is very similar, for both experiments. The fact that the initial ethanol concentration is not fully washed from the bioreactor

during the continuous cultivation at low dilution rates may be an indication that both strains are not in their full respiratory regimen, as ethanol production is not expected, or expected at low levels, in cases where yeast cells are on respiratory metabolism (Van Hoek *et al.*, 1998). However, the influence of initial ethanol concentration in continuous cultures in the respiratory capacity has not been, to our knowledge, previously reported and remains unexplained.

To evaluate the behavior of strain MG02 at low dilution rates and given the inconclusive results at dilution rate of 0.1 h^{-1}, a chemostat with a lower dilution rate of 0.05 h^{-1} was performed. Under these conditions, ethanol and acetate yields dropped to 0.1 $g.g_{glc}^{-1}$ and glycerol production was not detected, indicating that fermentative metabolism had a lower contribution. Even more, the biomass yield increased to 0.18 $g.g_{glc}^{-1}$ and follows an increasing pattern as the dilution rate decreases (Figure 3.3A). Nevertheless, at dilution 0.05 h^{-1}, MG yield on biomass significantly decreases to 8.21 $mg.g_{DW}^{-1}$ in comparison to 11.71 $mg_{MG}.g_{DW}^{-1}$ in dilution 0.1 h^{-1} (glucose-limited) (Table 3.3).

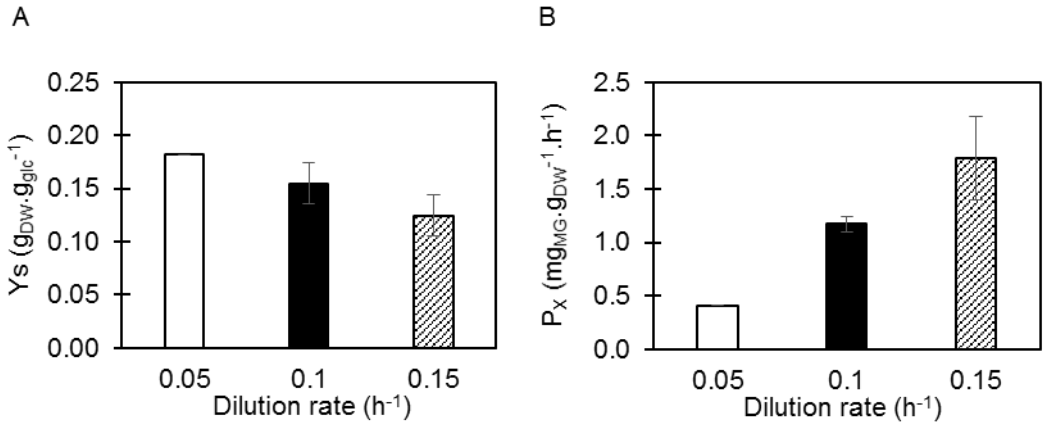

Figure 3.3 Effect of dilution rate in the formation of biomass and MG productivity for MG02 (mgsD ↑pmi40 ↑psa1). Cells were grown in a 2-L batch fermenter containing 0.8 L of synthetic media with 20 $g.L^{-1}$ glucose. A) biomass yield on substrate, represented as $g_{DW}.g_{glc}^{-1}$ for dilutions 0.05, 0.1 and 0.15 h^{-1} and B) MG productivity represented as $mg_{MG}.g_{DW}^{-1}.h^{-1}$ at dilutions 0.05, 0.1 and 0.15 h^{-1}. For dilution 0.1 and 0.15 h^{-1} data is the mean ± SD of two independent experiments; dilution 0.05 h^{-1} represents one experiment.

Cultivation with different dilution rates allowed the establishment of a direct correlation between dilution and the production rate of MG. By gradually decreasing the dilution rate in MG02, we verified a decrease in the production rate of MG (Figure 3.3B), not only due to the higher process rates but also due to decreased MG yields, especially when comparing the 0.05 h^{-1} dilution rate with the remaining ones. Continuous mode chemostat presents several advantages in comparison with batch cultures and it is a valuable tool to explore the best conditions that will be selected for industrial production. One advantage of continuous mode cultivation and that is explored in this work is the evaluation under controlled conditions of physiological parameters based on growth rates (Nielsen *et al.*, 2003). These parameters represent an important evaluation before scale-up and industrialization of a bioprocess. The cultivation of *S. cerevisiae* in respiro-fermentative metabolism holds some issues as it is accompanied by the accumulation of by-products, low

biomass yield and low glucose rentability. Additionally, ethanol production might interfere with the productivity of compounds by provoking toxicity (Stanley *et al.*, 2010). For this reason, industrial bioprocesses usually use a dilution rate of 0.1 h^{-1} or lower (Nielsen *et al.*, 2003). That is the case of resveratrol production (Vos *et al.*, 2015). Nonetheless, the respiro-fermentative metabolism favors the production of certain compounds. In a recent study, the cultivation conditions for ethylene production were optimized and were found to be coupled with the growth rate (Johansson *et al.*, 2013). In the present work, we also observed that MG accumulation is coupled to high growth rates. This can be explained by the fact that MG precursors, GDP-mannose and 3PG, derive both from glycolysis. As the respiro-fermentative metabolism of *S. cerevisiae* presents a higher glycolytic flux than in respiration, it is feasible to foresee that MG production would increase as it was observed. Additionally, it has been described that *PSA1* is an essential gene linked to the progression of phase G1 of the cell cycle, since the efficient glycosylation of proteins is a key-point for cell division (Zhang *et al.*, 1999). *PSA1* transcript levels will increase with higher growth rates, which may favor the synthesis of GDP-mannose and ultimately the synthesis of MG.

3.4 Conclusion

In this study, we have increased the production of mannosylgycerate (MG), a protein stabilizer with outstanding properties, by over-expressing *PSA1* and *PMI40* (GDP-mannose pathway) and by optimizing the mode of cultivation. The resulting strain - MG02 - represents a maximum of 2.2-fold improvement in MG yield on biomass, being able to produce a maximum of 1.79 mg of MG per gram of biomass per hour. It was also possible to establish a correlation between growth rate and MG production. New rounds of optimization are needed to produce MG in *S. cerevisiae* at adequate levels, i. e., well above the values reported for organisms that naturally produce this compound. To this end, a holistic approach should be considered, based on *in silico* driven metabolic engineering to find gene targets that will improve MG production.

Chapter 4

In silico design of *Saccharomyces cerevisiae* strains for high production of mannosylglycerate

Mannosylglycerate (MG) is one of the most notable compatible solutes accumulated by (hyper)thermophiles in response to either osmotic or thermal stresses. Its superior ability to stabilize proteins *in vitro* makes it a promising candidate for industrial applications wherein protein performance and robustness are important requisites. Also, MG is an efficient protector of aggregating-prone proteins in the intracellular milieu and this feature could set the basis for the development of therapeutics to combat neurodegenerative diseases, such as Parkinson´s and Alzheimer´s.

Recently, we engineered *S. cerevisiae* to synthesize MG by heterologous expression of the gene encoding the bifunctional enzyme mannosyl-3-phosphoglycerate synthase / phosphatase of *Dehalococcoides mccartyi*; moreover, following a strategy dictated by common sense, the biosynthetic pathway of one of the precursors, GDP-mannose, was over-expressed (Chapter 3). However, and despite some extent of fermentation optimization, the productivity was poor and the titers far below those required for large-scale industrial applications.

Systems biology is shaping the way science is progressing towards the development of efficient cell factories. In particular, genome-scale metabolic modeling is a strong asset to accelerate the construction of novel strains that will replace the current petrol-based processes. In this work, we used simulated annealing algorithm and different simulation methods to identify the best strategies for improving MG production in *S. cerevisiae*. The resulting strains were characterized in batch fermentation as well as continuous mode. There was a 1.6-fold increase in MG production in comparison with MG02 strain from Chapter 3, but the experimental results on MG yields were considerably lower than foreseen by simulations, suggesting the involvement of a regulatory response that the models were unable to predict.

C. Faria, N. Borges, H. Santos, I. Rocha. *In silico* design of *Saccharomyces cerevisiae* strains for high production of mannosylglycerate. (2017). Manuscript in preparation.

4.1 Introduction

Mannosylglycerate (MG) accumulates in thermophilic and hyperthermophilic prokaryotes primarily in response to osmotic stress (da Costa *et al.*, 1998; Lamosa *et al.*, 2007; Santos *et al.*, 2011; Santos & da Costa, 2002); accumulation of MG induced by thermal stress also occurs, though more rarely (Empadinhas *et al.*, 2001; Martins & Santos, 1995; Neves *et al.*, 2005). This heteroside was first encountered in red algae of the order Ceramiales, but in contrast with prokaryotes, MG plays no significant role in osmotic adaptation of those organisms (Eggert *et al.*, 2007). Therefore, the physiological function of MG in algae remains elusive.

In vitro experiments demonstrated that MG possesses a remarkable ability to protect proteins against thermal denaturation and aggregation (Faria *et al.*, 2008). Also, by engineering the biosynthesis of MG in a Parkinson's yeast model, we showed that the number of cells with α-synuclein inclusions was significantly reduced (Faria *et al.*, 2013). Moreover, MG increased the life-span of retroviral vectors (Cruz *et al.*, 2006) and improved the sensitivity of DNA microarrays (Mascellani *et al.*, 2007). These properties suggested many commercial applications for this solute and led to several patent applications (da Costa *et al.*, 2003; Lamosa *et al.*, 2006a; Santos *et al.*, 1996, 1998). In particular, this stress metabolite could be useful in the cosmetic industry as moisturizer and skin protector against UV damage and in analytical and clinical kits as general stabilizer of proteins and other biomaterials. Of upmost importance is the prospect to replace the cold-chain needed for long-term storage of vaccines, especially for delivery in hot climates. Moreover, in the agro industry, the heterologous expression of the MG pathway was demonstrated as a means to improve the resistance of plants to drought and other stresses (Scheller *et al.*, 2010).

Currently, MG is obtained from a natural producer, *Rhodothermus marinus*, and commercialized in a small scale by Bitop AG (Witten, Germany). A protocol for chemical synthesis has been established, but the method was not optimized for large scale production (Lamosa *et al.*, 2006b); moreover, the well-known bias of European consumers against synthetic products makes this process far less appealing than fermentation.

More and more, the development of improved cell factories relies on the rational design of strains. New approaches emerge as advances in the computational and synthetic biology areas continue to expand and come up with answers to complex metabolic problems (Lee *et al.*, 2009). In this work, we aim to improve the production of MG in *Saccharomyces cerevisiae* by using an engineering strategy supported by genome-scale modeling. *S. cerevisiae* is a robust industrial bio-platform for which a wealth of information has been gathered and an extensive repertoire of DNA manipulating techniques is available (Borodina & Nielsen, 2014; Goffeau *et al.*, 1996). Moreover, the accessibility of multiple computational models, including 13 genome-scale models, makes this yeast an ideal model organism for industrial as well as fundamental research.

Several methods are available to exploit the potentialities of genome-scale models in simulation and optimization of novel strains. The most widely used approach to simulate strain phenotypes is flux balance analysis (FBA) which calculates flux distributions by assuming, as objective

function, that microorganisms have evolved to maximize biomass (Orth *et al.*, 2010). This methodology can successfully predict gene essentiality, growth rates and knockout strategies for a given strain. Nonetheless, FBA can produce unrealistic results for knockout mutants since it does not account for certain perturbations generated by gene disruptions. To circumvent this limitation, a new method was formulated called Minimization Of Metabolic Adjustment (MOMA), which incorporates gene knockout effects on the network by setting the wild-type as reference (Segrè *et al.*, 2002). Although the predictions are usually more conservative, this method can, under the conditions evaluated, actually predict targets and gene essentiality with more accuracy than FBA. Besides simulation algorithms, optimization tools are useful to look for genetic modifications that will redirect fluxes towards a desired compound. Optknock is a bi-level optimization algorithm that accounts for maximization of growth and chemical production, delivering knockout solutions that accomplish these two objectives (Burgard *et al.*, 2003). New updated formulations were designed, including evolutionary methods that can solve large gene modification problems in less time and can be combined with simulation methods such as FBA or MOMA (Patil *et al.*, 2005). In this study, we use the genome-scale model of *S. cerevisiae i*MM904 (Mo *et al.*, 2009) to identify and select targets using Simulated Annealing algorithm (Rocha *et al.*, 2008b). Different simulation methods and objective functions were run to disclose the best set of solutions, including an under/over-expression plugin that searches for solutions with increased (over) or decreased (under) fluxes for a given reaction or expression of a given gene when compared with the reference strain (Gonçalves *et al.*, 2012).

MG can be synthesized by two pathways: in the single-step pathway the MG synthase (MGS) catalyzes the conversion of GDP-mannose and D-glycerate into MG (Martins *et al.*, 1999); and in the two-step pathway, 3-phosphoglycerate (3PG) and GDP-mannose are converted by mannosyl-3-phosphoglycerate synthase (MPGS) into mannosyl-3-phosphoglycerate, which is subsequently dephosphorylated by the mannosyl-3-phosphoglycerate phosphatase (MPGP), yielding MG (Empadinhas *et al.*, 2001; Martins *et al.*, 1999). This latter pathway is the most common one and was cloned in *S. cerevisiae* using the gene encoding mannosyl-3-phosphoglycerate synthase/phosphatase from *Dehalococcoides mccartyi* (Empadinhas *et al.*, 2004).

We incorporated this biosynthetic pathway in a genome-scale model of *S. cerevisiae* in order to define a model-based strategy obtained by simulation and optimization methods. By using this holistic strategy, we aim to increase MG production above the level reported in Chapter 3. The *in silico* results led to the construction of three mutant strains that were characterized and compared with the reference strain (only harboring a plasmid for MG synthesis). In the end, we were able to obtain a strain that produces 25.3 $mg.g_{DW}^{-1}$, which represents a 60% increase in comparison with strain MG02 described in Chapter 3.

4.2 Methods

4.2.1 Model and software

The genome-scale model used here is an updated version of *i*MM904 (Mo *et al.*, 2009; Pereira *et al.*, 2016) that was imported into OptFlux 3.07 (Rocha *et al.*, 2010). A lumped reaction representing the two steps of MG formation, i. e. the conversion of GDP-mannose and 3-phosphosglycerate to MG, GDP, and inorganic phosphate (R_MG: M_gdpmann_c + M_3pg_c → M_MG_c + M_pi_c + M_gdp_c), and a drain to MG were added to the model (R_EX_MG: MG_c → MG_e → MG_b). For optimization purposes, *in silico* environmental conditions were set to mimic minimal growth media supplemented with glucose under aerobic conditions (ammonia: unconstrained uptake, phosphate: unconstrained uptake, sulfate: unconstrained uptake, oxygen: unconstrained uptake, glucose uptake: 1.15 mmol.g_{DW}^{-1}.h^{-1}). All simulations were run within OptFlux 3.07 using IBM CPLEX Optimization Studio (Academic) as the linear and Mixed-Integer Linear Programming solvers. To analyze fermentative metabolism, oxygen uptake was set to 1 mmol.g_{DW}^{-1}.h^{-1}.

4.2.2 *In silico* optimization of mannosylglycerate production

The simulated annealing algorithm included in OptFlux was used for optimization purposes (Patil *et al.*, 2005; Rocha *et al.*, 2008b). pFBA (parsimonious FBA) (Lewis *et al.*, 2010; Ponce de León *et al.*, 2008) was used for the calculation of the reference (wild-type) flux distributions, with the maximization of biomass production as the objective function (Ponce de León *et al.*, 2008). MOMA (Segrè *et al.*, 2002), Linear MOMA (Becker *et al.*, 2007), MIMBL (Minimization of Metabolites Balance) (Brochado *et al.*, 2012) and ROOM (Regulatory On/Off Minimization of Metabolic Flux Changes) (Shlomi *et al.*, 2005b) were used as simulation methods. The objective functions considered in the algorithm were the Biomass-Product Coupled Yield (BPCY) and Yield (Patil *et al.*, 2005). The algorithm was run 3 times, setting the number of solution evaluations to 20,000 and the maximum number of strain modifications to 6.

4.2.3 DNA manipulation

PCR reactions were performed in a Thermal Cycler from Bio-Rad (Hercules, USA) using the protocols recommended by the manufacturers for each polymerase. Yeast genomic DNA for colony PCR and gene cloning was prepared as previously described Lõoke *et al.*, 2011 . Primers were purchased from Metabion (Germany). High fidelity PCR and Fusion PCRs were performed with Phusion polymerase from Thermoscientific (Waltham, USA); colony PCR was performed with DreamTaq DNA polymerase from Thermoscientific. Plasmid extraction, PCR product purification, and DNA gel extraction were carried out using Zymo Research kits (USA). All restriction enzymes were obtained in fast-digestion format from Thermoscientific.

Plasmid isolation and maintenance were carried out in *Escherichia coli* DH5α following the competence and transformation procedures developed by Dower *et al.*, 1988 . *S. cerevisiae* transformation was performed using the lithium acetate method as described by Gietz & Woods, 2002). Yeast cloning procedures used for cassette construction were performed as described earlier (Orr-Weaver & Szostak, 1983).

4.2.4 Strain maintenance and media

Selection and maintenance in *E. coli* of the plasmids constructed/used in this work was performed in LB medium containing 10 $g.L^{-1}$ of peptone, 10 $g.L^{-1}$ of NaCl, 5 $g.L^{-1}$ of yeast extract and supplemented with 100 $mg.L^{-1}$ of ampicillin. The solid version of this medium also included 20 $g.L^{-1}$ of agar. All cultivations of *E. coli* were performed at 37 ºC and 200 rpm agitation.
Cells of *S. cerevisiae* strain CENPK2-1C (MATa ura3-52 his3-Δ1 leu2-3,112 trp1-289, MAL2-8c SUC2) were cultivated in YPD medium, containing 10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ of peptone, 20 $g.L^{-1}$ of glucose, at 30 ºC with 160 rpm. Recombinant strains were selected on Synthetic Dextrose (SD) medium containing 6.7 $g.L^{-1}$ of yeast nitrogen base without amino acids from Difco (New Jersey, USA), 20 $g.L^{-1}$ of glucose and 20 $g.L^{-1}$ of agar. When necessary, histidine (20 $mg.L^{-1}$), tryptophan (20 $mg.L^{-1}$), uracil (20 $mg.L^{-1}$) and leucine (30 $mg.L^{-1}$), were added to cover auxotrophies. All cultivations of *S. cerevisiae* were performed at 30 ºC with 160 rpm agitation. Strains were grown overnight and stored at -80 ºC in selective media plus 30% (w/v) of glycerol.

