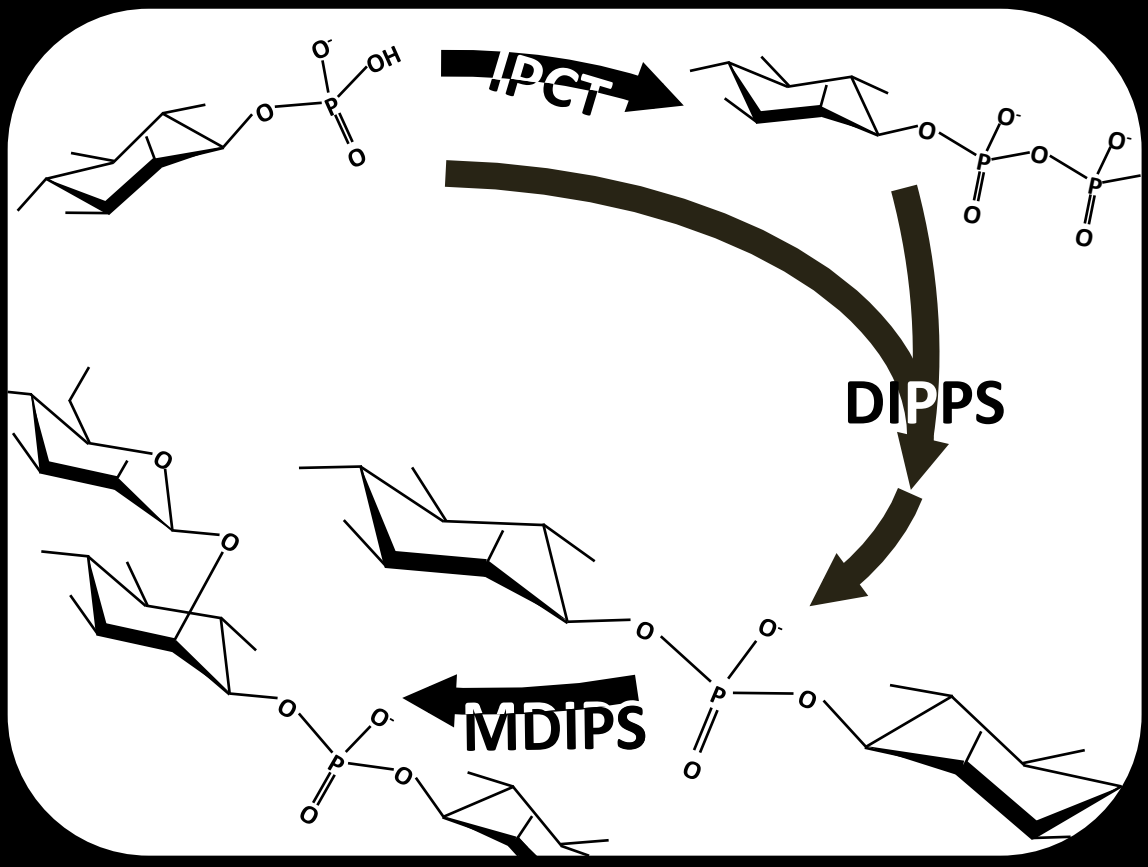
Heat stress adaptation in hyperthermophiles

Biosynthesis of inositol-containing compatible solutes

Marta Viseu Rodrigues



Dissertation presented to obtain the Ph.D. degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
July, 2011

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Heat stress adaptation in hyperthermophiles

Biosynthesis of inositol-containing compatible solutes

Marta Viseu Rodrigues

Dissertation presented to obtain the Ph.D. degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, July, 2011

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

From left to right: Prof. Carlos Romão, Prof. Isabel Sá-Correia, Prof. Helena Santos (Supervisor), Marta V. Rodrigues, Dr. Nuno Borges (Co-Supervisor), Prof. Beate Averhoff, Prof. Volker Müller. Oeiras, 15th of July 2011.

Apoio financeiro da Fundação para a Ciência e Tecnologia e do FSE no âmbito do Quadro Comunitário de apoio, Bolsa de Doutoramento com a referência SFRH / BD / 25539 / 2005.

Cover page by Dr. Nuno Borges: scheme of the pathway for the synthesis of di-*myo*-inositol phosphate and mannosyl-di-*myo*-inositol phosphate.

To my mother, father, and Nhanhã

Acknowledgments

First of all, I thank Prof. Helena Santos, my supervisor, for accepting me in her laboratory, for her guidance, support and motivation, and for all the effort invested in providing all the conditions necessary to perform this work. I thank for her patience during those long nights spent at the computer, but I must say that for me, those were learning periods by excellence. I thank for her persistence and for the confidence deposited in me. Prof. Helena Santos extreme dedication to science, scientific knowledge and rigor, were a major inspiration and driving force that greatly contributed to my education during these years.

To Dr. Nuno Borges, for being the best colleague ever, and for introducing me to the "bench" work by teaching me numerous experimental techniques. I thank him for the effort invested in the "work marathons" that we accomplished together, that made it much more easier to reach the end. His extreme attention to details and perfectionism made me a more meticulous scientist. I thank for the endless and unconditional support, for his drawing skills and for the critical and careful corrections that were essential to put together this dissertation.

To Dr. Pedro Lamosa, for his teaching and helpful discussions in many scientific matters. For his important contribution to this work especially with NMR related experiments.

To Dr. Luis (Gafeira) Gonçalves, for teaching me how to cultivate hyperthermophiles and for stimulating my interest for these difficult and

extraordinary organisms. I thank him for his great ideas and interest in this work that were essential for its development. I thank him for the help with NMR spectra, for teaching me many scientific matters and for the critical reading of this thesis. And for listening.

To Dr. David Turner, for the important scientific discussions and for correcting the abstract in this dissertation.

To Dr. Tiago Faria, for the experimental guidance and helpful discussions. For his unconditional availability and for being a good colleague. To Ana Mingote, the person with whom I shared the first days in this laboratory. For being a good friend and for contributing to a good working atmosphere.

To Dr. Rute Castro, Ana Lúcia, Patricia Almeida, Sónia Neto, Laura Paixão, and Teresa Maio, for their friendship, help, advices and just for being present. To Tiago Pais, Carla Jorge, Cristiana Faria, Ana Esteves, Dr. Luis Fonseca, Pedro Quintas, Dusica Rados, my other colleagues from the Cell Physiology & NMR group, for their help and for creating a good and friendly working environment. To Dr. Teresa Catarino, my girl guides "sister", for interesting discussions; to Dr. Paula Fareleira, for the support and scientific advices. To former members of the Cell physiology & NMR group Dr. Clélia Neves, Dr. Maria Manuel Sampaio, Dr. Margarida Santos, Dr. Melinda Noronha, Mafalda Henriques, Filipa Cardoso, Dr. Cláudia Sanchez, João Cavalheiro, and specially to Dr. Rasmus Larsen who taught me to work with lactic acid bacteria, and to Dr. Tony Collins and Dr. César Fonseca for their help and advices, and for contributing for a great working atmosphere on the 3rd floor. To Dr. Ana Rute Neves, Dr. Paula Gaspar, and Sandra Carvalho from the Lactic Acid Bacteria & In Vivo NMR group. To Filipe Almeida, a good next door neighbor.

I thank Luis Gonçalves, Miguel Loureiro, João Pires, Helena Matias, Fátima Madeira, Clara Fonseca, Isabel Baía, Anabela Bernardo, and Cristina Amaral for the help with practical issues.

To Fundação para a Ciência e Tecnologia for the financial support provided by the Ph.D. grant, and to Instituto de Tecnologia Química e Biólogica, for providing conditions to pursue scientific excellence.

Agradeço às minhas amigas de sempre Rita Carreira, Filipa Leão, Catarina Rosa, Margarida Sancho, Aurora Costa, Ana Nunes, Marta Abrantes, Margarida Matias e Lara Raquel e aos novos amigos para sempre Andreia Paiva, Rosário Abrantes e Daniel Carmo, pela amizade e por todo o apoio ao longo dos anos. Às minhas velhas amigas Guias Maria Rita, Ana Rita, Vera e Carminho, pela amizade incondicional, pela constante disponibilidade e por todo o percurso que fizémos juntas; à Ana Rute pela genuina amizade; às novas amigas Guias Ana Catarina, Beatriz, Maria e Saskia, pela sua amizade, energia, compreensão e apoio. Ao "Comi", Carolina Abrantes, Mariana Fernandes, Inês Morujo, Mariana Castro, Nina e Bárbara agradeço por terem apostado em mim mesmo nesta altura.

Agradeço à minha família, sobretudo aos meus Pais, à Nhanhã, e ao Rodrigo pelo verdadeiro amor, por acreditarem em mim e me apoiarem constante e incondicionalmente.

Abstract

The accumulation of low-molecular mass organic compounds, named compatible solutes, is an efficient, widespread strategy to counterbalance increases in the external osmolarity, thereby preserving cell viability. The intracellular accumulation of compatible solutes also occurs in response to supra-optimal temperatures, and this observation led to the assumption that they play a role in the thermoadaptation process. Hyperthermophiles, organisms with optimal growth temperatures above 80ºC, have been isolated from a variety of hot habitats. Many hyperthermophiles thrive in marine geothermal areas and are slightly halophilic. As a result, they have to cope with fluctuations in the salinity of the external medium and generally accumulate compatible solutes as a defense strategy. Interestingly, these hyperthermophilic organisms show a clear preference for negatively charged solutes, such as diglycerol phosphate, di-*myo*-inositol 1,3'-phosphate and mannosylglycerate, over neutral or zwitterionic solutes typically found in mesophiles (glycerol, trehalose, *myo*-inositol, and ectoines). The question then arises whether those charged solutes were selected by organisms adapted to grow at high temperatures because they are more suitable to protect proteins and other cell components against thermal denaturation.

Di-*myo*-inositol 1,3'-phosphate occurs in the most hyperthermophilic organisms known, *Pyrolobus fumarii* and *Pyrodictium occultum*, and it is widespread among hyperthermophilic archaea and bacteria; moreover, this polyol phosphodiester has never been found in organisms with optimal growth temperature below 60ºC. The accumulation of di-*myo*-inositol 1,3'-phosphate usually increases in response to growth at supra-optimal temperatures, hence a thermoprotective function for this solute has been often postulated.