4.2.5 Plasmid constructions

The plasmid pDES containing the gene *mgsD* from *D. mccartyi* under the control of the ENO2 promotor, was kindly provided by Milton S. da Costa (Coimbra University, Portugal). All primer sequences used in this study are shown in Table 4.1. pDES was modified by adding a Kosak sequence (5´- AAAAAA - 3') between the ENO2 promoter and *mgsD* starting codon. Additionally, a synthetic terminator (Ts) with the sequence 5´-TGGGTGGTATATATATATATATATATATATATAACTGTCTAGAAATAAAGAGTATCATCTTTCAAA-3' was cloned immediately after the stop codon of *mgsD* (Curran *et al.*, 2015). For this construction, *mgsD* was amplified with primers 1 and 2. Then, this PCR product was amplified by using primers 1 and 3 to complete the fusion of the terminator with *msgD*. This new construction and pDES were digested with BamHI / HindIII and then ligated to form pMG.

The plasmid pSP-GM (kindly provided by Jens Nielsen's group, Chalmers University of Technology, Sweden) was used to express genes *PMI40* and *PSA1* from *S. cerevisiae. PMI40* was amplified with primers 4 and 5 using as template genomic DNA from *S. cerevisiae*, digested with SacI / NotI and ligated in pSP-GM between promoter *TEF1* and terminator *ADH1* to yield plasmid pSP01. Next, *PSA1* was amplified with primers 6 and 7 using as template genomic DNA from *S. cerevisiae* and digested with HindIII / XmaI. pSP01 was digested with the same enzymes so that *PSA1* could be placed between promoter *PGK* and terminator *CYC1*. Ligation was

achieved by using T4 ligase, yielding plasmid pSP02. Positive colonies harboring pMG, pSP01, pSP02 were identified by colony PCR with the primers used in the amplification. Correctness of plasmid constructions was confirmed by restriction analysis and DNA sequencing (GATC Biotech, Germany). All plasmids used and constructed in this study are listed in Table 4.2.

Table 4.1 List of primers used to construct plasmids and integrative cassettes.

Primer	Sequence
1	CCAAGCTTAAAAAAATGCGCATTGAAAGCCTG
2	CAGTTATATATATATATATATATATATATACCACCCATTATTCCATGGGCAGTATTA
3	TTTGAAAGATGATACTCTTTATTTCTAGACAGTTATATATATATATATATATATATACCACC
4	CCGCGGCCGCAAAAAAATGTCCAACAAGCTGTTCAGG
5	CCGAGCTCCTAATTTGGTTCCACAAAGGC
6	GGCCCGGGAAAAAAATGAAAGGTTTAATTTTAGTCGG
7	CCAAGCTTTCACATAATAATAGCTTCCTTTGG
8	ATGACAAGCATTGACATTAACAACTTACAAAATACC
9	CGGGAATTGCCATGAAGCCGAAATTTCAGCGATGACCAATTCTGC
10	GGCGTTACCCAACTTAATCGCCAAATTGCAGGTGCTGCTTTGG
11	TTAATATAGCAATCTAATTGAGATCTTAGCAGAGG
12	TTCGGCTTCATGGCAATTCCCGGG
13	GGCGATTAAGTTGGGTAACGCCAGGG
14	ATGTCTTATTCAGCTGCCGATAATTTACAAG
15	CCCGGGAATTGCCATGAAGCCGAATCTCACCAATTACCAATTCTGCTACGG
16	GGCGTTACCCAACTTAATCGCCAAATTGCAGGTGCTGCTTTAGATG
17	TTAGTATAATAACCTGATGGAAACTTTGGCAG
18	TTAAACGGTAGAGACTTGCAAAGTGTTGG
19	CCTGGCGTTACCCAACTTAATCGCCCACCCAGACATCGGGCTTCCAC
20	TATATCGCATGAAGAATAACCAGAGTTTTTCTCCG
21	AATTGCCATGAAGCCGAAGATCAATTCTAACAAAAAAAAAATAAGG

4.2.6 Construction of DNA cassettes

The results obtained in the *in silico* optimization section suggested the elimination of the flux from 3-phosphoglycerate to serine. Two isoenzymes encoded by the genes *SER3* and *SER33* are involved in this conversion. *S. cerevisiae* genome was used as template for PCR procedures except when stated otherwise. To disrupt those genes we built a disruption cassette, which is composed by the first and the last 500 bp of each gene, intercalated with an *URA3* gene and the respective promoter, for selection purposes. Therefore, the upstream sequence of *SER3* was amplified with primers 8 and 9 (Table 4.1), and the downstream sequence was amplified with primers 10 and 11. *URA3* gene and the respective promoter from *Kluyveromyces lactis* was amplified with primers 12 and 13 from the plasmid pWJ1042 (kindly provided by Jens Nielsen´s group, Chalmers University of Technology, Sweden). All PCR products were purified and fused

by PCR using primers 8 and 11. For *SER33* cassette, the upstream sequence of *SER33* was amplified with primers 14 and 15 and the downstream sequence was amplified with primers 16 and 17. All PCR products were purified and fused by PCR using primers 14 and 17.

Another target flux identified to improve MG production was the reaction from phosphoenolpyruvate to pyruvate catalyzed by the pyruvate kinase and coded by gene *PYK1*. However, the disruption of this gene caused growth arrest in *S. cerevisiae* (Pearce *et al.*, 2001) and for this reason we set to construct a DNA cassette to provoke an under-expression of the gene by changing the promoter strength. This cassette was composed by an upstream region of *PYK1* (*PYK1*-up) linked with *URA3* gene and promoter for selection, the *PYK1* promoter lacking an upstream activating sequence (UAS), and the *PYK1* gene. The fused *PYK1p*$_{\Delta 653}$-*PYK1* (Pearce *et al.*, 2001) was amplified with primers 18 and 19. *URA3* gene and the respective promoter were amplified as stated above. Then, *PYK1*-up was amplified using primers 20 and 21 and fused with *URA3* using primers 13 and 21. Finally, *PYK1*-up-*URA3* was fused with *PYK1p*$_{\Delta 653}$-*PYK1* using primers 19 and 20 (see Figure 4.2).

Table 4.2 List of plasmids used in this study.

Plasmid	Description	Sources
pDES	p425:P_{ENO2}-*mgsD* LEU2	(Empadinhas *et al.*, 2004)
pMG	p425:P_{ENO2}-kosak-*mgsD*-T_S LEU2	This work
pSP-GM	Wild-type plasmid	(Partow *et al.*, 2010)
pSP01	pSP-GM:P_{TEF1}-*kosak*-pmi40-T_{ADH1} URA3	This work
pSP02	pSP-GM:P_{TEF1}-*kosak*-pmi40-T_{ADH1} P_{PGK1}-kosak-psa1-T_{CYC1} URA3	This work

4.2.7 Transformation of *S. cerevisiae* with the DNA cassettes

After amplification of the *SER3* cassette, yeast cells were transformed with 3 µg of DNA following the standard protocol (Gietz & Woods, 2002). The list of strains used in this study is shown in Table 4.3. Selection was made in SD-URA plates with agar (see section Cultivation Conditions above) and DNA insertion was confirmed by PCR. As described previously by Akada *et al.*, 2006 , the positive colonies were plated in YPD with 0.75 mg.L^{-1} of 5-fluoroorotic acid to select cells that have excised URA3 selection mark. This strain served as reference to disrupt gene SER33 (herein called strain S3) and to truncate PYK1 promotor (strain S4). After each disruption, the URA3 marker was recycled.

Table 4.3 List of strains used in this study.

Strain	Description	Plasmid	Sources
MG01	*MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2*	pDES	Chapter 3
S1	*MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2*	pMG	This work
S2	*MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2*	pMG, pSP02	This work
S3	*MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2; Δser3 Δser33*	pMG, pSP02	This work
S4	*MATa; his3D1; leu2-3_112; ura3-52; trp1-289; MAL2-8c; SUC2; Δser3 Δser33; PYK1pΔ_{653}-PYK1*	pMG, pSP02	This work

4.2.8 Cultivation in bioreactor

Batch fermentations were performed in SD medium containing 20 $g.L^{-1}$ of glucose. Cells were pre-grown overnight in shake flasks containing 50 mL of the same medium, at 30ºC and 160 rpm. Each fermenter was inoculated at an initial OD_{600} of 0.1. Optical density was measured using the spectrophotometer V-560 from Jasco (Japan). The batch fermentations were performed in the Eppendorf DASGIP Parallel Bioreactor System (Switzerland) coupled with 2-L culture vessels. The operating volume for the fermentations was 1-L, temperature was maintained at 30 ºC, airflow at 1 VVM, pH was kept at 5.5, controlled by addition of 2 M NaOH, and dissolved oxygen was kept above 30 % of saturation by feedback control of the stirring speed from 200 rpm up to 400 rpm. The concentration of O_2 and CO_2 in the exhaust gas was monitored by Bluesens gas analyzers (Germany).

For continuous cultures, cells were allowed to grow in batch mode until glucose was depleted from the medium. Then, a peristaltic pump was turned on to deliver new medium into the vessel at dilution rates of 0.05 and 0.1 h^{-1}. Cultivation broth was removed from the vessel using a leveled tube connected to a peristaltic pump. To guarantee that the culture reached a steady-state mode, at least 5 retention times were allowed.

4.2.9 Sampling and quantification of fermentation products

For each sampling, 50 mL of culture broth were removed from the bioreactor and centrifuged at 3600 × *g* for 10 min to separate cells from the medium. Next, the supernatant was filtered through a membrane filter with a pore size of 0.22 µm into HPLC vials and stored at -20 ºC until further analyzed for glucose, glycerol, acetate, and ethanol. MG is an intracellular compound and was extracted from the pellet centrifuged above by adding 5 mL of water, 5 mL of methanol and 10 mL of chloroform. This mixture was vigorously shaken with the help of a vortex and then centrifuged at 12,000 × *g* for 30 min at 4 ºC. The top part, corresponding to the aqueous phase, was carefully transferred to a new tube and the process of centrifugation was repeated after the addition of 5 mL of water. Once the aqueous phase was separated, samples were evaporated using a Savant™

SPD131DDA SpeedVac (Thermoscientific). Next, the evaporated samples were dissolved in 1 mL of water and transferred to an HPLC vial. Analysis was performed in an HPLC apparatus from Jasco (Japan) model LC-NetII/ADC equipped with UV-2075 Plus and RI-2031 Plus detectors, also from Jasco. The samples were analyzed using an Aminex HPX-87H column from Bio-Rad, which was kept at 50 ºC, and a solution of 0.01 M of H_2SO_4 was used as the mobile phase with a flow rate of 0.3 mL.min^{-1}. Quantitative analysis of glucose, MG, glycerol, acetate and ethanol was performed by comparison with a mixture of standards with known concentrations of each metabolite. Calibration curves were prepared using the peak areas of the RI detector for glucose, glycerol, acetate and ethanol and of the UV absorbance for MG. The cell dry weight was determined by filtering 5 mL of culture broth through a 0.22 µm pore Millipore filter and washing once with 5 mL of water. Filters were then dried for 10 min at 150 W in a microwave oven and weighted using an analytical balance.

4.2.10 Quantification of enzymatic activities

For the quantification of enzymatic activities, strains S2 and S4 were grown in shake flasks containing 50 mL of SD medium with 20 g.L^{-1} of glucose at 30°C and 160 rpm. Cells were harvested at mid-exponential phase and centrifuged at 3600 × *g* for 10 min. Then, cells were re-suspended in deionized water containing a protease inhibitor cocktail from Sigma (Sigma, USA) and were broken 3 times using a Fastprep 24 system (MP, USA). Supernatants were used to determine enzymatic activity following centrifugation at 13,500 × *g* for 5 min. Protein concentration was determined using the Bradford method (Bradford, 1976). All reagents were purchased from Sigma and are of the best grade available. The methods used to determine the enzymatic activities are described in Bergmeyer & Gawehn, 1974 . Briefly, the mannose-6-phosphate isomerase activity was determined by using a spectrophotometric method. The assay mixture contained the following components in a final volume of 1 mL: 50 mM Tris-HCl (pH 7), 5 mM mannose-6-phosphate, 8 mM $MgCl_2$, 2.8 mM NADP, 0.27 U of glucose-6-phosphate dehydrogenase and phosphoglucose isomerase and 0.1 mg of cell extract. Reactions were started by the addition of mannose-6-phosphate and the formation of NADPH was monitored at 340 nm for 5 min. The GDP-mannose pyrophosphorylase activity was determined using a spectrophotometric method. The assay mixture contained the following components in a final volume of 1 mL: 50 mM of buffer Tris-HCl (pH 7), 0.5 mM NaF, 10 mM $MgCl_2$, 0.1 mM ADP, 1 mM glucose, 1 mM NADP, 2 mM PPi, GDP-mannose at a concentration of 0.2 mM, 1 U of nucleoside kinase, hexokinase, and glucose-6-phosphate dehydrogenase and 0.1 mg (total protein) of cell extract. The reaction was started by the addition of GDP-mannose, and the formation of NADPH was recorded at 340 nm for 5 min. For the phosphomannomutase, the assay mixture contained the following components in a final volume of 1 mL: 50 mM of buffer Tris-HCl (pH 7), 0.5 mM $MgCl_2$, 0.25 mM NADP, 10 µg of yeast glucose-6-phosphate dehydrogenase, 0.1 mM mannose-1-phosphate, 1 µM mannose-1,6-bisphosphate, 10 µg phosphoglucose isomerase, 3.5 µg phosphomannose isomerase and 0.1 mg of cell extract. The reaction was started by the addition of mannose-1-phosphate and the

production of NADPH was measured at 340 nm for 5 min. Mannose-1,6-bisphosphate was synthetized by incubating for 2 h at 30°C a mixture of 50 mM Tris-HCl (pH 7), 0.3 mM mannose-1-phosphate, 0.2 mM glucose-1,6-bisphosphate, 5 mM $MgCl_2$, 0.25 mM NADP, 1 mM dithiothreitol, 5 µg/mL yeast glucose-6-phosphate dehydrogenase, and 10 µg/mL muscle phosphoglucomutase. Pyruvate kinase activity was determined using a spectrophotometric method and the assay mixture contained the following components in a final volume of 1 mL: 24 mM KH_2PO_4/K_2HPO_4 (pH 7.0), 150 µM NADH, 1 mM fructose-1,6-bisphosphate, 2.4 mM ADP, 25 U lactate dehydrogenase, 10 mM $MgSO_4$, 800 µM PEP and 0.1 mg of cell extract. The reaction was started by the addition of PEP, and the formation of NAD^+ was monitored at 340 nm for 5 min.

4.3 Results

4.3.1 *In silico* analysis

To accommodate the genome-scale model *i*MM904 we used the Optflux platform (Rocha *et al.*, 2010). The *i*MM904 model was modified to contain the reactions corresponding to MG biosynthesis, as described in the methods section. Using the simulated annealing optimization algorithm and the pFBA, MOMA, LMOMA, MIMBL and ROOM simulation methods, we set to identify which reactions and/or genes should be manipulated to increase the production of MG, either using BPCY (Biomass Product Coupled Yield) or YIELD (product flux with minimum biomass) as objective functions. For this study, glucose was constrained to 1.15 $mmol.g_{DW}^{-1}.h^{-1}$ and the limits for inorganic phosphate, oxygen, sulfate and ammonia were left unconstrained. In addition to knockout strategies, an over/under-expression plug-in that is implemented in the Optflux framework was used (Gonçalves *et al.*, 2012). This plug-in revealed to be very useful as it led us to the identification of candidate genes to over- and under-express, while also allowing to include knockout strategies. All solutions obtained by *in silico* methods were manually curated using *S. cerevisiae* databases with genomic, transcriptomic and physiological experimental data, namely Saccharomyces Genome Database (Cherry *et al.*, 2012), YEASTRACT (Teixeira *et al.*, 2014) and Kyoto Encyclopedia of Genes and Genomes (Kanehisa *et al.*, 2017). The maximal theoretical production of MG corresponds to a production flux of 0.7 $mmol.g_{DW}^{-1}.h^{-1}$ for the specified glucose uptake rate or a yield of 0.61 $mol_{MG}.mol_{glc}^{-1}$, when all carbon is channeled to MG production (thus with zero biomass).

The simulated annealing optimization algorithm was run three times, setting the number of solution evaluations to 20,000 and the maximum number of strain modifications allowed to 6. For each optimization, a set of solutions were obtained and the best four solutions of each simulation method were selected and categorized according to the MG production flux and the specific growth rate (Figure 4.1-A). The analysis of these results revealed that MOMA and LMOMA identified potential targets leading to high MG production, but with compromised growth. On the other hand, ROOM and MIMBL were used to find strategies with low MG production, and growth rates close

to the wild-type. A flux variability analysis is also represented in Figure 4.1-A as the dashed line, indicating the compromise between MG production and growth.

To verify if aerobic fermentation could favor MG production we did a new optimization round, this time by constraining O_2 to 1 mmol.g_{DW}^{-1}.h^{-1}. Results are presented in Figure 4.1-B and show that the respiratory metabolism produces higher yields of MG than the fermentative metabolism.

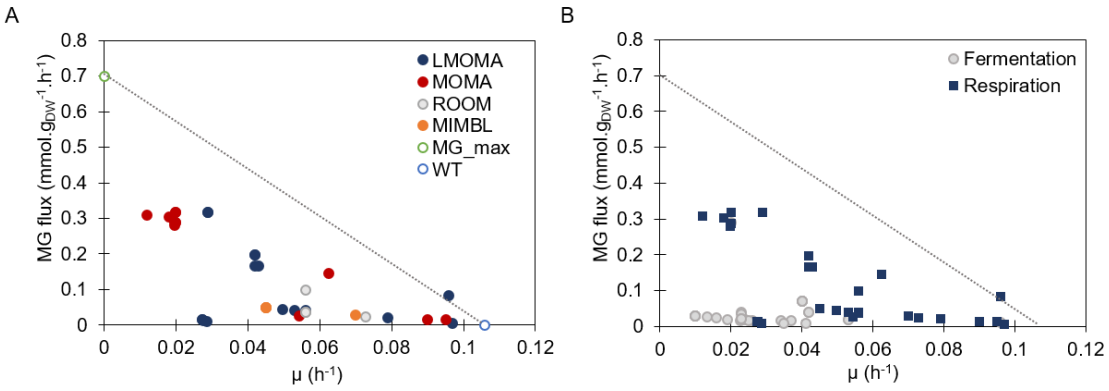

Figure 4.1 *In silico* optimization results for MG production. A - MG production flux versus the specific growth rate for solutions obtained using SA as optimization algorithm and MOMA, LMOMA, ROOM and MIMBL as simulation methods. MG_max corresponds to the maximal theoretical yield of MG, and WT is the wild-type simulation using pFBA. For building this plot, the best four solutions from each optimization setup were selected. B - Specific growth rate and MG production flux for solutions obtained with unconstrained (respiration) and constrained O_2 (fermentation), using the same protocol as described in A. The dashed lines represent the flux variability analysis that correlates MG production with growth.

The solutions displayed in Figure 4.1-A were analyzed and are summarized in Table 4.4 that contains knockout and over/under-expression targets for the best 6 solutions with regard to MG production. MG is formed from two compounds: 3-phosphoglycerate (3PG) and GDP-mannose. The main metabolic strategy identified included the inhibition of reactions that consume GDP-mannose and 3PG. To block 3PG spending, the algorithm suggested removing the flux from phosphoenolpyruvate (PEP) to pyruvate (PYR), which increased MG production, although impacting significantly biomass (solution 1 in Table 4.4). Another incremental strategy (solution 2 in Table 4.4) advises the elimination of two reactions that consume 3PG, namely its conversion to 2-phosphoglycerate (2PG) following glycolysis and to 3-phosphohydroxypyruvate (3PHP) that leads to the synthesis of serine. In addition, this strategy also advocates the elimination of a reaction involved in phospholipid synthesis that uses glycerol-3-phosphate as an intermediate to synthetize CDP-diacylglycerol. This solution increases 2-fold the MG production yield in comparison with solution 1, but also affects biomass production in 40% in relation to the wild-type. The results for solution 3 suggest the removal of the reaction (3PG → 2PG) to provoke 3PG accumulation, and the elimination of two reactions that spend GDP-mannose, both catalyzed by the mannose-inositol-phosphorylceramide synthase involved in the synthesis of multiple sphingolipids characteristic of yeasts. Although biomass production is predicted to be only 20% of that in wild-type, MG yield is 3.6-fold higher than in solution 2.

Naturally, the algorithms do not choose among the target genes that are *in silico* essential for growth, but it may happen that some of the targets are in fact essential *in vivo* due to limitations in the model. An example is the essentiality of pyruvate kinase to *S. cerevisiae* grown in glucose (Sprague, 1977) or the phosphoglycerate mutase (Papini *et al.*, 2010). Also, *in vivo*, the knockout of the gene that codes for the mannose-inositol-phosphorylceramide synthase presents a phenotype with sensitivity to temperature and decreased resistance to general stress (Cerbón *et al.*, 2005; Jenkins *et al.*, 1997).

Table 4.4 *In silico* strategies and corresponding biomass growth and MG yields obtained using the SA algorithm in conjunction with LMOMA and MOMA methods. μ, specific growth rate. †MG yield is expressed as $mol.mol_{glc}^{-1}$.‡Tmax, percentage of the theoretical maximal MG yield; PEP, phosphoenolpyruvate; ADP adenosine di-phosphate; Pi inorganic phosphate; 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; NAD^+, β-nicotinamide adenine dinucleotide; NADH, β-nicotinamide adenine dinucleotide, reduced form; 3PHP, 3-phosphohydroxypyruvate; CTP, cytidine triphosphate; CDPdiacylglycerol, cytidine diphosphate-diacylglycerol; M1P, mannose-1-phosphate; GDP, guanosine diphosphate; GDPman, GDP-mannose; PPi, pyrophosphate; F6P, fructose-6-phosphate; M6P, mannose-6-phosphate.

KNOCKOUTS				
ID	Reaction	Enzyme	MG yield† (Tmax‡)	μ (h^{-1})
	(wild-type)	none	0	0.10
1	PEP + ADP + Pi → PYR + ATP	Pyruvate kinase	0.035 (5%)	0.06
2	2PG → PEP 3PG + NAD^+ → 3PHP + NADH CTP + phosphatidate→ CDPdiacylglycerol 3PG → 2PG	Enolase 3-phosphoglycerate dehydrogenase CDP-diacylglycerol synthase Glycerate mutase	0.070 (10%)	0.04
3	IPC224 + GDPman → GDP + MPIC224 IPC326 + GDPman → GDP + MPIC326	Mannose-inositol-phosphoryl ceramide synthase	0.250 (36%)	0.02
OVER AND UNDER-EXPRESSION				
ID	Reaction (o/u value)	Enzyme	MG yield† (Tmax‡)	μ (h^{-1})
4	3PG → 2PG (0.5)	P-glycerate mutase	0.019 (3%)	0.08
5	M1P + GDP → GDPman + PPi (2.0) F6P → M6P (2.0) 3PG + NAD^+ → 3PHP + NADH (0)	M1P guanyltransferase M6P isomerase P-glycerate dehydrogenase	0.096 (14%)	0.08
6	M1P + GDP → GDPman + PPi (4.0)	M1P guanyltransferase	0.277 (40%)	0.03

Since all strategies were based on deleting key points of *S. cerevisiae* metabolism we used the Over/Under-expression (Gonçalves *et al.*, 2012) plug-in built for OptFlux v3.2.7 (Rocha *et al.*, 2010) that determines which reaction fluxes should be increased, decreased or eliminated to accomplish a given metabolic engineering aim. The best solutions obtained are described in Table 4.4 and point to three critical strategies: a) down-expression of enzymes that consume 3PG in the glycolytic pathway (solution 4); b) over-expression of the GDP-mannose pathway (solution 5 and 6) and c) elimination of reactions belonging to the serine pathway, which starts with the dehydrogenation of 3PG (solution 5). From these three strategies, solution 5 has a good specific

growth rate while MG yield is 14% of the theoretical maximum. On the other side, solution 6 produces 40% of MG maximum theoretical yield but with low growth (30% of μ_{wt}), making these two solutions similar in global BPCY. One common point for all solutions is that biomass and MG yield correlate inversely, meaning that when biomass is high, predicted MG yield is low and vice-versa. Therefore, strain design must consider a solution that conciliates biomass production and MG yields by pinpointing key reactions.

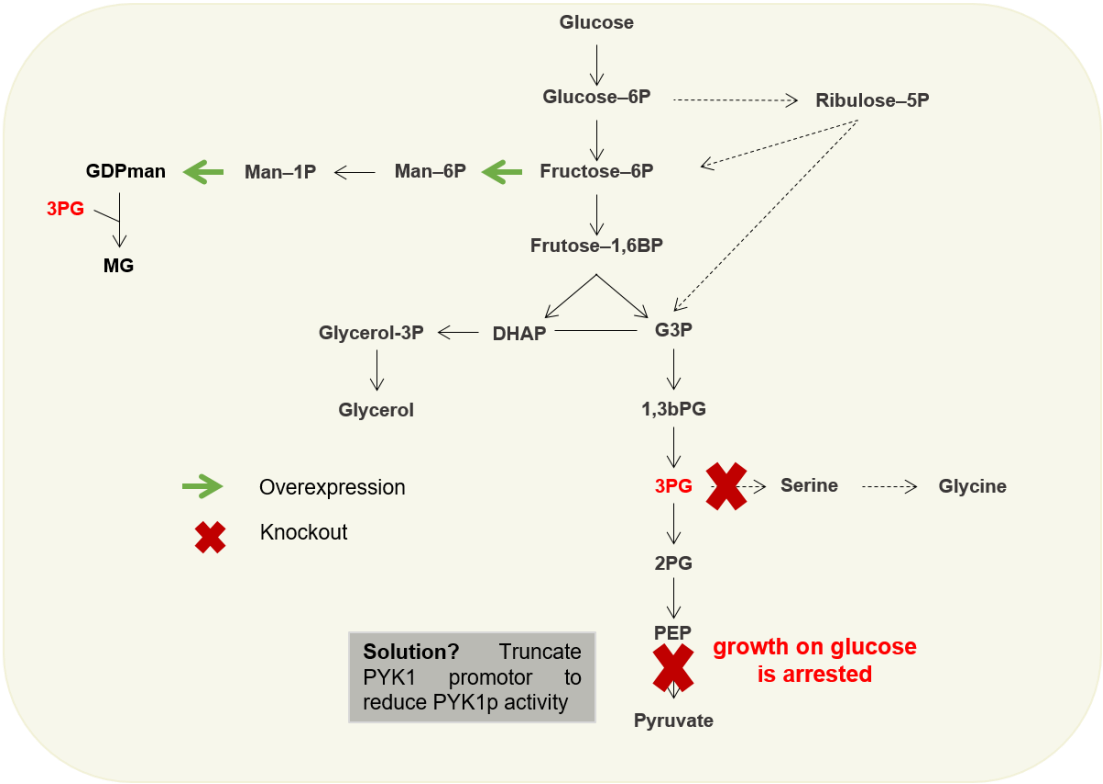

Figure 4.2 Schematic representation of the *in silico*-derived strategy to increase MG production. Green arrows represent the overexpression of the genes *PMI40* and *PSA1*. Red crosses indicate the knockouts of the genes for serine production – *SER3* and *SER33* – and for pyruvate kinase (gene *PYK1*). *PYK1* knockout is not possible to achieve in S. cerevisiae under growth on glucose. A way around this was to diminish *PYK1* promoter strength to down-express *PYK1*. Abbreviations: MG, mannosylglycerate; GDPman, GDP-mannose; Man-1P, mannose-1-phosphate; Man-6P, mannose-6-phosphate; G3P, glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; 1,3bPG, 1,3-bisphospho-D-glycerate; 3PG, 3-phosphoglycerate, 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate.

Based on these analyses, several strategies were selected with different complementary aims based on biological feasibility. One of the aims was to increase the GDP-mannose pool and this was sought by overexpressing *PSA1* and *PMI40,* coding the mannose-1-phosphate guanyltransferase (part of solution 4) and the mannose-6-phosphate isomerase (part of solution 5), respectively. A valid strategy to increase the 3PG pool consists in blocking its flux towards serine synthesis by knocking out *SER3* and *SER33* genes that code for two isoenzymes called 3-phosphoglycerate dehydrogenase (solution 5). Finally, to further increase the 3PG pool, an under-expression of the *PYK1* gene (pyruvate kinase) was also selected (Figure 4.2). The phenotype simulation of this combined solution using MOMA indicates a production of MG of 0.09

$mol_{MG}.mol_{glc}^{-1}$ (15% of the theoretical maximum) with a specific growth rate of 0.09 h^{-1}, very similar to what is obtained in the wild-type, i.e., 0.1 h^{-1}.

4.3.2 *In vivo* implementation

MG biosynthesis in *S. cerevisiae* was firstly observed by Empadinhas and co-workers who cloned the *mgsD* gene from *Dehalococcoides mccartyi* (formerly *Dehalococcoides ethenogenes*) to confirm the functionality of this gene (Empadinhas *et al.*, 2004). To improve *mgsD* transcription, a set of engineering strategies were implemented that included plasmid, promoter and terminator change (as referred in the Methods Section). *S. cerevisiae* harboring pMG, the newly constructed plasmid, was named strain S1 that was able to accumulate 12.91 $mg.g_{DW}^{-1}$ of MG in batch culture, corresponding to a 1.8-fold increase in comparison with pDES (see Chapter 3, Table 3.2). These results refer to cells grown in shake flasks with SD medium plus 20 $g.L^{-1}$ of glucose and the MG content was assessed upon glucose depletion.

To increase GDP-mannose availability, we transformed strain S1 with plasmid pSP02 carrying genes *PSA1* and *PMI40*, yielding strain S2 (Table 4.3). Additionally, the two incremental strategies advocated to increase the 3PG pool were implemented. First, the *SER3* and *SER33* genes were disrupted to block 3PG flux towards the production of serine and this strain was subsequently transformed with plasmids pMG and pSP02 (strain S3). Finally, an under-expression of *PYK1* was envisaged by constructing a *PYK1* promoter lacking an upstream activating sequence (UAS) (Nishizawa *et al.*, 1989). The resulting mutant was named S4 and contained the genetic alterations Δ*SER3* Δ*SER33 PYK1p*Δ_{653}*-PYK1* and the plasmids pMG and pSP02.

All strains were evaluated in controlled, aerobic batch bioreactors with 20 $g.L^{-1}$ of glucose in SD medium until growth on glucose was no longer observed. Over-production of enzymes leading to GDP-mannose (strain S2) exhibited a 1.35-fold improvement on MG titer (32.6 vs 44.1 $mg.L^{-1}$), 2-fold improvement in MG yield per biomass (12.9 vs 25.3 $mg.g_{DW}^{-1}$) and a reduction of 31% on biomass to substrate yield, in relation to strain S1 (Figure 4.3 and Table 4.5). Fermentation by-products such as glycerol, acetate and glycerol were also measured and show a similar profile of production. In strains S1 and S2, all glucose was exhausted.

Genetic alterations to increase the 3PG pool implemented in strains S3 and S4 did not increase MG titer or MG yield per biomass and reverted production to values obtained in the reference strain S1 (Figure 4.3 and Table 4.5). These strains did not consume glucose completely. The profile of by-product accumulation in strain S3 is similar to that observed for strains S1 and S2. In contrast, strain S4 exhibits a lower accumulation of acetate and ethanol. This was expected as a consequence of the down-regulation applied at the level of pyruvate kinase expression. The growth characteristics of the mutants were also studied and showed that the genetic modifications affect the duration of the lag phase, the growth rates and biomass production of all strains. For instance, strain S2 growth and biomass yield are highly impaired in relation to strain S1, which seems to correlate with the production of MG and its accumulation in the cell (Table 4.5). Cellular growth in strain S2 was arrested at 24 h upon inoculation. At this point, MG accumulation reached

26 mM (Table 5) and glucose was still present in the culture broth (Figure 4.3). After 34 h, glucose was completely exhausted but this metabolization did not translate into biomass production.

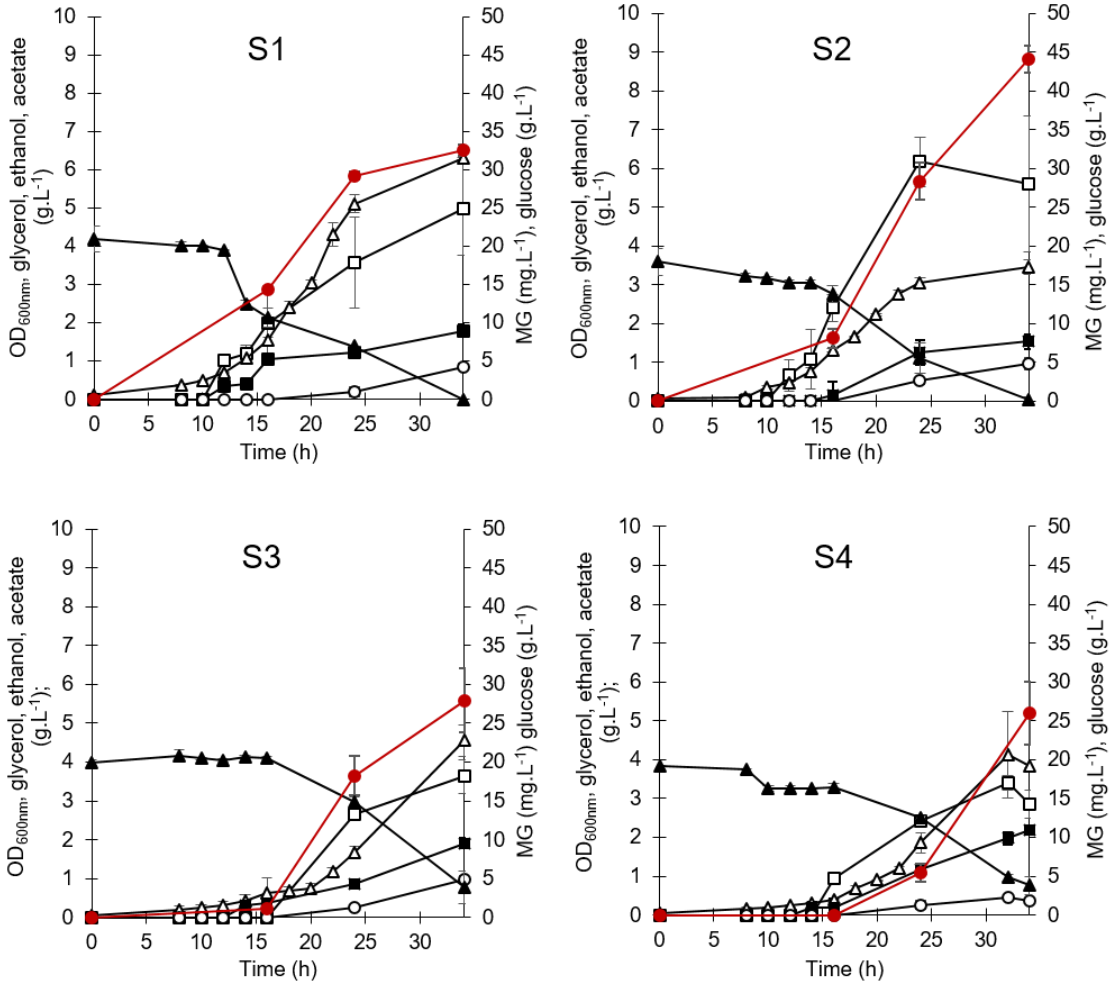

Figure 4.3 Growth curve, MG production and end-products of strains S1, S2, S3 and S4 in controlled bioreactors. S1 (mgsD), S2 (mgsD ↑pmi40 ↑psa1), S3 (mgsD ↑pmi40 ↑psa1 Δser*3* Δser33), and S4 (mgsD ↑pmi40 ↑psa1 Δser3 Δser33 pykp$_{\Delta 653}$-pyk), were cultivated in SD medium with 20 g.L^{-1} of glucose. Symbols are (▲) glucose, (●) MG, (□) ethanol, (Δ) OD$_{600nm}$, (■) glycerol and (○) acetate. Values represent mean ± SD of at least two independent experiments.