This thesis is focused on the elucidation of the biosynthetic pathways of three inositol-containing solutes typically found in hyperthermophiles: di-*myo*-inositol 1,3′-phosphate, glycerophospho-*myo*-inositol and 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate. The genes involved in the synthesis of these solutes were identified and the recombinant enzymes characterized. Furthermore, the stereochemistry of these compounds was firmly established by NMR by using substrates specifically labeled with ^{13}C. The archaeon *Archaeoglobus fulgidus* and the bacterium *Thermotoga maritima* were the two target hyperthermophilic organisms used in this work.

Archaeoglobus fulgidus has an optimal growth temperature of about 83ºC and grows optimally in a slightly saline medium, containing 1.9 % NaCl (wt/vol). The solute pool of this archaeon comprises diglycerol phosphate, di-*myo*-inositol 1,3′-phosphate, glycerophospho-*myo*-inositol, and minor amounts of glutamate. The accumulation of solutes is clearly dependent on the type of stress imposed: diglycerol phosphate is the major solute under osmotic stress conditions, while di-*myo*-inositol 1,3′-phosphate accumulates primarily under heat stress conditions. The pathway for the synthesis of di-*myo*-inositol 1,3′-phosphate was established from the relevant enzyme activities in cell extracts determined by using ^{31}P-NMR to monitor substrate consumption and/or product formation. The synthesis proceeds from glucose 6-phosphate via four steps: (1) glucose 6-phosphate was converted into L-*myo*-inositol 1-phosphate by L-*myo*-inositol 1-phosphate synthase; (2) L-*myo*-inositol 1-phosphate was activated to CDP-inositol at the expense of CTP; (3) CDP-inositol was coupled with L-*myo*-inositol 1-phosphate to yield a phosphorylated intermediate, di-*myo*-inosityl 1,3′-phosphate 1′-phosphate (DIPP); (4) finally, this product was dephosphorylated into di-*myo*-inositol 1,3′-phosphate by the action of a phosphatase. The identification of this novel pathway provided the first demonstration of CDP-inositol synthesis in any biological system.

The synthesis of glycerolphospho-*myo*-inositol proceeds via the condensation of CDP-glycerol with L-*myo*-inositol 1-phosphate yielding glycerolphospho-*myo*-inositol 1'-phosphate, which is ultimately dephosphorylated into the final product. The involvement of phosphorylated intermediates in the synthesis of di-*myo*-inositol 1,3'-phosphate and glycerolphospho-*myo*-inositol was firmly demonstrated by NMR analysis of the pure metabolites.

A genomic approach was used to identify the genes implicated in the synthesis of di-*myo*-inositol 1,3'-phosphate. The putative genes for CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase and DIPP synthase from several (hyper)thermophiles (*Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus*, and *Rubrobacter xylanophilus*) were cloned in *E. coli* and the presence of those activities was confirmed in the gene products. The cytidylyltransferase was absolutely specific for CTP and L-*myo*-inositol 1-phosphate; the DIPP synthase used only L-*myo*-inositol 1-phosphate as alcohol acceptor, but CDP-glycerol as well as CDP-L-*myo*-inositol, and CDP-D-*myo*-inositol were recognized as alcohol donors. Genome analysis showed homologues in all organisms known to accumulate di-*myo*-inositol 1,3'-phosphate and for which genome sequences were available. In most cases, the two activities (L-*myo*-inositol 1-phosphate cytidylyltransferase and DIPP synthase) were fused in a single gene product, but separate genes were predicted in *Aeropyrum pernix*, *Thermotoga maritima*, and *Hyperthermus butylicus*.

The solute pool of the hyperthermophilic bacterium *Thermotoga maritima* was comprised α- and β-glutamate, di-*myo*-inositol 1,3'-phosphate, 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (hereafter abbreviated as mannosyl-di-*myo*-inositol 1,3'-phosphate) and minor amounts of 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (abbreviated as

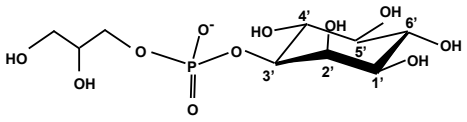
di-mannosyl-di-*myo*-inositol 1,3'-phosphate). *Thermotoga maritima* grows optimally at 80ºC with 2.7% NaCl (wt/vol); when cultivated at supra-optimal temperatures such as 88ºC, it accumulated preferentially mannosyl-di-*myo*-inositol 1,3'-phosphate and di-mannosyl-di-*myo*-inositol 1,3'-phosphate. The synthesis of the mannosylated derivatives of di-*myo*-inositol 1,3'-phosphate was investigated in this work. A putative gene for mannosyl-di-*myo*-inositol phosphate synthase (MDIP synthase) was identified in the genome of *Thermotoga maritima* and the activity was confirmed by functional expression in *E. coli*. The recombinant enzyme used di-*myo*-inositol 1,3'-phosphate and GDP-mannose for the synthesis of mannosyl-di-*myo*-inositol 1,3'-phosphate. The enzyme exhibited maximal activity at 95°C and apparent k_m values of 16 mM and 0.7 mM for di-*myo*-inositol 1,3'-phosphate and GDP-mannose, respectively. Moreover, this enzyme used mannosyl-di-*myo*-inositol 1,3'-phosphate as an acceptor of a second mannose residue, yielding the di-mannosylated derivative of di-*myo*-inositol 1,3'-phosphate. MDIP synthase is a β-1,2-mannosyltransferase, unrelated with known glycosyltransferases. Within the domain *Bacteria*, it is restricted to members of the two deepest lineages, *i.e.*, the *Thermotogales* and the *Aquificales*. Homologues of MDIP synthase were found only in the genomes of *Archaeoglobus profundus* and *Ferroglobus placidus*.

The stereochemical configuration of all the metabolites involved in di-*myo*-inositol 1,3'-phosphate synthesis was established by NMR analysis using L-*myo*-inositol 1-phosphate labeled at C_1 with carbon-13 as substrate for the native and recombinant enzymes. This substrate was produced from D-glucose labeled at C_6 by coupling the activities of hexokinase from *Thermoproteus tenax* and *myo*-inositol 1-phosphate synthase from *Archaeoglobus fulgidus*. We concluded that the two *myo*-inositol moieties in di-*myo*-inositol 1,3'-phosphate had different stereochemical configurations, in

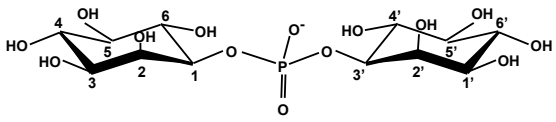
contradiction with reports in the literature. The phosphate group bridges carbon C_1 of one *myo*-inositol moiety with carbon C_3 of the second *myo*-inositol moiety. Thus, the use of the designation di-*myo*-inositol 1,3'-phosphate is recommended to facilitate tracing individual carbon atoms through metabolic pathways. By using a similar methodology the configurations of mannosyl-di-*myo*-inositol 1,3'-phosphate and di-mannosyl-di-*myo*-inositol 1,3'-phosphate were determined; it was firmly established that the mannosyl residue is linked at position 2 of the *myo*-inositol moiety whose carbon C_1 is bound to the phosphate group. The second mannosyl group in di-mannosyl-di-*myo*-inositol 1,3'-phoshate is linked to the first mannose via a C_1-C_2 glycosidic bond. The configuration of the *myo*-inositol moiety in glycerophospho-*myo*-inositol was also established: the phosphate group is bound to position 3 of myo-inositol. The stereochemistry of the glycerol moiety was not determined.

Interestingly, different forms of glycerophospho-*myo*-inositol accumulated in the bacterium *Aquifex pyrophilus* and in the archaeon *Archaeoglobus fulgidus*. Moreover, the configuration of glycerophospho-*myo*-inositol found in the bacterial representative (*Aquifex pyrophilus*) was identical to that of the glycerophospho-*myo*-inositol group in membrane phospholipids of eukaryotes. It is known that bacteria and eukarya use predominantly *sn*-glycerol 3-phosphate, while archaea use *sn*-glycerol 1-phosphate, hence we postulate that the different stereochemistry found in archaea and bacteria is due to the different enantiomers of glycerol phosphate used for the synthesis of CDP-glycerol. Finally, we obtained preliminary results on the heat-shock response of *Thermotoga maritima* with respect to the accumulation of compatible solutes and transcription profiles of genes encoding enzymes involved in the synthesis of di-*myo*-inositol 1,3'-phosphate and mannosylated derivatives.

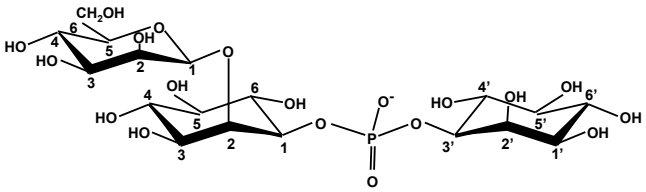
We trust that the results presented in this thesis represent an important step towards the final goal of understanding the whole molecular mechanism of heat stress adaptation, which goes from temperature sensing, to the regulation of gene expression, and finally to the accumulation of compatible solutes especially suited to protecting the cell machinery against heat damage.



Glycero-phospho-*myo*-inositol

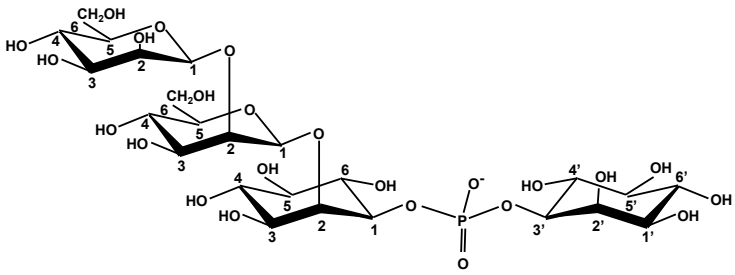


Di-*myo*-inositol 1,3'-phosphate



2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate

Mannosyl-di-*myo*-inositol 1,3'-phosphate



2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosy)-di-*myo*-inositol 1,3'-phosphate

Di-mannosyl-di-*myo*-inositol 1,3'-phosphate

Resumo

A acumulação de compostos orgânicos de baixa massa molecular, denominados solutos compatíveis, é uma estratégia amplamente utilizada na biosfera para contrabalançar flutuações na osmolaridade do meio exterior e assim preservar a viabilidade celular. Curiosamente, a acumulação intracelular de solutos compatíveis também ocorre em resposta a temperaturas de crescimento supra-óptimas, o que fundamenta o pressuposto de que estes compostos participem no processo de termo-adaptação. Organismos hipertermofílicos, *i.e.*, organismos cuja temperatura óptima de crescimento é superior a 80ºC, têm sido isolados de diversos ambientes quentes, quer naturais quer artificiais. Muitos hipertermófilos proliferam em áreas geotermais marinhas e, por isso, são ligeiramente halofílicos. Portanto, estes organismos têm que resistir a variações na salinidade do meio exterior, geralmente acumulando solutos compatíveis como estratégia de defesa. É interessante notar que organismos hipertermofílicos acumulam preferencialmente solutos carregados negativamente, tais como fosfato de diglicerol, fosfato de 1,3'-di-*myo*-inositol ou manosilglicerato, em vez de solutos neutros ou "zwiteriónicos" (glicerol, trealose, *myo*-inositol ou ectoínas), tipicamente encontrados em organismos adaptados a temperaturas moderadas. Neste contexto, é pertinente questionar se estes solutos aniónicos terão sido seleccionados durante o processo evolutivo pela sua eficácia superior na protecção de proteínas e outros componentes celulares contra os efeitos nocivos das altas temperaturas.