Interestingly, of all compounds analyzed, MG is the only one produced by strain S2 in the interval from 24 to 34 h, increasing up to 25.30 mg.g$_{DW}^{-1,}$ which corresponds to a concentration of 39.78 mM. This concentration is 2-fold higher than what is obtained in strain S1 and is probably related with a 33% decrease in the growth rate of strain S2 in comparison with S1 (Table 4.5). Strains S3 and S4 show a very different growth profile when compared with S1 and S2, characterized by a pronounced lag phase (Figure 4.3). Strain S3 has a growth rate similar to S2 but with a higher yield on biomass and final OD. The pyruvate bottleneck distinctively affected strain S4 by reducing the specific growth rate to 0.19 h^{-1} and the final OD. MG production for these two strains decreased

significantly to values observed in strain S1, either on yield per biomass or intracellular concentration (Table 4.5).

Table 4.5 Specific growth rates, biomass yield on substrate, MG yield on biomass and MG concentration for all strains studied. Cells were cultivated in controlled, aerobic batch bioreactors with 20 $g.L^{-1}$ of glucose in SD medium until growth on glucose was no longer observed.

Strain	μ_{max} (h^{-1})	$Y_{X/S}$ ($g_{DW}.g_{glc}^{-1}$)	$Y_{MG/X}$ ($mg.g_{DW}^{-1}$)	MG content (mM)
S1	0.33 ± 0.05	0.13 ± 0.016	12.91 ± 0.13	20.30 ± 2.72
S2	0.22 ± 0.02	0.09 ± 0.003	25.30 ± 1.91	39.78 ± 3.01
S3	0.23 ± 0.06	0.13 ± 0.002	12.81 ± 1.26	20.14 ± 1.99
S4	0.19 ± 0.03	0.12 ± 0.007	13.03 ± 1.89	20.49 ± 2.96

These results were not predicted by the simulations using the *i*MM904 genome-scale model, which supports the conclusion that the observed phenotypes for strains S3 and S4 are not related to flux constrains but to regulatory phenomena that cannot be predicted by stoichiometric models. To investigate further this hypothesis, the activities of the enzymes from the GDP-mannose pathway and pyruvate kinase were determined. The strains S2 and S4 were selected to compare the impact of disrupting genes *SER3* and *SER33* plus the modification in the *PYK1* promoter. Enzymatic activities were measured in cells at mid-exponential phase to show no significant differences between the two strains for the enzymes mannose-6-phosphate isomerase (PMI40p) and phosphomannomutase (SEC53p) (Figure 4.4). However, the activity of GDP-mannose pyrophosphorylase (PSA1p) decreased significantly for strain S4 in comparison to S2. This corresponds to a 4-fold reduction in the activity of the PSA1p, which must influence the GDP-mannose pool size.

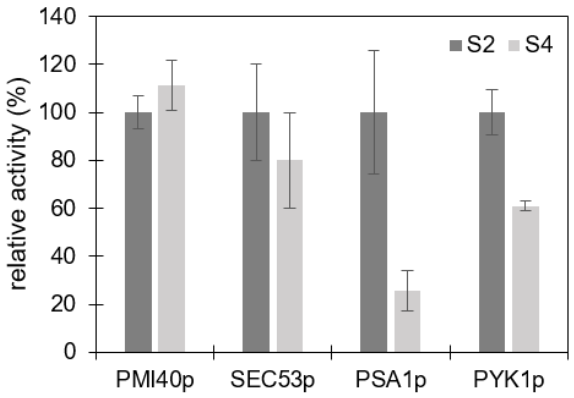

Figure 4.4 Activity of the enzymes mannose-6-phosphate isomerase (PMI40p), phosphomannomutase (SEC53p), GDP-mannose pyrophosphorylase (PSA1p) and pyruvate kinase (PYK1p) for strains S2 and S4. Enzymatic activities are shown as percentage of the respective activities determined for strain S2. Data are mean+/-SD from three independent measurements.

Pyruvate kinase activity was also measured to confirm the bottleneck applied by truncating the *PYK1* gene's promoter and revealed a 40% decrease for strain S4 in comparison to S2 (Figure 4.4). This decrease is consistent with literature reports (Nishizawa *et al.*, 1989), and proves the feasibility of this type of strategy for the manipulation of essential genes.

Table 4.6 Physiological parameters, MG titer and MG productivity for strains S2 and S4 cultivated in chemostat mode in SD medium with 20 $g.L^{-1}$ of glucose at dilutions of 0.1 and 0.05 h^{-1}. MG was extracted from cells with a methanol/chloroform mixture. Data represent mean ± SD in steady-state.

Strain	Dilution (h^{-1})	$Y_{X/S}$	Glucose ($g.L^{-1}$)	Ethanol ($g.L^{-1}$)	MG titer ($mg.L^{-1}$)	MG productivity ($mg.L^{-1}.h^{-1}$)
S2	0.10	0.10	5.3 ± 0.82	3.0 ± 0.4	34.2 ± 2.6	3.4 ± 0.3
	0.05	0.18	0	0	41.0 ± 5.1	2.0 ± 0.3
S4	0.10	0.17	1.14 ± 0.36	2.39 ± 0.4	11.2 ± 2.3	1.1 ± 0.2
	0.05	0.22	0	0	22.5 ± 1.8	1.0 ± 0.2

To assess MG productivity and physiological behavior of key strains, S2 and S4 were cultivated in overflow (D= 0.1 h^{-1}) and in full respiratory metabolism (D=0.05 h^{-1}). Chemostat cultivation is the best way to manipulate microbial metabolism for its simplicity to control different metabolic states (Nielsen *et al.*, 2003). For strain S2, overflow metabolism was achieved by cultivating cells at dilution 0.1 h^{-1}; there was production of ethanol (3 $g.L^{-1}$), acetate (0.7 $g.L^{-1}$) and glycerol (0.5 $g.L^{-1}$). Biomass yield in these conditions was 0.10 $g_{DW}.g_{glc}^{-1}$, MG titer and MG productivity were 34.2 $mg.L^{-1}$ and 3.4 $mg.L^{-1}.h^{-1}$, respectively (Table 4.6). No ethanol was observed when cells from strain S2 were cultivated at dilution 0.05 h^{-1}, although glycerol (0.9 $g.L^{-1}$) and acetate were produced (2.5 $g.L^{-1}$). Biomass yield and MG titer increased and MG productivity decreased in relation to dilution 0.1 h^{-1}. The same profile was obtained with strain S4 for biomass yield, acetate (1 vs 2.2 $g.L^{-1}$ for dilution 0.1 and 0.05 h^{-1}, respectively) and MG titer, glucose, ethanol and glycerol (1 vs 0.5 $g.L^{-1}$) as the dilution varies from 0.1 to 0.05 h^{-1}. MG productivity revealed to be independent of dilution (Table 4.6).

4.4 Discussion

The optimization of metabolic networks envisaging the production of target compounds often demands changes in different genes/pathways. As metabolic pathways are intimately connected, these alterations can often affect neighbor pathways, producing phenotypes that are difficult to predict. Genome-scale models are an excellent tool to predict and understand the implications of multiple genetic alterations. Multiple cases of success have been described for *in silico* driven solutions. That is the case of lycopene (Choi *et al.*, 2010) and L-valine (Park *et al.*, 2007) biosynthesis in engineered *E. coli* strains or cubebol (Asadollahi *et al.*, 2009), vanillin (Brochado *et al.*, 2010) and succinate (Otero *et al.*, 2013) in *S. cerevisiae*.

In this work, we used the genome-scale model *i*MM904 to identify target genes for improvement of MG yield. Different methods were used to search for the best strategy to achieve a higher MG yield. The curated solution aimed to improve flux towards GDP-mannose and to increase 3PG concentration in the cell. The over-expression of *PMI40* and *PSA1*, suggested by the simulating annealing algorithm, led to an increase of 2-fold on the MG production in comparison to strain S1. Strain S2 and MG02 from Chapter 3 have the same straightforward strategy of over-expressing *PMI40* and *PSA1*. However, S2 showed a better MG yield than MG02 (25.30 versus 15.86 $mg.g_{DW}^{-1}$, respectively). The difference in MG production between strain MG02 (Chapter 3) and strain S2 derives from the optimization in *mgsD* gene construction (see Methods section).

In addition to improve flux towards GDP-mannose, two strains were constructed with the intention of augmenting 3PG pool by disrupting the carbon flux towards the production of serine (Δ*SER3*Δ*SER33*) (strain S3) and by decreasing pyruvate kinase activity through downregulating *PYK1* expression (strain S4). By truncating the *PYK1* promoter we were able to reduce pyruvate kinase activity by 40% (Figure 4.4). Pearce and co-workers developed a similar construction to decrease *PYK1* gene transcription by substituting the *PYK1* promoter with a truncated version and called this strain YKC11 (Pearce *et al.*, 2001). YKC11 had reduced growth and glucose consumption and accumulated glucose-6-phosphate and fructose-6-phosphate while having reduced pools of fructose-1,6-biphosphate and trehalose-6-phosphate. In theory, the accumulation of glucose-6-phosphate and fructose-6-phosphate and the bottleneck on pyruvate kinase would favor the synthesis of MG by increasing the precursors 3PG and GDP-mannose (produced from fructose-6-phosphate). Strain YKC11 also presented a reduction on the production of ethanol and glycerol, as observed in this work with strain S4. However, this strategy of increasing 3PG concentration did not improve MG yield. To understand why MG production in strain S4 was lower than in strain S2, the activity of enzymes belonging to the GDP-mannose pathway were measured (Figure 4.4). PMI40p and SEC53p activity revealed to remain unchanged in strains S2 and S4, while the activity PSA1p showed a drastic decrease for strain S4, in comparison with S2. These results help to explain the reduced production of MG in strain S4, as a strong depression of PSA1p reduces the production of GDP-mannose in the cell (Janik *et al.*, 2003) that will ultimately compromise the production of MG. Activity levels of PSA1p for strain S3 were not measured but we suspect that this enzyme also has an reduced activity based on the decrease of MG production. In light of these results, it is reasonable to conclude that PSA1p activity in strains S3 and S4 is being affected by the alterations made on the glycolytic pathway.

An understanding of what might be influencing PSA1p activity is difficult to infer as information on PSA1p regulation is very limited. However, we postulate two scenarios. First, alterations on serine pathway and the reduced activity of PYK1p activated a regulatory network that targets PSA1p. The regulatory network of *S. cerevisiae* is a very complex system and can take place at different levels. Cells control protein synthesis based on feedback obtained from environmental or genetic conditions. This action starts with gene expression through modulation of the chromatin structure (Wyrick *et al.*, 1999) and gene transcription (O'Malley *et al.*, 1977). Next point of control is on mRNA pools which are regulated through modifications on mRNA and by the translation

machinery (Filipowicz *et al.*, 2008). Once proteins are synthesized, post-translational modifications dictate the fate of these molecules regarding activation / inhibition or even destruction (Mann & Jensen, 2003). PSAp activity can be influenced by each one of these regulation mechanisms or even a combination of more than one. Another explanation to PSA1p reduced activity may derive from the reduced growth rate of strains S3 and S4 in relation to strain S1 which might affect *PSA1* transcription levels, that are intimately connected to cell division. It is well known that PSA1p synthesizes GDP-mannose to be incorporated into N-linked and O-linked glycoproteins and its deletion is lethal to yeast cells (Hashimoto *et al.*, 1997). Moreover, hypomorphic mutants exhibit a variety of defects on cell wall biosynthesis, sensitivity to osmotic stress, lack of transformation competence and leakage of cell surface proteins (Yoda *et al.*, 2000). Transcription of *PSA1* is regulated by cyclins of G1 cell cycle that control budding division (Benton *et al.*, 1996). Even more, *PSA1* transcription is positively regulated by GTP levels inside the cell (Kuehner & Brow, 2008; Shimma *et al.*, 1997). The link between cell growth / proliferation and *PSA1* was established by Melamed and co-workers when transcript levels of cells were analyzed in response to high salinity. Under these conditions, transcript levels of *PSA1* and of other genes involved in cell-wall metabolism were highly reduced in comparison with optimal growth conditions (Melamed *et al.*, 2008). In the present work, we have observed that strains S3 and S4 show a prolonged lag phase and reduced growth rate, which, based on the studies referred above, may be directly related with lower levels of *PSA1* transcription.

Clearly, GDP-mannose pathway is highly regulated and connected to the glycolic pathway and cell growth. New strategies to improve the production of MG must have into consideration the impact on growth rate of *S. cerevisiae.*

Despite the success that computational tools have in the development of quality of solutions to be applied in metabolic engineering, they have failed to solve our metabolic problem. By applying the *in silico* solutions to *S. cerevisiae* we learned that MG synthesis is dependent of a regulatory network surrounding GDP-mannose. Genome-scale models are built on stoichiometric-based calculations and do not account for regulatory phenomena. However, it is important to stress out that the solutions obtained by the simutating annealing algorithm to increase MG yield were simple and robust and had a literature support, as *PSA1* over-expression increases GDP-mannose pool in *S. cerevisiae* (Janik *et al.*, 2003) and PYK1 down-expression increases fructose-6-phosphate pool (Pearce *et al.*, 2001).

In the future, it is expectable that genome-scale models evolve to harbor information regarding gene expression and regulation. The first ME-model for *E. coli* has been published (O'Brien *et al.*, 2013) and combines the stoichiometry of a genome-scale model with gene expression for a particular environmental condition or/and genetic alteration. These ME-models are a step to improve phenotype predictions and might be a solution to solve problems similar to what is presented here.

In this work, the *in silico* solutions predicted an increased MG production by targeting the over-expression of *PMI40* and *PSA1* and by disrupting reactions that consume 3PG (Table 4.4). From the applied strategies, only the overexpression of the GDP-mannose pathway (strain S2) resulted

in an enhanced MG production (Figure 4.3 and Table 4.5). Continuous mode production of MG with strain S2 shows better results with higher dilution rates, as observed in Chapter 3 (Table 4.6). These results reinforce the connection between cell division and GDP-mannose formation, as discussed above. Regarding strain S4, MG productivity in continuous mode seems to be unrelated with the dilution rate and may be due to the reduced activity measured for PSA1p. Clearly, while strain S2 served as a successful demonstration of systems biology-based strategies, new rounds of optimization are needed to increase MG productivity capable of competing with the existing processes.

Chapter 5

General discussion and future perspectives

The discovery of organisms capable of thriving in adverse environments was received by the scientific community with astonishment and excitement. Since then, multiple questions have been asked: How can these organisms survive and multiply? What are the adaptation strategies? A mechanism related with adaptation to extreme habitats is the synthesis of novel compounds with superior ability to protect cellular components against heat and osmotic stresses (da Costa *et al.*, 1998). These compounds are called compatible solutes and they accumulate intracellulary, sometimes in high concentration, without interfering with cell physiology. Mannosyglycerate (MG) is the most widespread solute found in marine (hyper)thermophiles and for its remarkable ability to stabilize proteins it became pertinent to develop efficient producers so that MG could be commercialized at competitive prices and widely used.

The main goals of this thesis are: i) to evaluate the *in vivo* role of MG regarding the stabilization of proteins; and ii) to construct a cell factory to permit a wider commercialization and utilization of MG.
To evaluate MG ability to protect proteins in the intracellular milieu, a new strain capable of producing MG and expressing α-synuclein (α-Syn) was constructed and compared to a strain expressing α-Syn only. Encouraged by these results, we set to develop a cost-effective bioprocess to produce MG. *Saccharomyces cerevisiae* was the industrial microorganism selected to accommodate the MG pathway.

MG acted as an inhibitor of the aggregation of α-synuclein (a protein involved in Parkinson´s disease) in the intracellular milieu of the yeast S. cerevisiae*.*

The first yeast model of Parkinson disease (PD) was published in 2003 by Outeiro and Lindquist (Outeiro & Lindquist, 2003) who showed that the mechanisms underlining α-Syn expression in humans can be partially represented in the unicellular fungi – *S. cerevisiae.* Although this model was developed with the purpose of finding therapies to treat PD, it became a tool to understand the phases of protein aggregation inside cells. For this reason, we selected this yeast model of PD to prove, for the first time, that MG is able to protect proteins inside the intracellular space, and to gain insight into the mechanisms underlying the ability of MG to reduce cell toxicity induced by α-Syn aggregates. Even more, we venture on the potential of MG as a lead compound for the development of drugs against misfolding-related neurodegenerative diseases.

As referred in Chapter 1, marine (hyper)thermophiles accumulate MG mainly in response to osmotic stress. Exceptions to this behavior are observed in *Rhodothermus marinus* and *Palaeococcus ferrophilus*; these organisms accumulate MG in response to osmotic and thermal stresses (Neves *et al.*, 2005; Silva *et al.*, 1999). However, *in vitro* studies have demonstrated that MG is capable of protecting a variety of model proteins against thermal denaturation and aggregation, such as lactate dehydrogenase, glutamate and alcohol dehydrogenase (Ramos *et*

al., 1997); malate dehydrogenase, nuclease A, and lysozyme (Faria *et al.*, 2008). Taking these results into consideration it is feasible to postulate a correlation between MG accumulation and protection of cell components against heat damage. This hypothesis is strengthened by the results obtained in Chapter 2. Cells expressing α-Syn and accumulating MG showed a clear reduction of the aggregation of this prone-aggregating protein. The same inhibition effect was observed when α-Syn was incubated with 0.1 M of MG *in vitro*.

The protein α-Syn is highly expressed in the human brain, especially in the dopaminergic neurons that control multiple brain functions, including voluntary movement and a broad array of behavioral processes such as mood, reward, addiction, and stress (Chinta & Andersen, 2005; Stefanis, 2012). Although its function remains unclear, α-Syn is localized specifically in the nerve terminal of the neuron and is associated with the localization of developing synapses, arriving once the synaptic vesicle is matured. Multiple factors seem to contribute to the development of PD, from genetic alterations to environmental conditions. However, the most critical factor appears to be age. As we grow old, mitochondrial dysfunction increases and the quality of protein degradation systems degrades (e.g. ubiquitin proteasome system and autophagy), disrupting cell homeostasis and promoting cell death (Reeve *et al.*, 2014). Familial forms of PD correlate with the existence of mutations and duplications of α-Syn gene and mutations in proteins from complex I of mitochondria (e.g. PINK1, PARK2 and PARK7 genes) or in protein quality control systems (e.g. LRRK2 gene) to cite a few examples (Trinh & Farrer, 2013).

Using a yeast model of PD we showed that MG significantly reduced α-Syn aggregation. Such effect can arise either by direct stabilization of the native α-Syn structure, or by stimulating protein degradation systems. The later hypothesis was tested by measuring the levels of chaperones which are able to refold aberrant forms of α-Syn. Additionally, α-Syn clearance and proteasome activity were analyzed once protein synthesis was arrested. In both experiments, MG accumulation did not affect chaperone expression or proteasome activity, strengthening the hypothesis that MG acted directly on α-Syn structure. In fact, it was demonstrated that MG is able to reduce α-Syn aggregation *in vitro*. A characteristic of PD at the cellular level is the significant increase of ROS levels due to elevated toxicity of α-Syn inclusions. This is indeed corroborated by our observations; most importantly, we showed that this stress indicator dropped in the presence of MG. We concluded that MG was able to reduce α-Syn fibrillation *in vitro* and to decrease ROS levels in yeast without affecting the protein degradation systems. So, the remaining question was: Could MG act as a molecular chaperone *in vivo*? To check this hypothesis, α-Syn oligomeric species were separated by size in cells with and without MG. Results show that cells accumulating MG have α-Syn oligomers of smaller size than cells without MG, indicating that this osmolyte interferes with the early stages of the fibrillation process preventing further polymerization.

In conclusion, this work brings new insights into the role of MG *in vivo*: This solute acts as a chemical chaperone and the stabilization mechanism involves direct solute/protein interactions. It also points to new fields for MG applications and emphasizes the need for development of an efficient industrial bioprocess.

Increasing flux towards GDP-mannose in S. cerevisiae *improves the production of mannosylglycerate*

Microorganisms can produce several compounds at high levels, which make them excellent hosts to produce desired chemicals at large-scale. Some of these microorganisms are natural-producers and to increase the production level, industrial bioprocesses are developed by optimizing growth conditions and medium composition. On the other hand, the development of genetic manipulation tools has opened the way for the construction of cells tailored to produce new compounds (Lee *et al.*, 2009). Some organisms are capable of growing fast up to high density with low medium requirements and show high resilience to stressful conditions, which makes them excellent cell factories. One of these organisms is *S. cerevisiae.* This budding yeast is used for centuries in the production of bread and beer, and is one of the most important cell factories. For its importance, a wide-range of molecular biology tools were developed and now are used to produce commodity compounds such as ethanol, or high-price pharmaceutical products such as artesimic acid (Kavšček *et al.*, 2015; Nevoigt, 2008).

In this thesis, we set to optimize the production of MG using *S. cerevisiae* as a host. Empadinhas and co-workers reported the accumulation of MG by engineering *S. cerevisiae* with a gene encoding the mannosyl-3-phosphoglycerate synthase/phosphatase from *Dehalococcoides mccartyi* (Empadinhas *et al.*, 2004). This mesophilic, neutrophilic, and gram-positive bacteria lives in a consortium with other species, where it has an important role in the bioremediation of groundwater (Loffler *et al.*, 2013). Homologues of mannosyl-3-phosphoglycerate synthase (MPGS) from *D. mccartyi* were found in different branches of the Tree of Life: from Bacteria like *Thermus thermophilus* or *Rhodothermus marinus* (MG industrial producer), to Fungi (e.g. *Sordaria macrospora* or *Actinoplanes friuliensis*) in which MG accumulation was never observed, or in two phyla of Archaea: Crenarchaeota (e.g. *Aeropyrum pernix* and *Staphylothermus marinus)* and Euryarchaeota (e.g. *Palaeococcus ferrophilus* or *Pyrococcus abyssi*) (Borges *et al.*, 2014).

Another critical task in this work was to choose the most appropriate strain of *S. cerevisiae*. The strain CEN.PK2-1C, a laboratory generated strain, was selected for having: i) high growth rates in complete and defined media under aerobic and anaerobic conditions in batch, fed-batch and continuous culture; ii) high mating and sporulation efficiency; iii) high transformation efficiency; and iv) good single-cell formation (no flocculation), which makes this strain the ideal candidate to study metabolic fluxes and to engineer for production of a variety of chemicals (Stark & Stansfield,

2007). MG is produced from GDP-mannose and 3-phosphoglycerate (3PG). Our first approach to improve MG production was to increase the flux to GDP-mannose production from fructose-6-phosphate. To achieve this goal we cloned the genes *PMI40* and *PSA1* (coding for mannose-6-phosphate isomerase and GDP-mannose pyrophosphorylase, respectively) in a multi-copy plasmid with stronger promoters and terminators. As the cultivation mode greatly influences the production of compounds we have tested growth in shake flasks, batch and continuous mode (Chapter 3). The overexpression of *PMI40* and *PSA1* led to a 1.5-fold increase on MG yield in shake flask and 2.2-fold increase in batch cultivation, corresponding to a maximal yield of 15.86 mg of MG per g of DW (0.13 µmol of MG per mg of protein).

The continuous mode cultivation of *S. cerevisiae* has been applied to the production of many compounds (Nielsen *et al.*, 2003). That is the case of resveratrol (Vos *et al.*, 2015), polyhydroxybutyrate (Kocharin & Nielsen, 2013) or vanillin (Brochado *et al.*, 2010). To the end of developing an efficient bioprocess to produce MG, we set to test dilution rates and determine the correspondent MG productivity. Our results show that MG productivity is positively connected with higher growth rates, as the best result was obtained with a dilution of 0.15, corresponding to 1.79 mg of MG per DW per hour. This correlation may be due to the fact that MG precursors are derived from glycolysis, as is the case of GDP-mannose (produced from fructose-6-phosphate), or are actually glycolytic intermediates, as for 3PG. Therefore, the synthesis of MG is favored by the high glycolytic fluxes obtained with elevated growth rates.

Improving MG production in S. cerevisiae *through in silico design*

The metabolic networks of cells are a product of thousands of millions of years of evolution “engineering”. Disrupting these optimized networks to serve our purposes is therefore a daunting goal. The construction of constrained-based models at a genomic scale allied with computational tools that try to mimic cell behavior is boosting the area by producing holistic strategies that will redirect metabolism towards the production of a particular product (Otero & Nielsen, 2010).

Multiple cases of success have been achieved by applying *in silico* design strategies to improve product titer in microorganisms. For this reason, we set to use the most robust computational tools to design a *S. cerevisiae* strain that would produce high levels of MG. These tools account for methods on simulation of phenotypes and optimization of strains which allied with evolutionary algorithms produce solutions to improve the *in silico* production of MG. To run these methods and algorithms we took advantage of Optlux (Rocha *et al.*, 2010). This is a computer program developed in Isabel Rocha group that has several tools implemented. The solutions obtained and described in chapter 4 are a result of curated outputs from different conjugations of method / algorithm (Gonçalves *et al.*, 2012; Maia *et al.*, 2016).

All methods tested pointed out two strategies: 1) increase the flux towards the formation of GDP-mannose; and 2) accumulation of 3PG. These two strategies were achieved by over-expression *PSA1* and *PMI40* genes using a multi-copy plasmid and by truncating the flux from phosphoenolpyruvate to pyruvate (Chapter 4). Additionally, the expression of the gene *mgsD* (pDES) was improved by adding a strong ribosome binding site and a synthetic terminator. The overexpression of the GDP-mannose pathway genes led to a 2-fold increase in MG production to 25.3 mg of MG per g of DW, in controlled batch mode.

Table 5.1 MG yield and productivity observed in the strains MG01 and MG02 (Chapter 3) and strains S2 and S4 (Chapter 4). Data from chemostat correspond to the highest dilution rates tested.

	Cultivation mode		
Strains	Batch ($mg_{MG}.g_{DW}$)	Chemostat ($mg.g_{DW}^{-1}.h^{-1}$)	Reference
MG01 (*mgsD*)	7.08	1.22	Chapter 3
MG02 (*mgsD* ↑*PSA1* ↑*PMI40*)	15.86	1.79	Chapter 3
S2 (*mgsD* ↑*PSA1* ↑*PMI40*)	25.30	2.31	Chapter 4
S4 (*mgsD* ↑*PSA1* ↑*PMI40* Δ*SER3* Δ*SER33* $Pykp_{\Delta 653}$-PYK)	13.03	0.34	Chapter 4

To improve the availability of 3PG, a disruption of the genes *SER3* and *SER33* (deviates flux towards serine production) and a down-expression of the *PYK1* (limits pyruvate formation) was performed. Strangely, MG production decreased to values obtained by the reference strain, which harbored only the *mgsD* gene encoding the bifuncional enzyme mannosyl-3-phosphoglycerate synthase/phosphatase. Curiously, changing the glycolytic flux at the level of pyruvate kinase seemed to have a profound effect on the GDP-mannose pathway. This was confirmed by evaluating the activity of the enzymes mannose-6-phosphate isomerase (PMI40p), phosphomannomutase (SEC53p) and GDP-mannose pyrophosphorylase (PSA1p). These enzymes are responsible for the conversion of fructose-6-phosphate from glycolysis into GDP-mannose. Indeed, the alterations in the glycolytic pathway led to 40% reduction in the activity of PSA1p, which likely affects the production of GDP-mannose and consequently explains the decrease in MG production. A list of the most relevant yields and productivities of MG obtained in these thesis is summarized in Table 5.1.

The phenotypic simulations of mutant strains using genome-scale models can only predict flux changes on the network and cannot foresee regulatory layers that are activated in response to genetic alterations. In order to predict the regulatory mechanisms from a cell, genome-scale models must integrate regulatory mechanisms acting on transcriptional, post-transcriptional, translational and post-translational levels (Faria *et al.*, 2014). The first model of integrated metabolic and regulatory networks of *E. coli* was developed by Covert and co-workers and

established that regulation significantly affects the quality of predictions for growth phenotypes and gene expression experiments (Covert *et al.*, 2004). Herrgård and co-workers also developed a model for *S. cerevisiae* that integrated a nutrient-controlled transcriptional regulatory data with a genome-scale model and identified regulatory cascades connecting transcription factors to target genes. This combined model is also capable of predicting growth phenotypes of knock out transcription factor strains (Herrgård, 2006). In the future, we expect that these computational models will enable catching the complexity of biological processes. The task of integrating these different models is challenging the scientific community. At the moment, metabolic and regulatory models are being combined from high-throughput data with the help of new algorithms. These models will be useful to understand the complexity of the biological processes, and will become an indispensable tool to design experimental protocols and also to increase the rate of biotechnology successes (Imam *et al.*, 2015).

MG is naturally produced by several organisms that grow optimally above 60ºC. Currently, MG is extracted from *Rhodothermus marinus* with a maximal yield of 0.8 $\mu mol.mg_{protein}^{-1}$ (Nunes *et al.*, 1995). A trehalose-deficient *Thermus thermophilus* RQ-1 mutant is able to accumulate 0.6 $\mu mol.mg_{protein}^{-1}$ of MG in a fed-batch cultivation (Egorova *et al.*, 2007). *S. cerevisiae* strain S2 (Chapter 4) produces 0.21 $\mu mol.mg_{protein}^{-1}$ of MG (considering 0.45 g protein per g of biomass (Lange & Heijnen, 2001)).

The MG-producing yeast (S2 strain) showed low productivity yield, but possesses several advantages in comparison to natural producers, namely: 1) growth at 30ºC, a moderate temperature that contrasts with the temperature used to cultivate the natural producers (between 65 and 100ºC); 2) the organism is easily grown with minimal nutrient requirements and 3) there are various bioprocesses already implemented in *S. cerevisiae* to product a wide range of products (Kavšček *et al.*, 2015; Krivoruchko & Nielsen, 2014)

Recommended future work

In this thesis, we were able to increase the production of MG in *S. cerevisiae*. However, the production is still far from that of the natural producers. New rounds of optimization are needed to improve MG production, such as:

1. Improving the flux to GDP-mannose by inserting new copies of the genes *PSA1* and *PMI40* in the genome. Additionally, a search for heterologous proteins with higher activity could also lead to an improved production of GDP-mannose;
2. Characterizing the regulatory network that controls the synthesis of GDP-mannose and 3PG. This data is important to select the best genetic manipulation to channel the glycolytic flux towards the MG precursors;

3. Searching and charactering new mesophilic mannosy-3-phosglycerate synthase/phosphatase to find enzymes with higher activity towards the synthesis of MG;
4. Construction and characterizing a *S. cerevisiae* expressing the mannosylglycerate synthase from plants (single-step pathway for MG). However, in this case we would also have to express a plant enzyme to convert 3PG into D-glycerate since there is no evidence that *S. cerevisiae* has such enzyme, either on SGD (Cherry *et al.*, 2012), YMDB (Jewison *et al.*, 2012) or Brenda (Schomburg *et al.*, 2004).
5. In terms of operating mode, it would be important to culture *S. cerevisiae* MG-producing strains in a fed-batch mode, and compare the yields with the continuous mode.

References

ActionAid. **(2010).** Meals per gallon. *Public Adm Dev* **17**, 142–410.

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I. & Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* **9,** (3): e1002980

Aguib, Y., Heiseke, A., Gilch, S., Riemer, C., Baier, M., Schätzl, H. M. & Ertmer, A. (2009). Autophagy induction by trehalose counteracts cellular prion infection. *Autophagy* **5**, 361–9.

Akada, R., Kitagawa, T., Kaneko, S., Toyonaga, D., Ito, S., Kakihara, Y., Hoshida, H., Morimura, S., Kondo, A. & Kida, K. (2006). PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*. *Yeast* **23**, 399–405.

Alarico, S., Empadinhas, N., Mingote, A., Simões, C., Santos, M. S. & da Costa, M. S. (2007). Mannosylglycerate is essential for osmotic adjustment in *Thermus thermophilus* strains HB27 and RQ-1. *Extremophiles* **11**, 833–40.

Alia, Hayashi, H., Sakamoto, A. & Murata, N. (1998). Enhancement of the tolerance of *Arabidopsis* to high temperatures by genetic engineering of the synthesis of glycinebetaine. *Plant J* **16**, 155–61.

Alper, H., Jin, Y.-S., Moxley, J. F. & Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* **7**, 155–64.

Anahas, A. M. P. & Muralitharan, G. (2015). Isolation and screening of heterocystous cyanobacterial strains for biodiesel production by evaluating the fuel properties from fatty acid methyl ester (FAME) profiles. *Bioresour Technol* **184**, 9–17.

Anastas, P. & Warner, J. (1998). *Green Chemistry: Principles and Practice*. Oxford University Press: New York.

Arora, A., Ha, C. & Park, C. B. (2004). Inhibition of insulin amyloid formation by small stress molecules. *FEBS Lett* **564**, 121–125.

Asadollahi, M. A., Maury, J., Patil, K. R., Schalk, M., Clark, A. & Nielsen, J. (2009). Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. *Metab Eng* **11**, 328–34.

Ascêncio, S. D., Orsato, A., França, R. A., Duarte, M. E. R. & Noseda, M. D. (2006). Complete 1H and 13C NMR assignment of digeneaside, a low-molecular-mass carbohydrate produced by red seaweeds. *Carbohydr Res* **341**, 677–682.

Atsumi, S., Hanai, T. & Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–9.

Bates, J. T., Chivian, D. & Arkin, A. P. (2011). GLAMM: Genome-Linked Application for Metabolic Maps. *Nucleic Acids Res* **39**, W400-5.

Bautista, P., Mohedano, A. F., Gilarranz, M. A., Casas, J. A. & Rodriguez, J. J. (2007). Application of Fenton oxidation to cosmetic wastewaters treatment. *J Hazard Mater* **143**, 128–34.

Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø. & Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* **2**, 727–38.

Becker, S. A. & Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4**, e1000082.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acids Res* **41**, D36-42.

Benton, B. K., Plump, S. D., Roos, J., Lennarz, W. J. & Cross, F. R. (1996). Over-expression of *S. cerevisiae* G1 cyclins restores the viability of alg 1 N-glycosylation mutants. *Curr Genet* **29**, 106–113.

Berger, R. G. (2007). *Flavours and Fragrances: Chemistry, Bioprocessing and Sustainability*, 1st edn. (R. G. Berger, Ed.). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bergmeyer, H. U. & Gawehn, K. (1974). *Methods of enzymatic analysis Volume 2* (V. Chemie, Ed.). Academic Press.

Bernier, V., Lagacé, M., Bichet, D. G. & Bouvier, M. (2004). Pharmacological chaperones: potential treatment for conformational diseases. *Trends Endocrinol Metab* **15**, 222–228.

Lo Bianco, C., Shorter, J., Régulier, E., Lashuel, H., Iwatsubo, T., Lindquist, S. & Aebischer, P. (2008). Hsp104 antagonizes α-synuclein aggregation and reduces dopaminergic degeneration in a rat model of Parkinson disease. *J Clin Invest* **118**, 3087–3097.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K. & other authors. **(1997).** The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–62.

Blöchl, E., Rachel, R., Burggraf, S., Hafenbradl, D., Jannasch, H. W. & Stetter, K. O. (1997). *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**, 14–21.

Bolen, D. & Baskakov, I. V. **(2001).** The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J Mol Biol* **310**, 955–963.

Borges, N., Ramos, A., Raven, N. D. H., Sharp, R. J. & Santos, H. (2002). Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes. *Extremophiles* **6**, 209–16.

Borges, N., Jorge, C. D., Gonçalves, L. G., Gonçalves, S., Matias, P. M. & Santos, H. (2014). Mannosylglycerate: structural analysis of biosynthesis and evolutionary history. *Extremophiles* **18**, 835–852.

Borodina, I. & Nielsen, J. (2014). Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. *Biotechnol J* **9**, 609–620.

Borwankar, T., Röthlein, C., Zhang, G., Techen, A., Dosche, C. & Ignatova, Z. (2011). Natural Osmolytes Remodel the Aggregation Pathway of Mutant Huntingtin Exon 1. *Biochemistry* **50**, 2048–2060.

Botstein, D., Chervitz, S. A. & Cherry, J. M. (1997). Yeast as a model organism. *Science* **277**, 1259–60.

Bouveng, H., Lindberg, B. & Wickberg, B. (1955). Low-molecular carbohydrates in algae. *Acta chem scand*.

Bradfield, M. F. A. & Nicol, W. (2016). Continuous succinic acid production from xylose by *Actinobacillus succinogenes*. *Bioprocess Biosyst Eng* **39**, 233–44.