O fosfato de 1,3'-di-*myo*-inositol é acumulado por *Pyrolobus fumarii* e *Pyrodictium occultum*, os dois organismos que detêm recordes de hipertermofilia, exibindo temperaturas óptimas de crescimento de 105ºC. Além disso, este soluto encontra-se em grande número de bactérias e

arqueões hiperthermofílicos, nunca tendo sido detectado em organismos com temperatura óptima de crescimento inferior a 60ºC. O conteúdo intracelular de fosfato de 1,3'-di-*myo*-inositol geralmente aumenta em resposta a agressão térmica, pelo que frequentemente lhe seja atribuída uma função termo-protectora. No entanto, ainda não existe evidência experimental suficiente para fundamentar convincentemente esta hipótese.

Esta tese tem por objectivo principal a elucidação das vias biossintéticas de três solutos derivados de inositol e tipicamente encontrados em hipertermófilos: fosfato de 1,3'-di-*myo*-inositol, glicero-fosfo-*myo*-inositol e fosfato de 2-(*O*-β-D-manosil)-1,3'-di-*myo*-inositol. Os genes envolvidos na síntese destes solutos foram identificados e expressos em *Escherichia coli*, e os respectivos produtos (enzimas biossintéticas) foram caracterizadas em detalhe. Com base nesta informação, a estereoquímica destes compostos foi firmemente estabelecida por NMR, usando substratos especificamente marcados com carbono-13.

Archaeoglobus fulgidus e *Thermotoga maritima*, representantes respectivamente dos Domínios *Archaea* e *Bacteria* na árvore filogenética, foram os dois hipertermófilos seleccionados como alvos principais deste estudo. *Archaeoglobus fulgidus* cresce optimamente a 83ºC, em meio ligeiramente salino, contendo cerca de 1,9% (m/v) de NaCl. Este arqueão acumula maioritariamente fosfato de diglicerol, fosfato de 1,3'-di-*myo*-inositol e glicero-fosfo-*myo*-inositol, para além de quantidades menores de glutamato. A acumulação de solutos foi claramente dependente do tipo de agressão imposta: enquanto que fosfato de diglicerol é o soluto maioritário em resposta a agressão salina, fosfato de 1,3'-di-*myo*-inositol é acumulado preferencialmente em condições de agressão térmica. A via de síntese do fosfato de 1,3'-di-*myo*-inositol foi estabelecida através da medição das actividades enzimáticas relevantes em extractos celulares, usando ^{31}P-NMR

para avaliar o consumo de substratos e/ou formação de produtos. A síntese processa-se a partir de glucose 6-fosfato em quatro passos reaccionais: (1) a glucose 6-fosfato é convertida em L-*myo*-inositol 1-fosfato na reacção catalisada pela sintase do L-*myo*-inositol 1-fosfato; (2) o L-*myo*-inositol 1-fosfato é activado a CDP-inositol à custa da energia de ligação de uma molécula de CTP; (3) CDP-inositol é ligado a L-*myo*-inositol 1-fosfato, dando origem ao intermediário fosforilado, fosfato de 1,3'-di-*myo*-inoitol 1'-fosfato (DIPP); (4) finalmente, este produto é desfosforilado por acção de uma fosfatase para produzir fosfato de 1,3'-di-*myo*-inositol. O presente trabalho constitui a primeira demonstração da síntese de CDP-inositol em sistemas biológicos.

A síntese de glicero-fosfo-*myo*-inositol processa-se através da condensação de CDP-glicerol com L-*myo*-inositol 1-fosfato para formar glicero-fosfo-*myo*-inositol 1'-fosfato, que por sua vez é desfosforilado, originando o produto final. O envolvimento de intermediários fosforilados na síntese de fosfato de 1,3'-di-*myo*-inositol e de glicero-fosfo-*myo*-inositol ficou definitivamente demonstrado através da caracterização estrutural dos respectivos metabolitos por NMR multinuclear.

Uma abordagem genómica permitiu propor genes presumivelmente envolvidos na síntese do fosfato de 1,3'-di-*myo*-inositol. Os melhores candidatos para codificarem as enzimas citidililtransferase: L-*myo*-inositol 1-fosfato e sintase de DIPP em vários (hiper)termófilos (*Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus* e *Rubrobacter xylanophilus*) foram clonados em *E. coli* tendo-se confirmado a presença das actividades enzimáticas suspeitadas. A citidililtransferase é absolutamente específica para os substratos CTP e L-*myo*-inositol 1-fosfato; a sintase do DIPP usa apenas L-*myo*-inositol 1-fosfato como aceitador do grupo álcool, mas reconhece CDP-glicerol, bem como CDP-L-*myo*-inositol e CDP-D-

myo-inositol como dadores do grupo álcool. Na maioria dos organismos, as duas actividades (citidililtransferase de L-*myo*-inositol 1-fosfato e sintase do DIPP) encontram-se fundidas num único gene, no entanto em *Aeropyrum pernix*, *Thermotoga maritima* e *Hyperthermus butylicus* existem em genes isolados.

O conjunto de solutos acumulados pela bactéria hipertermofílica *Thermotoga maritima* compreende α- e β-glutamato, fosfato de 1,3'-di-*myo*-inositol, fosfato de 2-(*O*-β-D-manosil)-1,3'-di-*myo*-inositol (abreviado como fosfato de manosil-1,3'-di-*myo*-inositol) e pequenas quantidades de fosfato de 2-(*O*-β-D-manosil-1,2-*O*-β-D-manosil)-1,3'-di-*myo*-inositol (abreviado como fosfato de di-manosil-1,3'-di-*myo*-inositol). *Thermotoga maritima* cresce optimamente a 80ºC com 2,7% (m/v) de NaCl; quando cultivada acima da temperatura óptima, por exemplo a 88ºC, acumula preferencialmente fosfato de manosil-1,3'-di-*myo*-inositol e fosfato de di-manosil-1,3'-di-*myo*-inositol. Neste trabalho foi investigada a síntese destes derivados manosilados de fosfato de 1,3'-di-*myo*-inositol. No genoma de *Thermotoga maritima* foi identificado um candidato promissor para codificar a sintase do fosfato de manosil-1,3'-di-*myo*-inositol (MDIP sintase) e a actividade foi confirmada através da produção da proteína funcional em *E. coli*. A enzima recombinante usa fosfato de 1,3'-di-*myo*-inositol e GDP-manose para a síntese de fosfato de manosil-1,3'-di-*myo*-inositol, num único passo reaccional. A mesma enzima tem actividade máxima a 95ºC e valores de k_m aparente de 16 mM e 0,7 mM para fosfato de 1,3'-di-*myo*-inositol e GDP-manose, respectivamente. Além disso, a enzima usa fosfato de manosil-1,3'-di-*myo*-inositol como aceitador de um segundo resíduo de manose, originando fosfato de di-manosil-1,3'-di-*myo*-inositol. A MDIP sintase é uma β-1,2-manosiltransferase totalmente distinta das outras glicosiltransferases conhecidas, o que torna a sua descoberta particularmente interessante. Esta enzima parece confinada a membros do

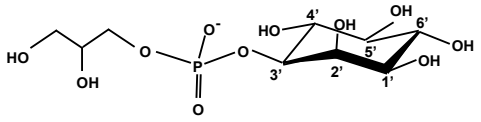
Domínio *Bacteria*, encontrando-se apenas em representantes das duas linhagens mais antigas, *i.e.*, os *Thermotogales* e os *Aquificales*. Sequências semelhantes à da MDIP sintase foram encontradas apenas nos genomas de *Archaeoglobus profundus* e de *Ferroglobus placidus*.

A configuração estereoquímica de todos os metabolitos envolvidos na síntese do fosfato de 1,3'-di-*myo*-inositol foi estabelecida por análise de NMR, usando L-*myo*-inositol 1-fosfato marcado com carbono-13 no C_1, como substrato para enzimas nativas e também para enzimas recombinantes. Este substrato foi obtido a partir de D-glucose 6-fosfato marcada no C_6, acoplando as actividades da hexocinase de *Thermoproteus tenax* e da sintase do L-*myo*-inositol 1-fosfato de *Archaeoglobus fulgidus*. Concluiu-se que as duas unidades de L-*myo*-inositol fosfato tinham configurações estereoquímicas diferentes, em discordância com resultados encontrados na literatura. O grupo fosfato está ligado ao C_1 de uma molécula de *myo*-inositol e ao C_3 da segunda molécula de *myo*-inositol. Recomenda-se aqui a designação fosfato de 1,3'-di-*myo*-inositol para facilitar o seguimento de átomos de carbono específicos ao longo das vias metabólicas. Metodologia semelhante foi usada para determinar a configuração do fosfato de manosil-1,3'-di-*myo*-inositol e do fosfato de di-manosil-1,3'-di-*myo*-inositol. Além disso, ficou firmemente estabelecido que o grupo manosilo está ligado na posição 2 da molécula de *myo*-inositol que por sua vez está ligada ao fosfato na posição C_1. O segundo grupo manosilo no fosfato de di-manosil-1,3'-di-*myo*-inositol está ligado ao primeiro através de uma ligação glicosídica C_1-C_2. A configuração da unidade *myo*-inositol no glicero-fosfo-*myo*-inositol foi também elucidada: o grupo fosfato está ligado na posição 3 do *myo*-inositol. A estereoquímica da molécula de glicerol permanece incerta.

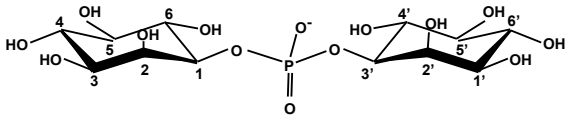
É interessante notar que diferentes formas de glicero-fosfo-*myo*-inositol são sintetizadas na bactéria *Aquifex pyrophilus* e no arqueão

Archaeoglobus fulgidus. Para além disso, a configuração do glicero-fosfo-*myo*-inositol encontrado na bactéria *Aquifex pyrophilus* é idêntica à da unidade glicero-fosfo-*myo*-inositol encontrada em fosfolípidos membranares de eucariontes. Sabe-se que bactérias e eucariontes usam preferencialmente *sn*-glicerol 3-fosfato, enquanto que os arqueões usam *sn*-glicerol 1-fosfato, pelo que postulamos que a diferença na estereoquímica de glicero-fosfo-*myo*-inositol proveniente de bactérias ou de arqueões é devida à utilização de diferentes enantiómeros de glicerol fosfato na síntese de CDP-glicerol. Por fim, obtivemos resultados preliminares para a resposta a agressão térmica em *Thermotoga maritima* com respeito à acumulação de solutos compatíveis e aos perfis de transcrição dos genes que codificam as enzimas envolvidas na síntese de fosfato de 1,3'-di-*myo*-inositol e dos derivados manosilados.

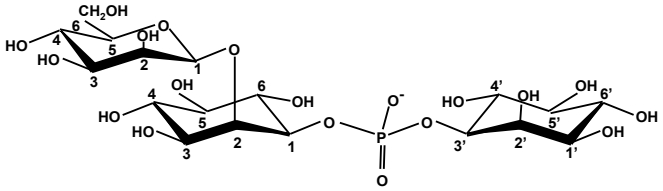
Pensamos que os resultados apresentados nesta tese constituem um passo importante para a compreensão cabal dos mecanismos moleculares subjacentes à adaptação a agressão térmica, desde a percepção da condição de agressão, à regulação da expressão génica, e finalmente à acumulação de solutos protectores da maquinaria celular contra danos induzidos por temperaturas elevadas.



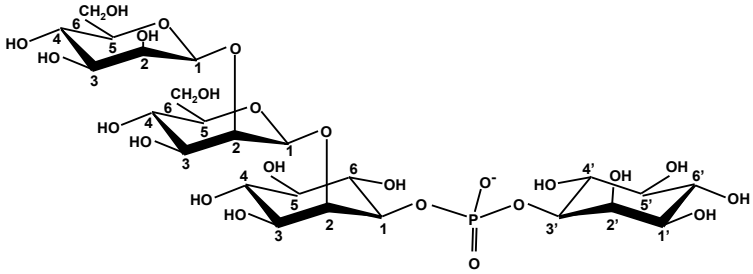
Glicero-fosfo-*myo*-inositol



Fosfato de 1,3'-di-*myo*-inositol



Fosfato de 2-(*O*-β-D-manosil)-1,3'-di-*myo*-inositol
(Fosfato de manosil-1,3'-di-*myo*-inositol)



Fosfato de 2-(*O*-β-D-manosil-1,2-*O*-β-D-manosil)-1,3'-di-*myo*-inositol
(Fosfato de di-manosil-1,3'-di-*myo*-inositol)

Contents

Abbreviations

ADP-glucose, adenosine 5’-diphosphoglucose

ATP, adenosine 5'-triphosphate

BnBr, benzyl bromide

CDP-glycerol, cytidine 5’-diphosphoglycerol

CDP-inositol, cytidine 5’-diphospho-*myo*-inositol

CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

CMP, cytidine 5’-monophosphate

COSY, homonuclear correlation spectroscopy

CTP, cytidine 5’-triphosphate

DCM, dichloromethane

DDM, n-dodecyl beta-D-maltoside

DGP, di-glycerol phosphate

DGPP, X,X’-diglycerol phosphate X’-phosphate

DHAP, dihydroxyacetonephosphate

DIP or di-*myo*-inositol phosphate or di-*myo*-inositol 1,3’-phosphate, L,L-di-*myo*-inositol 1,3’-phosphate

DIPP or di-*myo*-inositol phosphate phosphate or di-*myo*-inositol 1,3’-phosphate 1’-phosphate, L,L-di-*myo*-inositol 1,3’-phosphate 1’-phosphate

DIPPS or DIPP synthase, CDP-inositol:L-*myo*-inositol 1-phosphate transferase

DMF, dimethylformamide

DSS, 2,2-dimethyl-2-silapentane-5-sulfonate or 3-(trimethylsilyl)-1-propanesulfonic acid or 4,4-dimethyl-4-silapentane-1-sulfonic acid

EDTA, ethylenediaminetetraacetic acid

EtOH, ethanol

Fos-choline-12, n-dodecylphosphocholine

G1PDH, *sn*-glycerol 1-phosphate dehydrogenase

G3PDH, *sn*-glycerol 3-phosphate dehydrogenase

GCT, CTP:L-glycerol 3-phosphate cytidylyltransferase

GDP, guanosine 5′-diphosphate

GDP-glucose, guanosine 5′-diphosphoglucose

GDP-mannose, guanosine 5′-diphospho-D-mannose

Glc-6-P, glucose 6-phosphate

sn-gly-1-P, *sn*-glycerol 1-phosphate

sn-gly-3-P, *sn*-glycerol 3-phosphate

Gly-P, *rac*-glycerol 1-phosphate

GPI or glycero-phospho-*myo*-inositol, glycero(X)-phospho(3′)-L-*myo*-inositol

GPIP, glycero(X)-phospho(3′)-L-*myo*-inositol(1′)-phosphate

GTP, guanosine 5'-triphosphate

L-Inos-1-P, L-*myo*-inositol 1-phosphate

LDAO, N,N-dimethyldodecylamine N-oxide

HMBC, heteronuclear multiple bond correlation spectroscopy

HMQC, heteronuclear multiple quantum coherence spectroscopy

HSQC, heteronuclear single quantum coherence spectroscopy

IMP, inositol monophosphatase (EC 3.1.3.25)

IPCT/DIPPS, CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase/ CDP-inositol:L-*myo*-inositol 1-phosphate transferase

IPTG, isopropyl β-D-1-thiogalactopyranoside

MCPBA, *m*-chloroperbenzoic acid

MDIP or mannosyl-di-*myo*-inositol 1,3′-phosphate, 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate

MDIP synthase, mannosyl-di-*myo*-inositol 1,3′-phosphate synthase

Mega-9, nonanoyl-N-methylglucamide

MeOH, methanol

MIPS, L-*myo*-inositol 1-phosphate synthase or *myo*-inositol 1-phosphate synthase (EC 5.5.1.4)

MMDIP or di-mannosyl-di-*myo*-inositol 1,3′-phosphate, 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate

MOPS, 3-(*N*-morpholino)propanesulfonic acid

NAD+, nicotinamide adenine dinucleotide

NMR, nuclear magnetic resonance spectroscopy

Pi, inorganic phosphate

PPI, pyrophosphate

r.t., room temperature

SDS, sodium dodecyl sulfate

TBAF, tetrabutylammonium fluoride

TBDMSOTf, *tert*-butyldimethylsilyl trifluoromethanesulfonate

TDP-glucose, thymidine 5´-diphosphoglucose

TFA, trifluoroacetic acid

TOCSY, total correlation spectroscopy

tri-mannosyl-di-*myo*-inositol 1,3′-phosphate, 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl 1,2-*O*-β-D -mannosyl)-di-*myo*-inositol 1,3'-phosphate

p-TsOH, *p*-toluenesulfonic acid

Tris, 2-amino-2-(hydroxymethyl) 1,3-propanediol

Triton X-100, 4-(1,1,3,3-tetramethylbutyl)phenyl-polyethylene glycol

TSPSA, 3-(trimethylsisyl) propane sulfonic acid (sodium salt)

UDP-glucose, uridine 5′-diphosphoglucose

UTP, uridine 5'-triphosphate

CHAPTER 1

General introduction

Chapter 1 – Contents

Life in extreme environments

Many organisms are adapted to thrive in habitats with non-conventional conditions. Such hostile environments include physical (temperature, pressure or radiation), and geochemical (pH, salinity, desiccation or oxygen availability) extremes. The harsh conditions of those habitats totally abolish the concept of "normal" life, as we know it. These inhospitable conditions can be found in the deep sea, where the pressure is very high and the environmental temperature may be as low as 1ºC, or as high as 400ºC in the vicinity of hydrothermal vents. Other sites include hot springs and geysers, characterized by hot water, steam and sometimes low pH. Hypersaline environments include salt lakes, salt ponds, and marine salterns, like the Dead Sea in Israel or the Soda Lake in Egypt (Falb et al. 2008). Places where liquid water is a limiting factor include deserts that are extremely dry, and hot (Atacama Desert) or cold (Antarctica valleys) (Rothschild and Mancinelli 2001). Furthermore, astrobiology (or exobiology), a rapidly emerging field dedicated to the study of the origin, evolution, distribution and future of the universe, also "labels" the Cosmos as an extreme environment (Cavicchioli 2002, Pikuta and Hoover 2007). The unique organisms that thrive in these extreme conditions are called "extremophiles" and they are classified according to the optimal growth conditions.

Considering general environmental parameters, such as temperature, salinity, pH or pressure, extremophiles fall into the following categories: hyperthermophiles (optimal growth temperature above 80°C); thermophiles (optimal growth temperature between 50ºC and 80°C); psychrophiles (optimal growth temperature below 15°C); halophiles (optimal growth requiring between 0.2 and 1.7 M NaCl) or extreme halophiles (optimal growth requiring higher than 1.7 M NaCl); alkaliphiles (optimal growth requiring pH above 9);

acidophiles (optimal growth requiring pH below 4); and piezophiles (optimal growth requiring more than 40 MPa) (Rothschild and Mancinelli 2001, Pikuta and Hoover 2007, Bowers and Wiegel 2011).