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

Brethauer, S. & Studer, M. H. (2015). Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals – A Review. *Chim Int J Chem* **69**, 572–581.

Brochado, A. R., Matos, C., Møller, B. L., Hansen, J., Mortensen, U. H. & Patil, K. R. (2010). Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* **9**, 84.

Brochado, A. R., Andrejev, S., Maranas, C. D. & Patil, K. R. (2012). Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput Biol* **8**, e1002758.

Brock, T. D. (1967). Life at high temperatures. Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed. *Science* **158**, 1012–9.

Broeckling, C. D., Reddy, I. R., Duran, A. L., Zhao, X. & Sumner, L. W. (2006). MET-IDEA: data extraction tool for mass spectrometry-based metabolomics. *Anal Chem* **78**, 4334–41.

Brown, A. D. (1976). Microbial water stress. *Bacteriol Rev* **40**, 803–46.

Buenger, J. & Driller, H. (2004). Ectoin: an effective natural substance to prevent UVA-induced premature photoaging. *Skin Pharmacol Physiol* **17**, 232–7.

Burgard, A. P., Pharkya, P. & Maranas, C. D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84**, 647–57.

Caldas, T., Demont-Caulet, N., Ghazi, A. & Richarme, G. (1999). Thermoprotection by glycine betaine and choline. *Microbiology* **145**, 2543–8.

Carreto, L., Moore, E., Nobre, M. F., Wait, R., Riley, P. W., Sharp, R. J. & da Costa, M. S. (1996). *Rubrobacter xylanophilus* sp. nov., a New Thermophilic Species Isolated from a Thermally Polluted Effluent. *Int J Syst Bacteriol* **46**, 460–465.

Carroll, A. L., Desai, S. H. & Atsumi, S. (2015). Microbial production of scent and flavor compounds. *Curr Opin Biotechnol* **37**, 8–15.

Cerbón, J., Falcon, A., Hernández-Luna, C. & Segura-Cobos, D. (2005). Inositol phosphoceramide synthase is a regulator of intracellular levels of diacylglycerol and ceramide during the G1 to S transition in *Saccharomyces cerevisiae*. *Biochem J* **388**, 169–76.

Cerdà-Costa, N., Esteras-Chopo, A., Avilés, F. X., Serrano, L. & Villegas, V. (2007). Early kinetics of amyloid fibril formation reveals conformational reorganisation of initial aggregates. *J Mol Biol* **366**, 1351–63.

Chen, D.-H., Brkanac, Z., Verlinde, C. L. M. J., Tan, X.-J., Bylenok, L., Nochlin, D., Matsushita, M., Lipe, H., Wolff, J. & other authors. **(2003).** Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. *Am J Hum Genet* **72**, 839–49.

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S. & other authors. **(2012).** *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-5.

Cherubini, F. (2010). The biorefinery concept: Using biomass instead of oil for producing energy and chemicals. *Energy Convers Manag* **51**, 1412–1421.

Chinta, S. J. & Andersen, J. K. (2005). Dopaminergic neurons. *Int J Biochem Cell Biol* **37**, 942–946.

Choi, H. S., Lee, S. Y., Kim, T. Y. & Woo, H. M. (2010). *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* **76**, 3097–3105.

Christensen, B. & Nielsen, J. (2000). Metabolic network analysis. A powerful tool in metabolic engineering. *Adv Biochem Eng Biotechnol* **66**, 209–31.

Christensen, C. H., Rass-Hansen, J., Marsden, C. C., Taarning, E. & Egeblad, K. (2008). The Renewable Chemicals Industry. *ChemSusChem* **1**, 283–289.

Clark, J., Farmer, T., Hunt, A. & Sherwood, J. (2015). Opportunities for Bio-Based Solvents Created as Petrochemical and Fuel Products Transition towards Renewable Resources. *Int J Mol Sci* **16**, 17101–17159.

Claude, A., Bondu, S., Michaud, F., Bourgougnon, N. & Deslandes, E. (2009). X-ray structure of a sodium salt of digeneaside isolated from red alga *Ceramium botryocarpum*. *Carbohydr Res* **344**, 707–10.

Colijn, C., Brandes, A., Zucker, J., Lun, D. S., Weiner, B., Farhat, M. R., Cheng, T.-Y., Moody, D. B., Murray, M. & Galagan, J. E. (2009). Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol* **5**.

Colin, H. & Augier, J. (1939). Un glucide original chez les floridées du genre Polysiphonia le d-mannoside de l-glycérate de sodium. *C R Acad Sci* **208**, 1450–1453.

Collins, S. B., Reznik, E. & Segrè, D. (2012). Temporal expression-based analysis of metabolism. *PLoS Comput Biol* **8**, e1002781.

Consulting, P. M. (2010). *Report: Assessment of the Bio-based Products Market Potential for Innovation*.

da Costa, M. S., Empadinhas, N., Albuquerque, L. & Santos, H. (2003). Identification of a bifunctional gene for mannosylglycerate synthesis and development of a high-scale-production heterologous system based on *Saccharomyces cerevisiae*. Patente Internacional submetida em 17/10/2003 (processo nº 3398007Y) Requerent: BIOCANT. Abandonada.

da Costa, M. S., Santos, H. & Galinski, E. a. **(1998).** An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv Biochem Eng Biotechnol* **61**, 117–153.

Cotten, C. & Reed, J. L. (2013). Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnol J* **8**, 595–604.

Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96.

Crabtree, H. G. (1929). Observations on the carbohydrate metabolism of tumours. *Biochem J* **23**, 536–45.

Cruz, P. E., Silva, A. C., Roldao, A., Carmo, M., Carrondo, M. J. T. & Alves, P. M. (2006). Screening of Novel Excipients for Improving the Stability of Retroviral and Adenoviral Vectors. *Biotechnol Prog* **22**, 568–576.

Curran, K. a., Morse, N. J., Markham, K. a., Wagman, A. M., Gupta, A. & Alper, H. S. (2015). Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth Biol* **4**, 824–32.

Danzer, K. M. & McLean, P. J. (2011). Drug targets from genetics: α-synuclein. *CNS Neurol Disord Drug Targets* **10**, 712–23.

Das, R. K. & Brar, S. K. (2014). Enhanced fumaric acid production from brewery wastewater and insight into the morphology of *Rhizopus oryzae* 1526. *Appl Biochem Biotechnol* **172**, 2974–88.

Davies, J. E., Sarkar, S. & Rubinsztein, D. C. (2006). Trehalose reduces aggregate formation and delays pathology in a transgenic mouse model of oculopharyngeal muscular dystrophy. *Hum Mol Genet* **15**, 23–31.

Demain, A. L. & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* **27**, 297–306.

DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–6.

Dias, O., Rocha, M., Ferreira, E. C. & Rocha, I. (2015). Reconstructing genome-scale metabolic models with Merlin. *Nucleic Acids Res* **43**, 3899–3910.

Dower, W. J., Miller, J. F. & Ragsdale, C. W. (1988). High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Res* **16**, 6127–6145.

Edwards, J. S. & Palsson, B. O. (1999). Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem* **274**, 17410–6.

Eggert, A., Nitschke, U., West, J. A., Michalik, D. & Karsten, U. (2007). Acclimation of the intertidal red alga *Bangiopsis subsimplex* (Stylonematophyceae) to salinity changes. *J Exp Mar Bio Ecol* **343**, 176–186.

Egorova, K., Grudieva, T., Morinez, C., Kube, J., Santos, H., da Costa, M. S. & Antranikian, G. (2007). High yield of mannosylglycerate production by upshock fermentation and bacterial milking of trehalose-deficient mutant *Thermus thermophilus* RQ-1. *Appl Microbiol Biotechnol* **75**, 1039–45.

Elbein, A. D. (2003). New insights on trehalose: a multifunctional molecule. *Glycobiology* **13**, 17R–27.

Empadinhas, N., Marugg, J. D., Borges, N., Santos, H. & da Costa, M. S. (2001). Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*. Biochemical and genetic characterization of key enzymes. *J Biol Chem* **276**, 43580–8.

Empadinhas, N. (2004). *Pathways for the synthesis of mannosylglycerate in prokaryotes: genes, enzymes and evolutionary implications.* PhD thesis. University of Coimbra.

Empadinhas, N. & da Costa, M. S. (2011). Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ Microbiol* **13**, 2056–2077.

Empadinhas, N., Albuquerque, L., Henne, A., Santos, H. & da Costa, M. S. (2003). The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate. *Appl Environ Microbiol* **69**, 3272–9.

Empadinhas, N., Albuquerque, L., Costa, J., Zinder, S. H., Santos, M. a S., Santos, H. & Da Costa, M. S. (2004). A gene from the mesophilic bacterium *Dehalococcoides ethenogenes* encodes a novel mannosylglycerate synthase. *J Bacteriol* **186**, 4075–4084.

Empadinhas, N., Mendes, V., Simões, C., Santos, M. S., Mingote, A., Lamosa, P., Santos, H. & Costa, M. S. da. (2007). Organic solutes in *Rubrobacter xylanophilus*: the first example of di-myo-inositol-phosphate in a thermophile. *Extremophiles* **11**, 667–673.

Esteves, A. M., Chandrayan, S. K., McTernan, P. M., Borges, N., Adams, M. W. W. & Santos, H. (2014). Mannosylglycerate and di-myo-inositol phosphate have interchangeable roles during adaptation of *Pyrococcus furiosus* to heat stress. *Appl Environ Microbiol* **80**, 4226–33.

Faria, C., Jorge, C. D., Borges, N., Tenreiro, S., Outeiro, T. F. & Santos, H. (2013). Inhibition of formation of α-synuclein inclusions by mannosylglycerate in a yeast model of Parkinson's disease. *Biochim Biophys Acta* **1830**, 4065–72.

Faria, J. P., Overbeek, R., Xia, F., Rocha, M., Rocha, I. & Henry, C. S. (2014). Genome-scale bacterial transcriptional regulatory networks: reconstruction and integrated analysis with metabolic models. *Brief Bioinform* **15**, 592–611.

Faria, T. Q., Knapp, S., Ladenstein, R., Maçanita, A. L. & Santos, H. (2003). Protein stabilisation by compatible solutes: Effect of mannosylglycerate on unfolding thermodynamics and activity of ribonuclease A. *ChemBioChem* **4**, 734–741.

Faria, T. Q., Lima, J. C., Bastos, M., Maçanita, A. L. & Santos, H. (2004). Protein stabilization by osmolytes from hyperthermophiles: effect of mannosylglycerate on the thermal unfolding of recombinant nuclease a from *Staphylococcus aureus* studied by picosecond time-resolved fluorescence and calorimetry. *J Biol Chem* **279**, 48680–91.

Faria, T. Q., Mingote, A., Siopa, F., Ventura, R., Maycock, C. & Santos, H. (2008). Design of new enzyme stabilizers inspired by glycosides of hyperthermophilic microorganisms. *Carbohydr Res* **343**, 3025–3033.

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L. & Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7**, 129–43.

Feofilova, E. P. (1992). Trehalose, stress, and anabiosis (review). *Mikrobiologiya* **61**, 741–755.

Fernandes, C., Mendes, V., Costa, J., Empadinhas, N., Jorge, C., Lamosa, P., Santos, H. & da Costa, M. S. (2010). Two alternative pathways for the synthesis of the rare compatible solute mannosylglucosylglycerate in *Petrotoga mobilis*. *J Bacteriol* **192**, 1624–33.

Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* **2008**, 102–114.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A. & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.

Flower, T. R., Chesnokova, L. S., Froelich, C. A., Dixon, C. & Witt, S. N. (2005). Heat Shock Prevents Alpha-synuclein-induced Apoptosis in a Yeast Model of Parkinson's Disease. *J Mol Biol* **351**, 1081–1100.

Gietz, R. D. & Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, 87–96.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C. & other authors. **(1996).** Life with 6000 genes. *Science* **274**, 546, 563–7.

Goldemberg, J. (2007). Ethanol for a sustainable energy future. *Science* **315**, 808–10.

Gonçalves, E., Pereira, R., Rocha, I. & Rocha, M. (2012). Discovery of Optimal Targets for Gene Over/Underexpression. *J Comput Biol* **19**, 102–114.

Gonçalves, L. G., Huber, R., da Costa, M. S. & Santos, H. (2003). A variant of the hyperthermophile *Archaeoglobus fulgidus* adapted to grow at high salinity. *FEMS Microbiol Lett* **218**, 239–44.

Gonçalves, L. G., Lamosa, P., Huber, R. & Santos, H. (2008). Di-myo-inositol phosphate and novel UDP-sugars accumulate in the extreme hyperthermophile *Pyrolobus fumarii*. *Extremophiles* **12**, 383–9.

Graf, R., Anzali, S., Buenger, J., Pfluecker, F. & Driller, H. (2008). The multifunctional role of ectoine as a natural cell protectant. *Clin Dermatol* **26**, 326–33.

Gu, Y. & Jérôme, F. (2013). Bio-based solvents: an emerging generation of fluids for the design of eco-efficient processes in catalysis and organic chemistry. *Chem Soc Rev* **42**, 9550–70.

Günther, G. & Gottschalk, G. (1991). Limitation of growth and lactic acid production in batch and continuous cultures of *Lactobacillus helveticus*. *Appl Microbiol Biotechnol* **34**, 446–449.

Guo, A. C., Jewison, T., Wilson, M., Liu, Y., Knox, C., Djoumbou, Y., Lo, P., Mandal, R., Krishnamurthy, R. & Wishart, D. S. (2013). ECMDB: the *E. coli* Metabolome Database. *Nucleic Acids Res* **41**, D625-30.

Guthrie, C. & Fink, G. R. (1991). Guide to yeast genetics and molecular biology. In *Methods Enzymol , Vol 169*, p. 933pp. Academic Press, San Diego.

Han, M. J. & Lee, S. Y. (2003). Proteome profiling and its use in metabolic and cellular engineering. *Proteomics* **3**, 2317–2324.

Hanai, T., Atsumi, S. & Liao, J. C. (2007). Engineered synthetic pathway for isopropanol production in *Escherichia coli*. *Appl Environ Microbiol* **73**, 7814–8.

Hansen, E. H., Møller, B. L., Kock, G. R., Bünner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S. & Hansen, J. (2009). De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol* **75**, 2765–74.

Hashimoto, H., Sakakibara, A., Yamasaki, M. & Yoda, K. (1997). *Saccharomyces cerevisiae* VIG9 encodes GDP-mannose pyrophosphorylase, which is essential for protein glycosylation. *J Biol Chem* **272**, 16308–14.

Herrgård, M. J. (2006). Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* **16**, 627–635.

Hill, J. E., Myers, A. M., Koerner, T. J. & Tzagoloff, A. (1986). Yeast/*E. coli* shuttle vectors with multiple unique restriction sites. *Yeast* **2**, 163–7.

Van Hoek, P., Van Dijken, J. P. & Pronk, J. T. (1998). Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl Environ Microbiol* **64**, 4226–33.

Holmström, K. O., Somersalo, S., Mandal, A., Palva, T. E. & Welin, B. (2000). Improved tolerance to salinity and low temperature in transgenic tobacco producing glycine betaine. *J Exp Bot* **51**, 177–85.

Huh, W.-K., Falvo, J. V, Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S. & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**, 686–91.

Hyduke, D. R., Lewis, N. E. & Palsson, B. Ø. (2013). Analysis of omics data with genome-scale models of metabolism. *Mol BioSyst* **9**, 167–174.

Ignatova, Z. & Gierasch, L. M. (2006). Inhibition of protein aggregation in vitro and in vivo by a natural osmoprotectant. *Proc Natl Acad Sci U S A* **103**, 13357–61.

Imam, S., SchÃ¤uble, S., Brooks, A. N., Baliga, N. S. & Price, N. D. (2015). Data-driven integration of genome-scale regulatory and metabolic network models. *Front Microbiol* **6**, 409.

Janik, A., Sosnowska, M., Kruszewska, J., Krotkiewski, H., Lehle, L. & Palamarczyk, G. (2003). Overexpression of GDP-mannose pyrophosphorylase in *Saccharomyces cerevisiae* corrects defects in dolichol-linked saccharide formation and protein glycosylation. *Biochim Biophys Acta - Gen Subj* **1621**, 22–30.

Jenkins, G. M., Richards, A., Wahl, T., Mao, C., Obeid, L. & Hannun, Y. (1997). Involvement of yeast sphingolipids in the heat stress response of *Saccharomyces cerevisiae*. *J Biol Chem* **272**, 32566–32572.

Jensen, P. A. & Papin, J. A. (2011). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27**, 541–7.

Jensen, P. A., Lutz, K. A. & Papin, J. A. (2011). TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Syst Biol* **5**, 147.

Jewison, T., Knox, C., Neveu, V., Djoumbou, Y., Guo, A. C., Lee, J., Liu, P., Mandal, R., Krishnamurthy, R. & other authors. **(2012).** YMDB: the Yeast Metabolome Database. *Nucleic Acids Res* **40**, D815-20.

Johansson, N., Quehl, P., Norbeck, J. & Larsson, C. (2013). Identification of factors for improved ethylene production via the ethylene forming enzyme in chemostat cultures of *Saccharomyces cerevisiae*. *Microb Cell Fact* **12**, 89.

Jones, D. T. & Woods, D. R. (1986). Acetone-butanol fermentation revisited. *Microbiol Rev* **50**, 484–524.

Jones, G. M., Stalker, J., Humphray, S., West, A., Cox, T., Rogers, J., Dunham, I. & Prelich, G. (2008). A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. *Nat Methods* **5**, 239–41.

Jorge, C. D., Lamosa, P. & Santos, H. (2007). α-d-Mannopyranosyl-(1→2)-α-d-glucopyranosyl-(1→2)-glycerate in the thermophilic bacterium *Petrotoga miotherma* − structure, cellular content and function. *FEBS J* **274**, 3120–3127.

Jorge, C. D., Borges, N., Bagyan, I., Bilstein, A. & Santos, H. (2016). Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. *Extremophiles* **20**, 251–9.

Kanapathipillai, M., Ku, S. H., Girigoswami, K. & Park, C. B. (2008). Small stress molecules inhibit aggregation and neurotoxicity of prion peptide 106–126. *Biochem Biophys Res Commun* **365**, 808–813.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353–D361.

Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., Kaipa, P., Gilham, F., Spaulding, A. & other authors. **(2010).** Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11**, 40–79.

Karsten, U. & West, J. (1993). Ecophysiological Studies on Six Species of the Mangrove Red Algal Genus Caloglossa. *Aust J Plant Physiol* **20**, 729.

Karsten, U., Görs, S., Eggert, A. & West, J. A. (2007). Trehalose, digeneaside, and floridoside in the Florideophyceae (Rhodophyta) – a reevaluation of its chemotaxonomic value. *Phycologia* **46**, 143–150.

Katajamaa, M., Miettinen, J. & Oresic, M. (2006). MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics* **22**, 634–6.

Kavšček, M., Stražar, M., Curk, T., Natter, K. & Petrovič, U. (2015). Yeast as a cell factory: current state and perspectives. *Microb Cell Fact* **14**, 94.

Khan, S. H., Ahmad, N., Ahmad, F. & Kumar, R. (2010). Naturally occurring organic osmolytes: From cell physiology to disease prevention. *IUBMB Life* **62**, 891–895.

Kim, H. U., Kim, W. J. & Lee, S. Y. (2013). Flux-coupled genes and their use in metabolic flux analysis. *Biotechnol J* **8**, 1035–42.

Kim, J. & Reed, J. L. (2010). OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol* **4**, 53.

Kocharin, K. & Nielsen, J. (2013). Specific growth rate and substrate dependent polyhydroxybutyrate production in *Saccharomyces cerevisiae*. *AMB Express* **3**, 18.

Kremer, B. P. (1980). Taxonomic implications of algal photoassimilate patterns. *Br Phycol J* **15**, 399–409.

Krivoruchko, A. & Nielsen, J. (2014). Production of natural products through metabolic engineering of *Saccharomyces cerevisiae*. *Curr Opin Biotechnol* **35C**, 7–15.

Kuehner, J. N. & Brow, D. A. (2008). Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. *Mol Cell* **31**, 201–211.

Kumar, R. R. & Prasad, S. (2011). Metabolic Engineering of Bacteria. *Indian J Microbiol* **51**, 403–409.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–5. Nature Publishing Group.

Lamosa, Martins, Da Costa MS & Santos. **(1998).** Effects of temperature, salinity, and medium composition on compatible solute accumulation by *Thermococcus* spp. *Appl Environ Microbiol* **64**, 3591–8.

Lamosa, P., Faria, T. Q., Ventura, M. R., Maycock, C. D. & Santos, H. (2009). Utilization of compatible solutes to improve the performance of the techniques using immobilized biologic materials. Applicants: Stab Vida, Bitop.

Lamosa, P., Faria, T. Q., Ventura, M. R., Maycock, C. D. & Santos, H. (2006a). Derivados sintéticos de manosilglicerato para a estabilização e/ou preservação de biomateriais. Patente PAT 103442 P. Requerente: STAB VIDA, Lda. Abandonada.

Lamosa, P., Burke, A., Peist, R., Huber, R., Liu, M. Y., Silva, G., Rodrigues-Pousada, C., LeGall, J., Maycock, C. & Santos, H. (2000). Thermostabilization of proteins by diglycerol phosphate, a new compatible solute from the hyperthermophile *Archaeoglobus fulgidus*. *Appl Environ Microbiol* **66**, 1974–1979.

Lamosa, P., Gonçalves, L. G., Rodrigues, M. V, Martins, L. O., Raven, N. D. H. & Santos, H. (2006b). Occurrence of 1-glyceryl-1-myo-inosityl phosphate in hyperthermophiles. *Appl Environ Microbiol* **72**, 6169–73.

Lamosa, P., Faria, T. Q., Borges, N., Neves, C. & Santos, H. (2007). The Physiological Role, Biosynthesis, and Mode of Action of Compatible Solutes from (Hyper)Thermophiles. In *Physiol Biochem Extrem*, pp. 86–103. American Society of Microbiology.

Lander, E. S. (1999). Array of hope. *Nat Genet* **21**, 3–4.

Lange, H. C. & Heijnen, J. J. (2001). Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **75**, 334–44.

Larsson, C., von Stockar, U., Marison, I. & Gustafsson, L. (1993). Growth and metabolism of *Saccharomyces cerevisiae* in chemostat cultures under carbon-, nitrogen-, or carbon- and nitrogen-limiting conditions. *J Bacteriol* **175**, 4809–16.

Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S. & Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol* **8**, 536–46.

Lee, S. Y., Kim, H. U., Park, J. H., Park, J. M. & Kim, T. Y. (2009). Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov Today* **14**, 78–88.

Lee, S. Y., Park, J. M. & Kim, T. Y. (2011). Chapter four – Application of Metabolic Flux Analysis in Metabolic Engineering. In *Methods Enzymol*, pp. 67–93.

Leuenberger, H. G. W. (1972). Cultivation of *Saccharomyces cerevisiae* in continuous culture. *Arch Mikrobiol* **83**, 347–358.

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O. & other authors. **(2010).** Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* **6**, 390.

Li, X., Liu, Y., Yang, Y., Zhang, H., Wang, H., Wu, Y., Zhang, M., Sun, T., Cheng, J. & other authors. **(2014).** High levels of malic acid production by the bioconversion of corn straw hydrolyte using an isolated *Rhizopus delemar* strain. *Biotechnol Bioprocess Eng* **19**, 478–492.

Liu, C., Wang, H., Karim, A. M., Sun, J. & Wang, Y. (2014). Catalytic fast pyrolysis of lignocellulosic biomass. *Chem Soc Rev* **43**, 7594–7623.

Liu, R., Bassalo, M. C., Zeitoun, R. I. & Gill, R. T. (2015a). Genome scale engineering techniques for metabolic engineering. *Metab Eng* **32**, 143–54.

Liu, W., Wu, A., Pellegrini, M. & Wang, X. (2015b). Integrative analysis of human protein, function and disease networks. *Sci Rep* **5**, 14344.

Liu, Y. & Bolen, D. W. (1995). The Peptide Backbone Plays a Dominant Role in Protein Stabilization by Naturally Occurring Osmolytes. *Biochemistry* **34**, 12884–12891.

Loffler, F. E., Yan, J., Ritalahti, K. M., Adrian, L., Edwards, E. A., Konstantinidis, K. T., Muller, J. A., Fullerton, H., Zinder, S. H. & Spormann, A. M. (2013). *Dehalococcoides mccartyi* gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidia* classis nov., order *Dehalococcoidales* ord. nov. and famil. *Int J Syst Evol Microbiol* **63**, 625–635.

Longo, C. M., Wei, Y., Roberts, M. F. & Miller, S. J. (2009). Asymmetric Syntheses of l,l - and l,d - Di- myo -inositol-1,1′-phosphate and their Behavior as Stabilizers of Enzyme Activity at Extreme Temperatures. *Angew Chemie Int Ed* **48**, 4158–4161.

Lõoke, M., Kristjuhan, K. & Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* **50**, 325–8.

Lun, D. S., Rockwell, G., Guido, N. J., Baym, M., Kelner, J. A., Berger, B., Galagan, J. E. & Church, G. M. (2009). Large-scale identification of genetic design strategies using local search. *Mol Syst Biol* **5**, 296.

Ma, F. & Milford, A. **H. (1999).** Biodiesel production: a review. *Bioresour Technol* **70**, 1–15.

Madeo, F., Fröhlich, E., Ligr, M., Grey, M., Sigrist, S. J., Wolf, D. H. & Fröhlich, K. U. (1999). Oxygen stress: a regulator of apoptosis in yeast. *J Cell Biol* **145**, 757–67. The Rockefeller University Press.

Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. (2004). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **33**, D54–D58.

Maia, P., Rocha, M. & Rocha, I. (2016). *In silico* Constraint-Based Strain Optimization Methods: the Quest for optimal cell factories **80**, 45–67.

Mann, M. & Jensen, O. N. (2003). Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255–261.

Marini, A., Reinelt, K., Krutmann, J. & Bilstein, A. (2014). Ectoine-containing cream in the treatment of mild to moderate atopic dermatitis: a randomised, comparator-controlled, intra-individual double-blind, multi-center trial. *Skin Pharmacol Physiol* **27**, 57–65.

Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* **21**, 796–802.

Martins, L. O. & Santos, H. (1995). Accumulation of Mannosylglycerate and Di-myo-Inositol-Phosphate by *Pyrococcus furiosus* in Response to Salinity and Temperature. *Appl Environ Microbiol* **61**, 3299–303.

Martins, L. O., Carreto, L. S., Da Costa, M. S. & Santos, H. (1996). New compatible solutes related to Di-myo-inositol-phosphate in members of the order *Thermotogales*. *J Bacteriol* **178**, 5644–51.

Martins, L. O., Empadinhas, N., Marugg, J. D., Miguel, C., Ferreira, C., da Costa, M. S. & Santos, H. (1999). Biosynthesis of mannosylglycerate in the thermophilic bacterium *Rhodothermus marinus*. Biochemical and genetic characterization of a mannosylglycerate synthase. *J Biol Chem* **274**, 35407–14.

Mascellani, N., Liu, X., Rossi, S., Marchesini, J., Valentini, D., Arcelli, D., Taccioli, C., Helmer Citterich, M., Liu, C.-G. & other authors. **(2007).** Compatible solutes from hyperthermophiles improve the quality of DNA microarrays. *BMC Biotechnol* **7**, 82.

Maschio, G., Lucchesi, A. & Stoppato, G. (1994). Production of syngas from biomass. *Bioresour Technol* **48**, 119–126.

McCloskey, D., Palsson, B. Ø. & Feist, A. M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol* **9**, 661. EMBO Press.

McLean, P. J., Kawamata, H., Shariff, S., Hewett, J., Sharma, N., Ueda, K., Breakefield, X. O. & Hyman, B. T. (2002). TorsinA and heat shock proteins act as molecular chaperones: suppression of alpha-synuclein aggregation. *J Neurochem* **83**, 846–54.

McLean, P. J., Klucken, J., Shin, Y. & Hyman, B. T. (2004). Geldanamycin induces Hsp70 and prevents α-synuclein aggregation and toxicity in vitro. *Biochem Biophys Res Commun* **321**, 665–669.

Melamed, D., Pnueli, L. & Arava, Y. (2008). Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *RNA* **14**, 1337–1351.

Milne, C. B., Kim, P.-J., Eddy, J. A. & Price, N. D. (2009). Accomplishments in genome-scale *in silico* modeling for industrial and medical biotechnology. *Biotechnol J* **4**, 1653–1670.

Mo, M. L., Palsson, B. Ø. & Herrgård, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* **3**, 37.

Monk, J., Nogales, J. & Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nat Biotechnol* **32**, 447–52.

Morello, J.-P., Salahpour, A., Laperrière, A., Bernier, V., Arthus, M.-F., Lonergan, M., Petäjä-Repo, U., Angers, S., Morin, D. & other authors. **(2000).** Pharmacological chaperones rescue cell-surface expression and function of misfolded V2 vasopressin receptor mutants. *J Clin Invest* **105**, 887–895.

Moriya, Y., Shigemizu, D., Hattori, M., Tokimatsu, T., Kotera, M., Goto, S. & Kanehisa, M. (2010). PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res* **38**, W138-43.

Mortimer, R. K. (2000). Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Res* **10**, 403–9.

Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renew Sustain Energy Rev* **14**, 578–597.

Nakamura, C. E. & Whited, G. M. (2003). Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* **14**, 454–9.

Nat Biotech, P. (2014). How big is the bioeconomy? *Nat Biotechnol* **32**, 598–598.

Navid, A. & Almaas, E. (2012). Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst Biol* **6**, 150.

Neckers, L. (2002). Hsp90 inhibitors as novel cancer chemotherapeutic agents. *Trends Mol Med* **8**, S55-61.

Neves, C., da Costa, M. S. & Santos, H. (2005). Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: osmoadaptation and thermoadaptation in the order thermococcales. *Appl Environ Microbiol* **71**, 8091–8.

Nevoigt, E. (2008). Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **72**, 379–412.

Nielsen, J., Villadsen, J. & Lidén, G. (2003). *Bioreaction Engineering Principles*. *First Ed Bioreaclion Eng Princ*. Boston, MA: Springer US.

Nishizawa, M., Araki, R. & Teranishi, Y. (1989). Identification of an upstream activating sequence and an upstream repressible sequence of the pyruvate kinase gene of the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **9**, 442–51.

Nobre, A., Empadinhas, N., Nobre, M. F., Lourenço, E. C., Maycock, C., Ventura, M. R., Mingote, A. & da Costa, M. S. (2013). The plant *Selaginella moellendorffii* possesses enzymes for synthesis and hydrolysis of the compatible solutes mannosylglycerate and glucosylglycerate. *Planta* **237**, 891–901.

Nunes, O. C., Manaia, C. M., Da Costa, M. S. & Santos, H. (1995). Compatible Solutes in the Thermophilic Bacteria *Rhodothermus marinus* and '*Thermus thermophilus*'. *Appl Environ Microbiol* **61**, 2351–7.

O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* **9**, 693.

O'Malley, B. W., Towle, H. C. & Schwartz, R. J. (1977). Regulation of Gene Expression in Eucaryotes. *Annu Rev Genet* **11**, 239–275.

OECD. **(2009).** *The Bioeconomy to 2030 Main Findings and Policy Conclusions*.

OECD. **(2011a).** Industrial Biotechnology and Climate Change.

OECD. **(2011b).** *Future Prospects for Industrial Biotechnology*.

Okoniewski, M. J. & Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics* **7**, 276.

Olsson, L. & Hahn-Hägerdal, B. (1993). Fermentative performance of bacteria and yeasts in lignocellulose hydrolysates. *Process Biochem* **28**, 249–257.

Orr-Weaver, T. L. & Szostak, J. W. (1983). Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc Natl Acad Sci* **80**, 4417–4421.

Orth, J. D., Thiele, I. & Palsson, B. Ø. (2010). What is flux balance analysis? *Nat Biotechnol* **28**, 245–8.

Osterlund, T., Nookaew, I. & Nielsen, J. (2012). Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Biotechnol Adv* **30**, 979–988. Elsevier Inc.

Otero, J. M. & Nielsen, J. (2010). Industrial systems biology. *Biotechnol Bioeng* **105**, 439–60.

Otero, J. M., Cimini, D., Patil, K. R., Poulsen, S. G., Olsson, L. & Nielsen, J. (2013). Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS One* **8**, e54144.

Outeiro, T. F., Klucken, J., Bercury, K., Tetzlaff, J., Putcha, P., Oliveira, L. M. A., Quintas, A., McLean, P. J. & Hyman, B. T. (2009). Dopamine-Induced Conformational Changes in Alpha-Synuclein. *PLoS One* **4**, e6906.

Outeiro, T. F. & Lindquist, S. (2003). Yeast cells provide insight into alpha-synuclein biology and pathobiology. *Science* **302**, 1772–5.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V, Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T. & other authors. **(2005).** The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**, 5691–702.

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A. & other authors. **(2013).** High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–32.

Paddon, C. J. & Keasling, J. D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Microbiol* **12**, 355–67.

Pais, T. M., Lamosa, P., Matzapetakis, M., Turner, D. L. & Santos, H. (2012). Mannosylglycerate stabilizes staphylococcal nuclease with restriction of slow β-sheet motions. *Protein Sci* **21**, 1126–1137.

Pais, T. M., Lamosa, P., Garcia-Moreno, B., Turner, D. L. & Santos, H. (2009). Relationship between Protein Stabilization and Protein Rigidification Induced by Mannosylglycerate. *J Mol Biol* **394**, 237–50.

Paley, S. M. & Karp, P. D. (2006). The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res* **34**, 3771–8.

Papini, M., Nookaew, I., Scalcinati, G., Siewers, V. & Nielsen, J. (2010). Phosphoglycerate mutase knock-out mutant *Saccharomyces cerevisiae*: Physiological investigation and transcriptome analysis. *Biotechnol J* **5**, 1016–1027.

Papp, E. & Csermely, P. (2006). Chemical Chaperones: Mechanisms of Action and Potential Use. In *Mol Chaperones Heal Dis*, pp. 405–416. Berlin/Heidelberg: Springer-Verlag.

Park, J. H., Lee, K. H., Kim, T. Y. & Lee, S. Y. (2007). Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci U S A* **104**, 7797–802.

Partow, S., Siewers, V., Bjørn, S., Nielsen, J. & Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **27**, 955–964.

Patil, K. R., Åkesson, M. & Nielsen, J. (2004). Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* **15**, 64–69.

Patil, K. R., Rocha, I., Förster, J. & Nielsen, J. (2005). Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* **6**, 308.

Patnaik, R. (2008). Engineering complex phenotypes in industrial strains. *Biotechnol Prog* **24**, 38–47.

Pearce, A. K., Crimmins, K., Groussac, E., Hewlins, M. J. E., Dickinson, J. R., Francois, J., Booth, I. R. & Brown, a. J. P. (2001). Pyruvate kinase (Pyk1) levels influence both the rate and direction of carbon flux in yeast under fermentative conditions. *Microbiology* **147**, 391–401.

Pellarin, R. & Caflisch, A. (2006). Interpreting the Aggregation Kinetics of Amyloid Peptides. *J Mol Biol* **360**, 882–892.

Pereira, R., Nielsen, J. & Rocha, I. (2016). Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*. *Metab Eng Commun* **3**, 153–163.

Pham, H. T. B., Larsson, G. & Enfors, S. O. (1998). Growth and energy metabolism in aerobic fed-batch cultures of *Saccharomyces cerevisiae*: Simulation and model verification. *Biotechnol Bioeng* **60**, 474–482.

Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Sampson, R., Houlgatte, R., Soularue, P. & Auffray, C. (1996). Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res* **6**, 492–503.

Ponce de León, M., Cancela, H. & Acerenza, L. (2008). A strategy to calculate the patterns of nutrient consumption by microorganisms applying a two-level optimisation principle to reconstructed metabolic networks. *J Biol Phys* **34**, 73–90.

Postma, E., Verduyn, C., Scheffers, W. A. & Van Dijken, J. P. (1989). Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **55**, 468–77.

Putcha, P., Danzer, K. M., Kranich, L. R., Scott, A., Silinski, M., Mabbett, S., Hicks, C. D., Veal, J. M., Steed, P. M. & other authors. **(2010).** Brain-permeable small-molecule inhibitors of Hsp90 prevent alpha-synuclein oligomer formation and rescue alpha-synuclein-induced toxicity. *J Pharmacol Exp Ther* **332**, 849–57.