Some organisms can live in environments where more than one extreme condition exists. Acidophilic hyperthermophiles include species of the genus *Sulfolobus* whose physiology is characterized by growth at low pH and high temperature, or species of the genus *Acidianus*, which grow at pH 2 and 90ºC (Grogan 1989, Cavicchioli 2002). Haloalkaliphiles include several species of the genus *Natronococcus* that live optimally in habitats with 20% (wt/vol) NaCl and pH 9, or *Natronomonas pharaonis* which inhabits an environment with 35% (wt/vol) NaCl and a pH of 8.5 (Kanal et al. 1995, Xu et al. 1999, Falb et al. 2008). One of the most recent examples of an organism living in two extremes is *Pyrococcus yayanosii*, the first obligate piezophilic hyperthermophile isolated from a deep-sea hydrothermal vent, showing optimal pressure and temperature for growth of 52 MPa and 98ºC, respectively (Birrien et al. 2011). It is also possible to find organisms with extreme resistance to ionizing radiation like *Deinococcus radiodurans,* which was the first radiotolerant bacterium to be isolated, or the hyperthermophile *Thermococcus gammatolerans*, which tolerates high doses of γ-radiation (Rothschild and Mancinelli 2001, Pikuta and Hoover 2007). As mentioned above, water limitation can also be considered as an extreme condition, and yet it is possible to find life in sites with extreme dryness. Organisms that thrive in extremely dry environments (desiccation conditions) are called "xerophiles", and as examples we can find the classical cases of cacti and some fungi (Rothschild and Mancinelli 2001). Extremophiles have representatives in the three domains of Life, the *Archaea*, the *Bacteria* and the *Eukarya*. Nevertheless, the record holders of extremophily are found in the domain *Archaea*.

The three domains of Life

The proposal for the division of Life on Earth into a higher order taxa, the "domain", was put forward in 1990 by Carl R. Woese and colleagues. Based on molecular structures and sequences that evidenced the existence of profound differences particularly among prokaryotes, these authors proposed the division of Life into three domains: *Bacteria*, *Eukarya*, and *Archaea* (Woese et al. 1990). The molecular information available at that time to fundament this division, namely the striking differences found in the small subunit of rRNA sequences and structures of several organisms, has been validated by comparative genomics and currently, the division of the biological world into three domains is widely accepted (Woese 1993, Gribaldo and Brochier-Armanet 2006). Some distinctive characteristics of the three domains are summarized in Table 1.1.

In a simplistic view, it can be considered that the *Archaea* operate in a bacterial-like environment using eukaryotic tools. A remarkable example is the DNA replication apparatus. Many of the mechanisms used by the *Archaea* to process information are eukaryotic-related and totally different from their bacterial counterparts. On the other hand, archaea share many characteristics with bacteria, as size and organization of their chromosomes or the presence of polycistronic mRNAs (Gribaldo and Brochier-Armanet 2006).

Table 1.1. Particular characteristics of the three domains of Life on Earth (adapted from Purves et al. 1998).

Characteristic	Domain		
	Bacteria	***Archaea***	***Eukarya***
Membrane-enclosed nucleus or organelles	-	-	+
N-acetylmuramic acid in the cell wall	+	-	-
Membrane lipids	Ester-linked (Unbranched)	Ether-linked (Branched)	Ester-linked (Unbranched)
Ribosomes	70S	70S	80S
Operons	+	+	-
Initiator tRNA	Formylmethionine	Methionine	Methionine

+, present; -, absent.

However, archaea have earned their position as a distinctive domain not only due to the distinct features of the small subunit rRNA sequence and structure, but also due to the presence of unique genomic signatures and biochemical properties. For example, archaeal organisms harbor a unique type of membrane polar lipids; in the *Archaea*, the phospholipid backbone is built on *sn*-glycerol 1-phosphate while in *Bacteria* and *Eukarya* the stereoisomer *sn*-glycerol 3-phosphate takes place. Furthermore, archaeal lipids have methyl-branched isoprenoid side chains bound to glycerol phosphate by ether-linkages, while in *Bacteria* and *Eukarya* the side chains of membrane phospholipids are made of straight chain fatty-acids linked to the glycerol phosphate backbone by ester linkages (Koga and Morri 2007, Peretó et al. 2004).

Based on phylogenetic analyses of 16S rRNA sequences, originally the *Archaea* were dividied into two major phyla, the *Euryarchaeota* and the *Crenarchaeota*. The *Euryarchaeota* is a phenotypically heterogeneous group, comprising organisms that inhabit a broad range of environments. Such organisms include a few hyperthermophiles, methanogens, thermoacidophiles,

and halophiles. On the contrary, the *Crenarchaeota* include organisms isolated exclusively from thermophilic niches, and comprise hyperthermophiles, thermoacidophiles and sulfur-dependent organisms (Woese et al. 1990, Grabowski and Kelman 2003). Recent metagenomic analyses revealed that some archaea share *Crenarchaeota* and *Euryarchaeota*-specific genomic traits. Hence, there was need for additional phyla divisions. The first one to emerge was the phylum *Korarchaeota*, composed by deeply-branched archaea, including Candidatus (Ca.) *Korarchaeum cryptofilum* (Barns et al. 1996, Elkins et al. 2008, Nunoura et al. 2010). The emergence of new groups of uncultivated mesophilic *Crenarchaeota* and the sequenced genome of Ca. *Cenarchaeum symbiosum* led to the proposal of another phylum, the *Thaumarchaeota*. The proposal for this phylum has been further supported by the genome sequences of other mesophilic archaea, such as Ca. *Nitrosopumilus maritimus* and Ca. *Nitrososphaera gargensis* (Brochier-Armanet et al. 2008, Nunoura et al. 2010). Another phylum, the *Nanoarchaeota*, yet with a debatable position, has been proposed based on phylogeny of the smallest known living cell, the *Nanoarchaeum equitans*, a tiny hyperthermophile that grows and divides at the surface of crenarchaeal *Ignicoccus* species, and can not be cultivated independently (Huber et al. 2002, Brochier et al. 2005, Paper et al. 2007).

Without any doubt the domain *Archaea* harbors a complex and diverse group of organisms, whose position so close to the root of the Universal Tree of Life, might well reveal some clues about the origin of life on Earth. Furthermore, this domain comprises the majority of extreme organisms, which are sources of extreme biocatalysts with unique and useful properties for industrial applications.

Thermophiles and hyperthermophiles

The first hyperthermophile was discovered in the early eighties. During a trip to Iceland, Karl Stetter and colleagues collected several samples of boiling water and mud from which they isolated *Methanothermus fervidus*, the first organism ever isolated that grew up to 97ºC and exhibited the fastest growth at 82ºC (Stetter et al. 1981, Stetter 2006). Until then, the record was held by *Sulfolobus acidocaldarius*, isolated in 1972, that although it was able to grow up to 87ºC, the optimal growth temperature was 75ºC (Brock et al. 1972). The way to the world of hyperthermophily had just been opened. The boom in the isolation and characterization of hyperthermophiles happened in the following 25 years, especially by the group of Karl Stetter, who isolated more than 50 species of hyperthermophiles (Stetter 2006). Presently, over 100 species have been described and around 80 have their genomes fully sequenced.

Hyperthermophiles are located near the root of a 16S rRNA-based phylogenetic tree (Fig. 1.1), representing all short and deep-branches within the archaeal and bacterial domains, indicating a slow rate of evolution. Due to their location in the Tree of Life and their primitive lifestyle, it has been argued that the last universal common ancestor may have been indeed a hyperthermophile (Forterre 1996, Di Giulio 2003). However, this is a matter of intense controversy and the position of the last universal common ancestor is yet to be determined since updated theories are constantly emerging (Glansdorff et al. 2008, Di Giulio 2011).

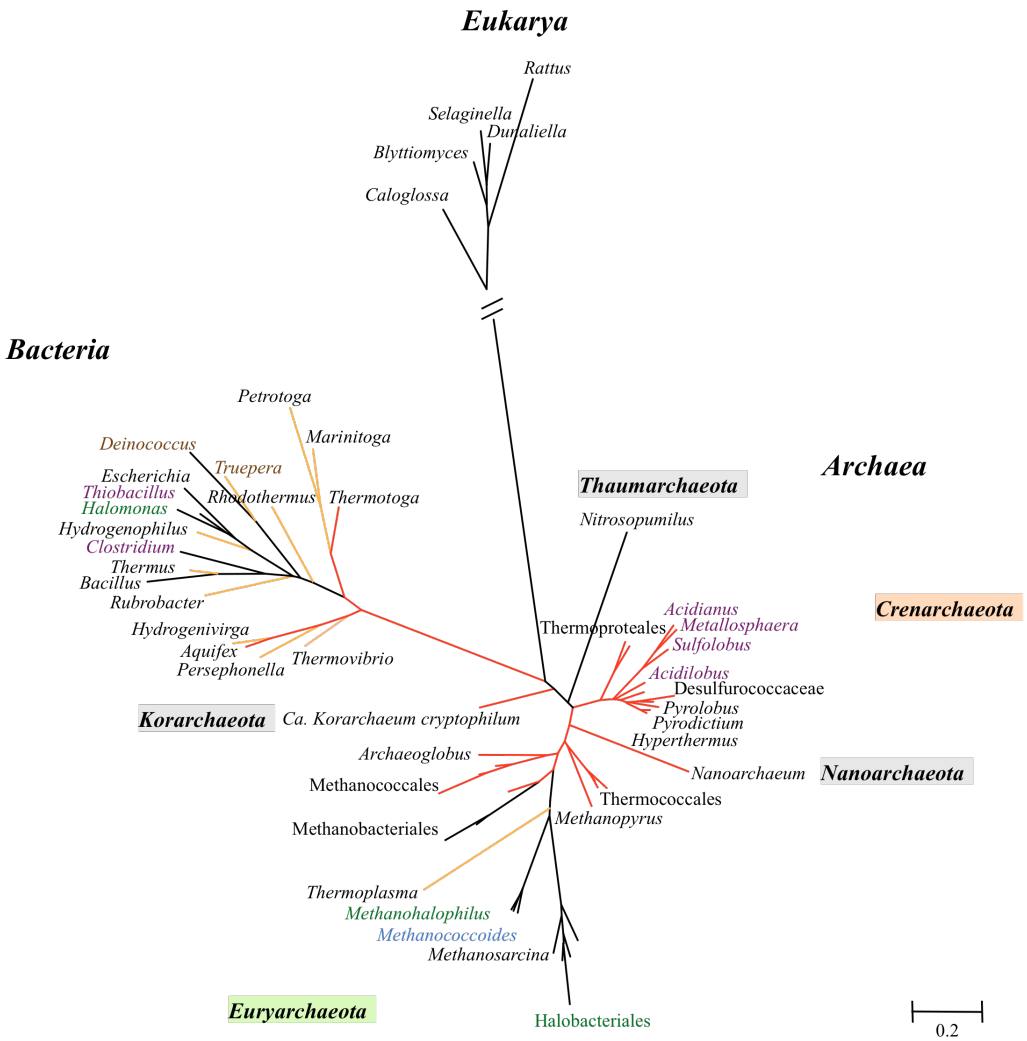


Figure 1.1. Tree of Life based on 16S rRNA sequences of *Bacteria*, *Archaea* and *Eukarya*. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The 59 nucleotide 16S rRNA sequences were retrieved from Silva database (http://www.arb-silva.de/) and the analysis involved one species of the genus represented in the Tree. Sequences were aligned using MUSCLE and the maximum likelihood tree was computed by PHYML. Hyperthermophilic and thermophilic branches are represented in red and orange, respectively. Thermoproteales include the genera *Caldivirga*, and *Pyrobaculum*; Desulfurococcaceae include the genera *Ignisphaera*, *Aeropyrum*, *Ignicoccus,* and *Staphylothermus*; Thermococcales include the genera *Palaeococcus*, *Thermococcus*, and *Pyrococcus*; Methanococcales include the genera *Methanocaldococcus*, *Methanotorris*, and *Methanococcus*; Methanobacteriales include the genera *Methanosphaera*,

Methanobacterium, and *Methanothermus*; Halobacteriales include the genera *Halobacterium*, *Natronobacterium*, *Haloferax*, *Natronococcus,* and *Halococcus*.

Within the domain *Bacteria*, hyperthermophiles are restricted to members of the orders *Aquificales* and *Thermotogales*, more specifically to the genera *Aquifex* and *Thermocrinis* for the *Aquificales*, and the genus *Thermotoga* for the *Thermotogales*. *Thermotoga maritima* and *Aquifex pyrophilus* withstand the highest growth temperatures, 90ºC and 95ºC, respectively (with optimal growth temperatures of 80ºC for *Thermotoga maritima* and 85ºC for *Aquifex pyrophilus*). The position of these bacteria near the root of the Tree of Life makes them good candidates for evolutionary studies (Vieille and Zeikus 2001, Stetter 2006). One interesting aspect that arouse from the sequencing of the *Thermotoga maritima* genome was the abundance of foreign genes originating from archaeal species or other bacterial species, which supported the occurrence of lateral gene transfer between (hyper)thermophilic bacteria and archaea (Nelson et al. 1999, Mongodin et al. 2005).

Hyperthermophiles are much better represented among the domain *Archaea* than in the domain *Bacteria*. Species of the genera *Thermococcus*, *Pyrolobus*, *Hyperthermus*, and *Sulfolobus* are examples of hyperthermophiles that belong to the phylum *Crenarchaeota*; *Thermococcus*, *Pyrococcus*, and *Archaeoglobus* are examples of genera belonging to the phylum *Euryarchaeota*, and the phylum *Nanoarchaeota* is represented by *Nanoarchaeum equitans* (Stetter 2006, Burgess et al. 2007). *Pyrolobus fumarii*, with an optimal growth temperature of 106ºC is one of the most hyperthermophilic organisms known to date (Blöchl et al. 1997). This archaeon is optimally adapted to superheated habitats since it can grow up to 113ºC and does not grow below 90ºC. In 2003, it was reported the isolation

of a new archaeal microbe, strain 121, which supposedly was able to grow up to 121ºC (Kashefi and Lovley 2003), however, this result could not be validated by the scientific community since the strain was not made available. The domain *Eukarya* is represented by the peculiar "worm" *Alvinella pompejana*, isolated from hydrothermal vents, which withstands temperatures above 80ºC and live symbiotically with hyperthermophilic bacteria that grow on its tail (Cary et al. 1998, Burgess et al. 2007).

Hyperthermophiles can thrive in a diverse array of biotopes and within each type several factors such as pH, pressure, or NaCl concentration, introduce additional variety. Natural habitats include terrestrial hot springs, geysers, sulfataras, marine hydrothermal vents or even subsurface environments as petroleum reservoirs or geothermally heated lakes. Artificial water-containing hot environments include hot outflows from geothermal power plants, coal refuse piles or even household compost piles (Burgess et al. 2007). The bottleneck for the discovery of new species does not reside in sampling, because nowadays it is possible to access hyperthermophilic biotopes almost routinely. The major problem in studying hyperthermophilic physiology has been, and still is, isolation and cultivation of new species. A good example is the bacterium *Thermocrinis rubber*, described by Brock in the sixties, which was successfully cultivated by Huber and co-workers only in 1998 (Brock 1967, Huber et al. 1998).

The energy sources and lifestyle of hyperthermophiles are very simple. These organisms are mainly chemolithoautotrophs, using inorganic compounds as energy source and CO_2 as the single carbon source to build up organic cell material. Molecular hydrogen (H_2) is an important electron donor, but sulfide, sulfur, and ferrous iron can be used for that matter as well. Energy yielding reactions are mainly anaerobic and include nitrate (NO_3^-), ferric iron ($Fe(OH)_3$), sulfate (SO_4^{2-}), sulfur (S^0), and carbon dioxide (CO_2) as

electron acceptors. However, microaerophilic respiration can be found for example in *Aquifex pyrophilus*, which may use oxygen as electron acceptor. Some hyperthermophiles are facultative heterotrophs, using organic material (from dying cells, for example) instead of inorganic nutrients as electron donors, and conserve energy either by aerobic respiration, different types of anaerobic respiration, or even by fermentation (Stetter 2006).

Hyperthermophiles are adapted to distinct environmental factors such as pH, redox potential, salinity, minerals, gas composition and temperature. These extraordinary organisms are unable to grow below 80ºC but amazingly they can survive at ambient temperature for decades (Stetter 2006). The primitive and yet extraordinary physiology and biochemistry of these organisms has drawn the attention of scientists to several lines of investigation, like the evolutionary aspects of Life on Earth, adaptation strategies to extreme environments, and to the biotechnological potential of hyperthermophiles (Huber et al. 2000).

Extreme biocatalysts

Extremophiles have deserved great attention, as they are a valuable source of new enzymes with great potential for industrial applications. Extremophilic organisms have adapted to diverse extreme environments and thus, it is expected that their enzymes work in similar conditions (van den Burg 2003). Many ordinary available enzymes do not withstand the conditions of industrial reactions, but biocatalysts of extremophiles are able to circumvent that limitation and are suitable for utilization under the common harsh conditions of industrial processes. For example, halophilies provide enzymes that are often used in reactions performed in non-aqueous medium due to their capacity to work in reduced water activity conditions, while psychrophiles are

sources of proteases, lipases or amylases for utilization in detergent, food or cosmetics industry (Antranikian et al. 2005). The degradation process of polymers such as starch or cellulose is improved if performed at high temperature, for the sake of solubility improvement and consequent better substrate accessibility. Hence enzymes that are resistant and active at high temperatures are preferred for such application. Furthermore, a few studies were also conducted for the utilization of live cultures of extremophiles in industrial processes, namely desulfurization of ground rubber with the hyperthermophilic *Pyrococcus furiosus*, bioremediation of waters following water spills by psychrophilic Antarctic bacteria, or production of β-carotenes and glycerol by the halophilic *Dunaliella salina* (Hubber and Stetter 1998, Bredberg et al. 2001, Rothschild and Mancinelli 2001).

In particular, thermophiles and hyperthermophiles, enriched the biotechnological industry with a variety of enzymes, including lipases, esterases, glycosyl hydrolases (*e.g.* cellulases, chitinases, xylanases), proteases or, the remarkable case of DNA-polymerases. Targets of these enzymes include detergent industry, starch, chitin or cellulose processing, paper bleaching, and molecular biology applications (Egorova and Antranikian 2005). By performing industrial processes at high temperature there is reduced risk of contamination, superior transfer rates, lower viscosity and higher solubility of substrates (van den Burg 2003). On the other hand, the utilization of cultures of hyperthermophiles in industrial processes is hampered by two major drawbacks: the poor growth yields and the scarcity of tools to genetically engineer these organisms to better suit our goals. Thus far, gene disruption is only possible in *Thermococcus kodakarensis* and *Sulfolobus solfataricus* (Sato et al. 2003, Worthington et al. 2003). Finding methods to genetically manipulate hyperthermophiles is a field with a very slow progress and certainly deserves more attention. Meanwhile, deciphering the strategies

and the molecular mechanisms adopted by these organisms to support life under extreme conditions is particularly interesting and poses a major challenge to the scientific community.

Adaptation to extreme environments

Intrinsic factors for thermoadaptation

The driving forces that permit extremophiles to thrive in environments with extremely low pH, incredibly high salt concentration or temperatures around 100ºC remain unclear. The key for success of these organisms is the ability to keep the external conditions outside the cell. For example, in acidic environments besides low pH, the metal solubility is very high, so acidophiles not only have to handle high concentration of H^+ in the extracellular environment but also resist high metal concentration. These organisms have biochemical mechanisms to circumvent these problems, which include the presence of extracellular proteins adapted to low pH, membranes with special permeability characteristics, efficient proton pumps, and several mechanisms that promote metal tolerance. Together, these strategies allow for the maintenance of a neutral cytoplasm in acidophiles and the growth of these organisms in strongly acidic habitats (Johnson 2007). However, thermophiles and hyperthermophiles cannot escape heat, hence they must have all cell components resistant to high temperature. Nucleic acids, lipids and proteins are usually very heat sensitive, but apparently these heat-loving organisms have adopted special ways to cope with these conditions. Although the molecular basis for thermoadaptation is still under investigation, some key factors have been pointed as intrinsic adaptation mechanisms. It is difficult to separate the features that are associated with hyperthermophily from those

that are related with phylogeny, namely due to the close relationship between hyperthermophiles and the domain *Archaea*.

i) Lipids

The maintenance of membrane fluidity, ion gradients or transport systems across the cellular membrane is vital for cell growth. However, high temperature affects chemical and physical properties of this barrier, causing increased membrane fluidity and permeability. Deficient ion gradients are sufficient to abolish cell growth. In fact, membrane permeability may as well be one of the most important factors in determining the upper temperature limit for life, given the vulnerability of ion gradients at elevated temperature (van de Vossenberg et al. 1998). Hyperthermophiles have evolved mechanisms for the maintenance of membrane integrity at high temperatures, assuring an appropriate matrix for proteins and thus allowing the generation of concentration gradients across the membrane. Such mechanisms include alterations in the length and branching of the acyl chains, saturation, cyclization, or the arrangement of phospholipids across the membrane (Daniel and Cowan 2000, Ulrih et al. 2009).