Ramos, A., Raven, N. D. H., Sharp, R. J., Bartolucci, S., Rossi, M., Cannio, R., Lebbink, J., Vanderoost, J., Devos, W. M. & Santos, H. (1997). Stabilization of enzymes against thermal stress and freeze- drying by mannosylglycerate. *Appl Env Microbiol* **63**, 4020–4025.

Ranganathan, S., Suthers, P. F. & Maranas, C. D. (2010). OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* **6**, e1000744.

Reeve, A., Simcox, E. & Turnbull, D. (2014). Ageing and Parkinson's disease: Why is advancing age the biggest risk factor? *Ageing Res Rev* **14**, 19–30.

Riedel, M., Goldbaum, O., Schwarz, L., Schmitt, S. & Richter-Landsberg, C. (2010). 17-AAG Induces Cytoplasmic α-Synuclein Aggregate Clearance by Induction of Autophagy. *PLoS One* **5**, e8753.

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S. & other authors. **(2006).** Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943.

Roberts, M. F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems* **1**, 5.

Robertson, D. E., Roberts, M. F., Belay, N., Stetter, K. O. & Boone, D. R. (1990). Occurrence of beta-glutamate, a novel osmolyte, in marine methanogenic bacteria. *Appl Environ Microbiol* **56**, 1504–8.

Rocha, I., Förster, J. & Nielsen, J. (2008a). Design and application of genome-scale reconstructed metabolic models. *Methods Mol Biol* **416**, 409–31.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C. & Rocha, M. (2010). OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst Biol* **4**, 45.

Rocha, M., Maia, P., Mendes, R., Pinto, J. P., Ferreira, E. C., Nielsen, J., Patil, K. R. & Rocha, I. (2008b). Natural computation meta-heuristics for the *in silico* optimization of microbial strains. *BMC Bioinformatics* **9**, 499.

Rodrigues, M. V, Borges, N., Almeida, C. P., Lamosa, P. & Santos, H. (2009). A unique beta-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-myo-inositol phosphate as the mannosyl acceptor. *J Bacteriol* **191**, 6105–15.

Ronzon, T., Santini, F. & M'Barek, R. (2015). The Bioeconomy in the European Union in numbers. Facts and figures on biomass, turnover and employment. *Eur Comm Jt Res Centre, Inst Prospect Technol Stud Spain, 4p*.

Rossell, S., Huynen, M. A. & Notebaart, R. A. (2013). Inferring metabolic states in uncharacterized environments using gene-expression measurements. *PLoS Comput Biol* **9**, e1002988.

Rothschild, L. J. & Mancinelli, R. L. (2001). Life in extreme environments. *Nature* **409**, 1092–101.

Ryu, J., Kanapathipillai, M., Lentzen, G. & Park, C. B. (2008). Inhibition of β-amyloid peptide aggregation and neurotoxicity by α-d-mannosylglycerate, a natural extremolyte. *Peptides* **29**, 578–584.

Saha, R., Chowdhury, A. & Maranas, C. D. (2014). Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr Opin Biotechnol* **29**, 39–45.

Sampaio, M. M., Santos, H. & Boos, W. (2003). Synthesis of GDP-mannose and mannosylglycerate from labeled mannose by genetically engineered *Escherichia coli* without loss of specific isotopic enrichment. *Appl Environ Microbiol* **69**, 233–240.

Sancenon, V., Lee, S.-A., Patrick, C., Griffith, J., Paulino, A., Outeiro, T. F., Reggiori, F., Masliah, E. & Muchowski, P. J. (2012). Suppression of α-synuclein toxicity and vesicle trafficking defects by phosphorylation at S129 in yeast depends on genetic context. *Hum Mol Genet* **21**, 2432–49.

Santos, H., Lamosa, P., Faria, T. Q., Pais, T. M., de la Paz, M. L. & Serrano, L. (2007). Compatible solutes of (hyper)thermophiles and their role in protein stabilization. In *Thermophiles*, pp. 9–24. Edited by R. F. Antranikian G, Driesen A. USA: CRC Taylor and Francis.

Santos, H., Ramos, A. & da Costa, M. S. (1996). Utilização de manosilglicerato na termoestabilização, osmoprotecção e protecção contra a desidratação de componentes celulares e células. Instituto Nacional da Propriedade Industrial. Patente de Invenção. Requerente: IBET.

Santos, H., Ramos, A. & da Costa, M. S. (1998). Thermostabilization, osmoprotection, and protection against desiccation of enzymes, cell components, and cells by mannosylglycerate.

Patente Europeia submetida em 14/03/97 com o nº 97670002.1 e publicada em 07/01/98 Bulletin 1998/02 (EP 816509 A2, A3). Requerente: IBET. Licenciada à bitop AG em 2002.

Santos, H. & da Costa, M. S. (2002). Compatible solutes of organisms that live in hot saline environments. *Environ Microbiol* **4**, 501–509.

Santos, H., Lamosa, P. & Borges, N. (2006). Characterization and Quantification of Compatible Solutes in (Hyper)thermophilic Microorganisms. In *Methods Microbiol*, pp. 173–199.

Santos, H., Lamosa, P., Borges, N., Gonçalves, L. G., Pais, T. & Rodrigues, M. V. (2011). Organic Compatible Solutes of Prokaryotes that Thrive in Hot Environments: The Importance of Ionic Compounds for Thermostabilization. In *Extrem Handb*, pp. 497–520. Tokyo: Springer Japan.

Sarkar, S., Davies, J. E., Huang, Z., Tunnacliffe, A. & Rubinsztein, D. C. (2007). Trehalose, a novel mTOR-independent autophagy enhancer, accelerates the clearance of mutant huntingtin and alpha-synuclein. *J Biol Chem* **282**, 5641–52.

Schein, C. H. (1990). Solubility as a function of protein structure and solvent components. *Biotechnology* **8**, 308–17.

Scheller, H., Harholt, J. & Ulvskov, P. (2010). Stress-Tolerant Plants Expressing Mannosylglycerate-Producing Enzymes. European Patent Office.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–70.

Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson, B. O. & Hyduke, D. R. (2013). GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics* **29**, 2900–2908.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. & Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* **32**, 431D–433.

Schwarz, T. (2003). Use of beta-mannosylglycerate and derivatives in cosmetic and dermatological formulations. Patent no: WO/2002/015867. Applicants: BITOP GMBH.

Segrè, D., Vitkup, D. & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* **99**, 15112–7.

Seki, T., Abe-Seki, N., Kikawada, T., Takahashi, H., Yamamoto, K., Adachi, N., Tanaka, S., Hide, I., Saito, N. & Sakai, N. (2010). Effect of trehalose on the properties of mutant {gamma}PKC, which causes spinocerebellar ataxia type 14, in neuronal cell lines and cultured Purkinje cells. *J Biol Chem* **285**, 33252–64.

Sheldon, R., Arends, I. & Hanefeld, U. (2007). Introduction: Green Chemistry and Catalysis. In *Green Chem Catal*, pp. 1–47. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Sheldon, R. a. (2000). Atom utilisation, E factors and the catalytic solution. *Comptes Rendus l'Académie des Sci - Ser IIC - Chem* **3**, 541–551.

Sheldon, R. A. (2014). Green and sustainable manufacture of chemicals from biomass: state of the art. *Green Chem* **16**, 950–963.

Shima, S., Hérault, D. A., Berkessel, A. & Thauer, R. K. (1998). Activation and thermostabilization effects of cyclic 2, 3-diphosphoglycerate on enzymes from the hyperthermophilic *Methanopyrus kandleri*. *Arch Microbiol* **170**, 469–72.

Shimma, Y., Nishikawa, A., bin Kassim, B., Eto, A. & Jigami, Y. (1997). A defect in GTP synthesis affects mannose outer chain elongation in *Saccharomyces cerevisiae*. *Mol Gen Genet* **256**, 469–80.

Shlomi, T., Berkman, O. & Ruppin, E. (2005a). Regulatory on/off ¯minimization of metabolic flux. *Pnas* **102**, 7695–7700.

Shlomi, T., Berkman, O. & Ruppin, E. (2005b). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* **102**, 7695–700.

Silva, Z., Borges, N., Martins, L. O., Wait, R., da Costa, M. S. & Santos, H. (1999). Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by *Rhodothermus marinus* and *Rhodothermus obamensis*. *Extremophiles* **3**, 163–72.

Silva, Z., Alarico, S., Nobre, A., Horlacher, R., Marugg, J., Boos, W., Mingote, A. I. & da Costa, M. S. (2003). Osmotic adaptation of *Thermus thermophilus* RQ-1: lesson from a mutant deficient in synthesis of trehalose. *J Bacteriol* **185**, 5943–52.

Singer, M. A. & Lindquist, S. (1998). Multiple effects of trehalose on protein folding in vitro and in vivo. *Mol Cell* **1**, 639–48.

Singh Dhillon, G., Kaur Brar, S., Verma, M. & Tyagi, R. D. (2011). Recent Advances in Citric Acid Bio-production and Recovery. *Food Bioprocess Technol* **4**, 505–529.

Sivasamy, A., Cheah, K. Y., Fornasiero, P., Kemausuor, F., Zinoviev, S. & Miertus, S. (2009). Catalytic applications in the production of biodiesel from vegetable oils. *ChemSusChem* **2**, 278–300.

Sonderegger, M., Schümperli, M. & Sauer, U. (2004). Metabolic engineering of a phosphoketolase pathway for pentose catabolism in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **70**, 2892–7.

Song, C. W., Kim, D. I., Choi, S., Jang, J. W. & Lee, S. Y. (2013). Metabolic engineering of *Escherichia coli* for the production of fumaric acid. *Biotechnol Bioeng* **110**, 2025–2034.

Sonnleitner, B. & Käppeli, O. (1986). Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis. *Biotechnol Bioeng* **28**, 927–37.

Sprague, G. F. (1977). Isolation and characterization of a *Saccharomyces cerevisiae* mutant deficient in pyruvate kinase activity. *J Bacteriol* **130**, 232–41.

Stanley, D., Bandara, A., Fraser, S., Chambers, P. J. & Stanley, G. A. (2010). The ethanol stress response and ethanol tolerance of *Saccharomyces cerevisiae*. *J Appl Microbiol* **109**, 13–24.

Stark, M. J. & Stansfield, I. (2007). *Yeast Gene Analysis*. *2nd Ed San Diego*. Academic Press.

Stefanis, L. (2012). α-Synuclein in Parkinson's Disease. *Cold Spring Harb Perspect Med* **2**, a009399.

Stephanopoulos, G. N., Aristidou, A. A. & Nielsen, J. (1998). The Essence of Metabolic Engineering. In *Metab Eng*, pp. 1–20.

Swainston, N., Smallbone, K., Mendes, P., Kell, D. & Paton, N. (2011). The SuBliMinaL Toolbox: automating steps in the reconstruction of metabolic networks. *J Integr Bioinform* **8**, 186.

Tatzelt, J., Prusiner, S. B. & Welch, W. J. (1996). Chemical chaperones interfere with the formation of scrapie prion protein. *EMBO J* **15**, 6363–73.

Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Gonçalves, J. P., Mira, N. P., dos Santos, S. C., Cabrito, T. R., Palma, M., Costa, C. & other authors. **(2014).** The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **42**, D161-6.

Tepper, N. & Shlomi, T. (2010). Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* **26**, 536–43.

Theobald, U., Mailinger, W., Baltes, M., Rizzi, M. & Reuss, M. (1997). *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnol Bioeng* **55**, 305–16.

Thiele, I. & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* **5**, 93–121.

Timasheff, S. N. (1992). Water as ligand: Preferential binding and exclusion of denaturants in protein unfolding. *Biochemistry* **31**, 9857–9864.

Töpfer, N., Jozefczuk, S. & Nikoloski, Z. (2012). Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. *BMC Syst Biol* **6**, 148.

Trinh, J. & Farrer, M. (2013). Advances in the genetics of Parkinson disease. *Nat Rev Neurol* **9**, 445–454.

Trojanowski, J. Q. & Lee, V. M.-Y. (2003). Parkinson's disease and related alpha-synucleinopathies are brain amyloidoses. *Ann N Y Acad Sci* **991**, 107–10.

Uversky, V. N., Li, J. & Fink, A. L. (2001). Trimethylamine-N-oxide-induced folding of α-synuclein. *FEBS Lett* **509**, 31–35.

Varshney, R. K., Bansal, K. C., Aggarwal, P. K., Datta, S. K. & Craufurd, P. Q. (2011). Agricultural biotechnology for crop improvement in a variable climate: hope or hype? *Trends Plant Sci* **16**, 363–71.

Vemuri, G. N. & Aristidou, A. a. **(2005).** Metabolic Engineering in the -omics Era: Elucidating and Modulating Regulatory Networks Metabolic Engineering in the -omics Era: Elucidating and Modulating Regulatory Networks. *Society* **69**, 197–216.

Vlysidis, A., Binns, M., Webb, C. & Theodoropoulos, C. (2011). A techno-economic analysis of biodiesel biorefineries: Assessment of integrated designs for the co-production of fuels and chemicals. *Energy* **36**, 4671–4683.

Vos, T., de la Torre Cortés, P., van Gulik, W. M., Pronk, J. T. & Daran-Lapujade, P. (2015). Growth-rate dependency of de novo resveratrol production in chemostat cultures of an engineered *Saccharomyces cerevisiae* strain. *Microb Cell Fact* **14**, 133.

Wang, C., Yoon, S.-H., Jang, H.-J., Chung, Y.-R., Kim, J.-Y., Choi, E.-S. & Kim, S.-W. (2011). Metabolic engineering of *Escherichia coli* for α-farnesene production. *Metab Eng* **13**, 648–55.

Wang, Y., Eddy, J. A. & Price, N. D. (2012). Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol* **6**, 153.

Wang, Z., Gerstein, M. & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63.

Westerhoff, H. V & Palsson, B. O. (2004). The evolution of molecular biology into systems biology. *Nat Biotechnol* **22**, 1249–1252.

Wiechert, W. & de Graaf, A. A. (1996). *In vivo* stationary flux analysis by 13C labeling experiments. *Adv Biochem Eng Biotechnol* **54**, 109–54.

Wiemken, A. (1990). Trehalose in yeast, stress protectant rather than reserve carbohydrate. *Antonie Van Leeuwenhoek* **58**, 209–217.

Willingham, S., Outeiro, T. F., DeVit, M. J., Lindquist, S. L. & Muchowski, P. J. (2003). Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein. *Science* **302**, 1769–72.

Wintermute, E. H., Lieberman, T. D. & Silver, P. A. (2013). An objective function exploiting suboptimal solutions in metabolic networks. *BMC Syst Biol* **7**, 98.

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D. & other authors. **(1999).** Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–6.

Witt, S. N. & Flower, T. R. (2006). alpha-Synuclein, oxidative stress and apoptosis from the perspective of a yeast model of Parkinson's disease. *FEMS Yeast Res* **6**, 1107–16.

Wittmann, C. (2007). Fluxome analysis using GC-MS. *Microb Cell Fact* **6**, 6.

Woolcock, P. J. & Brown, R. C. (2013). A review of cleaning technologies for biomass-derived syngas. *Biomass and Bioenergy* **52**, 54–84.

Wyrick, J. J., Holstege, F. C., Jennings, E. G., Causton, H. C., Shore, D., Grunstein, M., Lander, E. S. & Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418–21.

Xu, Z., Zheng, P., Sun, J. & Ma, Y. (2013). ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *PLoS One* **8**, e72150.

Yamamoto, S., Gunji, W., Suzuki, H., Toda, H., Suda, M., Jojima, T., Inui, M. & Yukawa, H. (2012). Overexpression of genes encoding glycolytic enzymes in *Corynebacterium glutamicum* enhances glucose metabolism and alanine production under oxygen deprivation conditions. *Appl Environ Microbiol* **78**, 4447–57.

Yang, E.-C., Scot, J., West, J. A., Yoon, H.-S., Yokoyama, A., Karsten, U., De Goer, S. L. & Orlova, E. (2011a). *Erythrolobus australicus* sp. nov. (Porphyridiophyceae, Rhodophyta): a description based on several approaches. *ALGAE* **26**, 167–180.

Yang, L., Cluett, W. R. & Mahadevan, R. (2011b). EMILiO: a fast algorithm for genome-scale strain design. *Metab Eng* **13**, 272–81.

Yin, X., Li, J., Shin, H., Du, G., Liu, L. & Chen, J. (2015). Metabolic engineering in the biotechnological production of organic acids in the tricarboxylic acid cycle of microorganisms: Advances and prospects. *Biotechnol Adv* **33**, 830–41.

Yoda, K., Kawada, T., Kaibara, C., Fujie, A., Abe, M., Hitoshi, Hashimoto, Shimizu, J., Tomishige, N. & other authors. **(2000).** Defect in cell wall integrity of the yeast *Saccharomyces cerevisiae* caused by a mutation of the GDP-mannose pyrophosphorylase gene VIG9. *Biosci Biotechnol Biochem* **64**, 1937–41.

Yu, Z., Du, G., Zhou, J. & Chen, J. (2012). Enhanced α-ketoglutaric acid production in *Yarrowia lipolytica* WSH-Z06 by an improved integrated fed-batch strategy. *Bioresour Technol* **114**, 597–602.

Zabrocki, P., Pellens, K., Vanhelmont, T., Vandebroek, T., Griffioen, G., Wera, S., Van Leuven, F. & Winderickx, J. (2005). Characterization of alpha-synuclein aggregation and synergistic toxicity with protein tau in yeast. *FEBS J* **272**, 1386–400.

Zamboni, N., Fendt, S.-M., Rühl, M. & Sauer, U. (2009). 13C-based metabolic flux analysis. *Nat Protoc* **4**, 878–892.

Zhang, N., Gardner, D. C., Oliver, S. G. & Stateva, L. I. (1999). Down-regulation of the expression of PKC1 and SRB1/PSA1/VIG9, two genes involved in cell wall integrity in *Saccharomyces cerevisiae*, causes flocculation. *Microbiology* **145**, 309–16.

Zhang, X., Shanmugam, K. T. & Ingram, L. O. (2010). Fermentation of glycerol to succinate by metabolically engineered strains of *Escherichia coli*. *Appl Environ Microbiol* **76**, 2397–401.

van Zyl, W. H., Lynd, L. R., den Haan, R. & McBride, J. E. (2007). Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Adv Biochem Eng Biotechnol* **108**, 205–35.