It is important to bear in mind that many of these distinctive characteristics are not restricted to hyperthermophiles nor are ubiquitous among them. Usually membranes of prokaryotes contain phospholipids with a core structure of a glycerol moiety with attached fatty acyl chains. Hydrocarbon chains in archaeal lipids are ether-linked to the glycerol moiety in contrast to bacterial and eukaryal lipids that are ester-linked. The ether-linkage is more resistant to hydrolysis and oxidation, hence more stable, and this advantage may help these organisms to withstand extreme environmental conditions (Daniel and Cowan 2000, Driessen and Albers 2007). Archaeal

hydrocarbon chains are highly methyl-branched while its bacterial counterparts are mainly composed by straight chain fatty acids. Most of the archaeal isoprenoid lipid acyl chains are fully saturated, and so they are more resistant to hydrolysis and oxidation. Membrane spanning (monolayer) tetraether lipids (C_{40} isoprenoid acyl chains) are usually found in extreme thermophiles and in acidophiles. In contrast, diether lipids (C_{20} isoprenoid acyl chains) are present in moderate archaea and halobacteria, forming bilayers similar to those found in the bacterial ester-linked lipids (Koga and Morii 2007). The addition of cyclic structures to the transmenbrane portions results in enhanced stability and reduced membrane fluidity (Daniel and Cowan 2000).

Interestingly, lipids in hyperthermophilic bacteria appear to mimic archaeal lipids. Although some exceptions can be found within the domain *Bacteria*, the presence of membrane-spanning lipids and mixed ether/ester lipids in hyperthermophilic bacteria of the genera *Thermotoga* and *Aquifex*, corroborates the idea that these features may be associated with life at high temperature (Damsté et al. 2007).

ii) Nucleic acids

At high temperatures nucleic acids are more susceptible to chemical modification/degradation and denaturation. The thermal resistance of nucleic acids has been studied for the past 30 years and some main features that contribute to DNA and RNA stability have been pointed out (Grosjean and Oshima 2007). The DNA molecule, as long as it remains covalently closed, is quite resistant to high temperatures. Circular plasmids of double stranded DNA have been found at temperatures as high as 107ºC (Marguet and Forterre 1994). It is known that due to additional hydrogen bonds, the G-C

pair is more stable and thus, more heat resistant than the A-T pair. However, a correlation between high G+C content in DNA and hyperthermophily is not observed. Some mesophiles have genomes with higher G+C content than hyperthermophiles and even amongst hyperthermophiles there is a huge variation of genomic G+C contents. For example, *Pyrolobus fumarii* (T_{opt} 106ºC), *Pyrococcus furiosus* (T_{opt} 100ºC) and *Aquifex aeolicus* (T_{opt} 85ºC) have genomic G+C contents of 53, 41, and 44%, respectively. However, the same is not true for RNA. In fact, studies performed in the nineties revealed that the secondary structure of tRNA and rRNA are stabilized by an increased number of G+C pairs near the stem areas and that the increase in G+C content correlates positively with the optimal growth temperature of the organism (Galtier and Lorby 1997). These studies have been confirmed by recent comparative genomic analysis (Dutta and Chaudhuri 2010). Additionally, RNA molecules in hyperthermophiles are stabilized by post-translational covalent modification of nucleosides. These modifications involve increased methylation of the ribose, thiolation and acetylation or replacement of N by C atoms in adenine rings (Kowalak et al 1994, Dutta and Chaudhuri 2010). Salts protect DNA and RNA from thermal degradation: *in vitro* studies have proved that Na^+, K^+ and Mg^{2+} protect DNA from chemical degradation by shielding the negatively charged phosphate backbone of the nucleic acids molecules, hence protecting the phosphodiester bond and aiding the correct folding of these molecules (Hensel and Konig 1988, Marguet and Forterre 2001, Serebov et al 2001). The same effect could be justified by the presence of high intracellular concentrations of salts (Molar range) or the presence of intracellular polycationic polyamines in some hyperthermophiles (Terui et al. 2005, Grosjean and Oshima 2007).

The thermal resistance of the DNA molecule is also enhanced by association with cationic proteins and supercoiling. Numerous histone-like

proteins, that compact DNA and have *in vitro* stabilizing effects, have been identified in hyperhermophiles (Daniel and Cowan 2000). Reverse gyrase is a unique DNA topoisomerase responsible for generating positive supercoils in dsDNA. Reverse gyrase is present in all hyperthermophiles and may be an important hallmark for thermoadaptation (Lopes-Garcia and Forterre 2000, Grosjean and Oshima 2007). However, the *in vivo* role of reverse gyrase is not completely understood. A *Thermococcus kodakarensis* mutant with deletion on the reverse gyrase gene revealed that growth was only compromised at extreme temperatures. These results and other recent data suggest that positive supercoiling is not essential for hyperthermophily and that the role of this enzyme may be associated with protection of the DNA molecule against damage such as depurination or DNA breakage (Atomi et al. 2004, Brochier-Armanet and Forterre 2007, Heine and Chandra 2009).

iii) Proteins

Obviously, temperature also affects protein structure and function. Proteins from hyperthermophiles not only are conformationally more stable at high temperatures, but are also more resistant to other extreme conditions like pH, salt, pressure, organic solvents, or detergents. Although these are general trends, it is important to note that not all proteins from hyperthermophiles are hyperstable. Nevertheless, structural data and comparative genomic studies of hyperthermophilic proteins allowed the identification of some molecular determinants important for protein stability in hyperthermophiles (Luke et al. 2007, Robb and Newby 2007). Intrinsic factors like amino acid composition, hydrophobic interactions, ions pairs and ion networks, packing and reduction of solvent exposed surfaces, fewer and shorter surface loops, or oligomerization, have been pointed out (Unsworth et al. 2007). Comparative

studies of meso-, thermo-, and hyperthermophilic proteins show that with increasing temperatures there is a general tendency for an increase in the amount of charged amino acids (Glu, Asp, Arg and Lys), and a decreased frequency of uncharged polar amino acids (Ser, Gln, Tyr and Asn). Furthermore, recent work revealed a positive correlation between the number of intra-subunit ionic interactions and the growth temperature. It is clear that charged residues can promote ion-pair formation, but the exact role that ion-pairs have in the thermal stability of proteins is still unclear (Elcock 1998, Das and Gerstein 2000). In general, stable proteins have more compact structures with fewer internal spaces, a consequence of smaller loops, but also form oligomeric assemblies and inter-subunit interactions (Arnott et al. 2000). Extrinsic factors such as molecular chaperones, intracellular salt concentration, compatible solutes, metabolites, or cofactors may also contribute to protein stability (Ladenstein and Antranikian 1998).

The key for success at high temperatures is a correct balance between flexibility and stability. Proteins are dynamic structures with conformational flexibility, and it is important that at high temperatures correct protein flexibility is achieved to allow for enzyme activity. This concept is of extreme importance, for example in industrial applications, since it is crucial the existence of a balance between protein stability and functionality for optimal enzyme activity.

iv) Metabolites

A more important problem for cellular activity at high temperatures may be the weak stability of protein substrates and cofactors. It has been suggested that the temperature limit for life is dictated by small molecules, like coenzymes or metabolites like ATP, NAD/P(H), acetyl-CoA or acetyl-phosphate

(Stetter 1999, Cowan 2004). Again, Nature appears to have found alternative ways to bypass the problem of thermal instability of metabolites and coenzymes. Such mechanisms include micro-environmental compartmentalization or metabolic channeling, rapid turnover or increased catalytic efficiency, utilization of alternative more stable metabolites, or alternative biochemical pathways (Daniel and Cowan 2000, Massant 2007). Interestingly, the clustering of functionally related genes into operons, a feature exclusive to bacteria and archaea, facilitates the formation of multienzyme complexes and mutual stabilization of proteins, as well as the channeling of thermolabile metabolites (Glasdorff 1999).

The replacement of the thermolabile NAD^+/P for more heat stable non-haem iron proteins in redox reactions has been observed in several members of the domain *Archaea*. The same observation was made in hyperthermophilic bacteria of the genus *Thermotoga*, suggesting that although it may be a characteristic of primitive organisms there is also a correlation with hyperthermophily (Daniel and Cowan 2000). Furthermore, hyperthermophilic archaeal kinases show some preference for the utilization of more stable metabolites like pyrophosphate or ADP instead of ATP, as is the case of *Pyrococcus furiosus* hexokinase and phosphofructokinase, which are exclusively ADP-dependent (Daniel and Cowan 2000).

Functional genomics studies, arising from the whole-genome sequencing era, have provided global methods (microarrays, proteomics, metabolomics or mutational analysis) to investigate the molecular basis of heat adaptation. High-throughput studies are presently being carried out for extremophiles such as *Thermotoga maritima*, *Sulfolobus solfataricus*, *Pyrococcus abyssi* or *Pyrococcus furiosus* (Conners et al. 2006, Trauger et al. 2008, Lee et al. 2009, Groisillier et al. 2010, Zaparty et al. 2010).

The heat-shock response

Cellular stress is caused by a sudden change in the surrounding environment of the cell. The type of stress may be physical, like temperature variation, or chemical, such as changes in pH, salinity or oxygen concentration. In all cases, the stress response includes protein denaturation, down-regulation of housekeeping genes and activation of heat-shock genes. The stress proteins encoded by the heat-shock genes are multifunctional, ubiquitous, and serve functions as molecular chaperones, proteases, rRNA processing and signal transduction (Macario et al. 1999).

The heat-shock response has been first described in *Drosophila melanogaster* (Ritossa 1962) and then in *Escherichia coli* (Yamamori and Yura 1980), but this is a ubiquitous protective strategy (Schumann 2007). The beginning of the whole process involves thermosensors that will recognize the temperature shift, and trigger a stress response. This response is usually transitory and involves the expression of specific genes that will counteract the external physical effects allowing the cell to reach a new steady state.

Sensing mechanisms involve alterations in membrane fluidity, changes in DNA curvature or supercoiling, and/or in mRNA and protein conformation (Lopez-Garcia and Forterre 2000, Hurme and Rhen 1998, Klinkert and Narberhaus 2009). Heat-shock leads to a transient increase in DNA positive supercoiling; in mesophiles like *Escherichia coli*, *Salmonella* spp. or *Bacillus subtilis* this process is mediated by gyrase and topoisomerase I and recovery to negative supercoiling is done by a concerted action between gyrase, a nucleotide binding protein HU, and a molecular chaperone DnaK (Klinkert and Narberhaus 2009). DNA topology acts as a sensor under heat-shock conditions and is an important factor for regulation of gene expression, especially because it controls the access to RNA polymerase (Lopes-Garcia

and Forterre 2000, Klinkert and Narberhaus 2009). At low temperature the Shine-Dalgarno sequence and the start codon are trapped in a hairpin structure within the mRNA molecules. This prevents binding of the ribosome and thus translation initiation. Upon an increase in the temperature the secondary structure is destabilized and the ribosome binding site becomes accessible (Schumann 2007, Klinkert and Narberhaus 2009).

Protein thermosensors include transcription regulators, chemosensory proteins, chaperones or proteases. Regulators of transcription include the repressor proteins TlpA, RheA, and Phr found in *Salmonella enterica*, *Streptomyces albus*, and *Pyrococcus furiosus*, respectively (Servant et al. 2000, Hurme et al. 1997, Vierke et al. 2003). Phr was the first transcription factor of *Archaea* to be discovered and it acts as a repressor of heat-shock genes at physiological temperature, but upon heat-shock (107ºC) it is released from the promoters (Liu et al. 2007, Klinkert and Narberhaus 2009). Molecular chaperones and proteases are part of more complex thermosensor systems and their transcription is induced under heat-shock conditions. These proteins participate in protein folding, assembly, transport and repair under stress conditions (Schumann 2007, Klinkert and Narberhaus 2009). Not all heat-shock proteins are molecular chaperones and likewise, not all molecular chaperones are heat-shock proteins. In fact, molecular chaperones are present and functional in the cell under normal conditions (Lopez-Garcia and Forterre 2000).

The heat-shock response has been extensively studied in *Escherichia coli* (Guisbert et al. 2008). The master regulon σ^{32}, encoded by the *rpoH* gene, is composed of more than 30 heat-shock proteins. Many heat-shock proteins are molecular chaperones (like GroE or DnaK) or ATP-dependent proteases (like ClpX or ClpP) that assist protein folding, repair and degradation. The induction of heat-shock proteins, observed after a sudden

increase in the temperature, occurs due to an increase in the levels of the σ^{32} regulon, which directs RNA polymerase to transcribe from the heat-shock promoters. At optimal physiological conditions (*e.g.,* at 30ºC), σ^{32} is relatively unstable and the intracellular concentration is kept very low, but upon temperature up-shift (*e.g.,* from 30ºC to 42ºC) the intracellular concentration of σ^{32} transiently increases (nearly 20-fold in 5 min). During this time the transcription of some heat-shock genes is activated. The increase in σ^{32} levels is transient and rapidly starts to decline reaching a steady-state level approximately 2 or 3-fold higher than its original value. After a temperature downshift (*e.g.* from 42ºC to 30ºC) the transcription of heat-shock genes decreases as a result of the decrease in the activity of σ^{32}, mediated by the heat-shock proteins. Hence, transcription of heat-shock genes can be regulated by controlling the synthesis, degradation or activity of σ^{32} (Yura et al. 2000, Guisbert et al. 2008). While σ^{32} offers protection against cytoplasmatic stress, another σ factor, σ^{E}, was found to be involved in the extracytoplasmatic stress response of *Escherichia coli*. This second regulon provides cytoplasmatic and periplasmatic heat-shock proteins that are essential for cell survival at very high temperatures (*e.g.*, from 45ºC to 50ºC) (Yura et al. 2000).

Little is known about the heat-shock response of hyperthermophilic organisms. So far, the heat-shock response was investigated in the archaea *Pyrococcus furiosus*, *Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, *Methanocaldococcus jannaschii* and in the bacterium *Thermotoga maritima*, (Shockley et al 2003, Pysz et al. 2004, Rohlin et al. 2005, Boonyaratanakornkit et al. 2005, 2007, Tachdjian and Kelly 2006). Hyperthermophiles also have heat-shock proteins, most of them identified by acrylamide gel electrophoresis after stress induction or by sequence homology to other known heat-shock

proteins. Some aspects of the archaeal heat-shock response are similar to their bacterial counterparts, but there are some distinctive features namely the heat-shock protein families. TF55, also known as thermosome, is a type II Hsp60 chaperonin more related to the eukaryal TCP-1 chaperonin family than to the bacterial GroES/GroEL system (type I Hsp60 chaperonins). This protein is unique and ubiquitous among archaea and is predominant among hyperthermophiles. It is the major heat-shock protein among hyperthermophilic archaea and its localization near the membrane suggests that it plays a role in the maintenance of the membrane integrity under heat-stress. Nevertheless, TF55 is also abundant in the cell under normal growth conditions (Tachdjian et al. 2008, Lopez-Garcia and Forterre 2000). Archaea have another eukarya-related molecular chaperone, named prefoldin, which interacts with Hsp60 and is absent in bacteria. On the contrary, the common molecular chaperones Hsp40 (DnaJ), Hsp70 (DnaK), and Hsp23 (GrpE) are absent from hyperthermophilic archaea, but not from hyperthermophilic bacteria (Holden et al 1999, Pysz et al 2004, Tachdjian et al. 2008). Small heat-shock proteins that act as chaperones to prevent aggregation of denatured proteins, and proteolytic systems, namely a system similar to the Lon protease, are present in the three domains of Life, including hyperthermophiles (determined by sequence homology) (Shockley et al. 2003, Lopez-Garcia and Forterre 2000). Another distinctive feature of hyperthermophilic archaea is the absence of proteases of the Clp, HlsV and FtsH families, and the presence of large proteolytic systems, the proteosome. Other factors like biofilm formation or the presence of high amounts of small organic solutes that accumulate under stress conditions may be seen as an active part of the hyperthermophilic heat-shock response (Holden et al 1999, Tachdjian et al. 2008).

Generally, the transcriptional heat-shock response of hyperyhermophilic archaea has some common trends: in *Pyrococcus furiosus*, *Methanococcus janaschii*, and *Archaeoglobus fulgidus* the thermosome, small heat-shock proteins, and the proteosome are up-regulated. However, small differences can be found and other genes are differentially regulated: while in *Methanococcus janaschii* genes encoding an α-subunit of the prefoldin is highly up-regulated, in *Archaeoglobus fulgidus* all the prefoldin related genes are down-regulated upon heat-shock (Shockley et al 2003, Rohlin et al. 2005, Boonyaratanakornkit et al. 2005). The heat-shock response of *Sulfolobus solfataricus* showed that genes encoding small heat-shock proteins were up-regulated, but genes encoding the thermosome and proteases were not affected or were down-regulated by the temperature increase. Interestingly, several DNA-binding proteins were highly up-regulated as well as many insertion sequence elements (Tachdjian and Kelly 2006). The heat-shock response of the hyperthermophilic bacteria *Thermotoga maritima* was also investigated: it shares some common elements with mesophilic bacteria like the up-regulation of chaperones such as DnaK, but also significant differences were noticed, specially the down-regulation of ATP-dependent proteases (Pysz et al. 2004).

Although the action/presence of intrinsic factors that contribute to cellular stability of hyperthermophiles has been known for a long time, the mechanisms involved in the heat-shock response of these organisms only began to be investigated in 2003 in *Pyrococcus furiosus* (Shockley et al. 2003). Hence the relative contribution of chaperones, proteases, organic solutes or other factors involved in the heat-shock response of hyperthermophiles still needs to be determined.

Compatible solutes

The content of salt or sugar in the environment influences the amount of water available to microbes and, the availability of water affects the survival and growth of microorganisms. Most organisms were isolated from aqueous environments where salts and sugars are relatively diluted, but some can be found in extremely saturated brines or concentrated sugar solutions. In either case, microorganisms are able to adjust to osmotic variations that occur in the surroundings, by a process called osmoadaptation. Alterations in the external osmolarity of the medium, usually due to changes in the external Na^+ concentration, rapidly triggers a cascade of responses, including water movement across the cellular membrane, that could result in cell swelling and bursting in hypotonic environments, or cell dehydratation in hypertonic environments.

Many hyperthermophiles originate from marine habitats where the salt concentration is similar to that of seawater. Many of these organisms require low levels of NaCl for survival and growth, but like any other microorganism living in a saline environment they are challenged with alterations in the salinity of the medium and must be able to adjust to such changes. There are two main strategies for osmoadaptation in archaea and bacteria to face increasing osmolarities: the "salt-in-the-cytoplasm" strategy, which involves the influx of inorganic ions like Na^+, K^+ or Cl^-, into the cytoplasm; or the accumulation of low-molecular mass organic osmolytes. The "salt-in-the-cytoplasm" strategy is commonly found in extreme halophilic archaea and in halotolerant bacteria whose macromolecules and entire metabolic pathways are perfectly adapted to high intracellular salt concentration. In fact, these organisms evolved in a way that most of their enzymes are dependent on those ions for activity (da Costa et al. 1998, Roeβler and Müller 2001,

Empadinhas and da Costa 2008). However, the majority of prokaryotes accumulate preferentially small organic osmolytes, *i.e.*, compatible solutes, to cope with hypertonic environments.

The notion of "compatible solute" was put forward in the early seventies by Brown (Brown and Simpson 1972, Brown 1976) and referred to organic or inorganic compounds that could accumulate to high concentrations and did not interfere with normal cell function. The accumulation of compatible solutes does not require modifications of the genetic or enzymatic machinery and is an efficient way to adjust to osmotic fluctuations. Perhaps due to the specific requirements implied in the definition of compatible solutes, only a limited number of compounds meet these criteria and apparently they were widely adopted as osmoprotectants. In fact, this strategy is widespread in Nature and is present in the three domains of Life (da Costa et al. 1998, Roeβler and Müller 2001).

Diversity and distribution of compatible solutes

With the discovery of new organisms, specially extremophiles, the collection of compatible solutes has been increasing; some compatible solutes like trehalose, proline, glycine betaine, and ectoine are widespread in the microbial world, while others, such as di-*myo*-inositol phosphate or mannosylglycerate, appear to be restricted to heat adapted organisms. Compatible solutes fall into three main categories: amino acids and derivatives, sugars, and polyols. In general, compatible solutes are polar and highly soluble molecules that do not have a net charge at physiological pH, however, many exceptions can be found within hyperthermophiles, as will be discussed below. Figure 1.2 shows the distribution of common solutes within the three domains of Life.

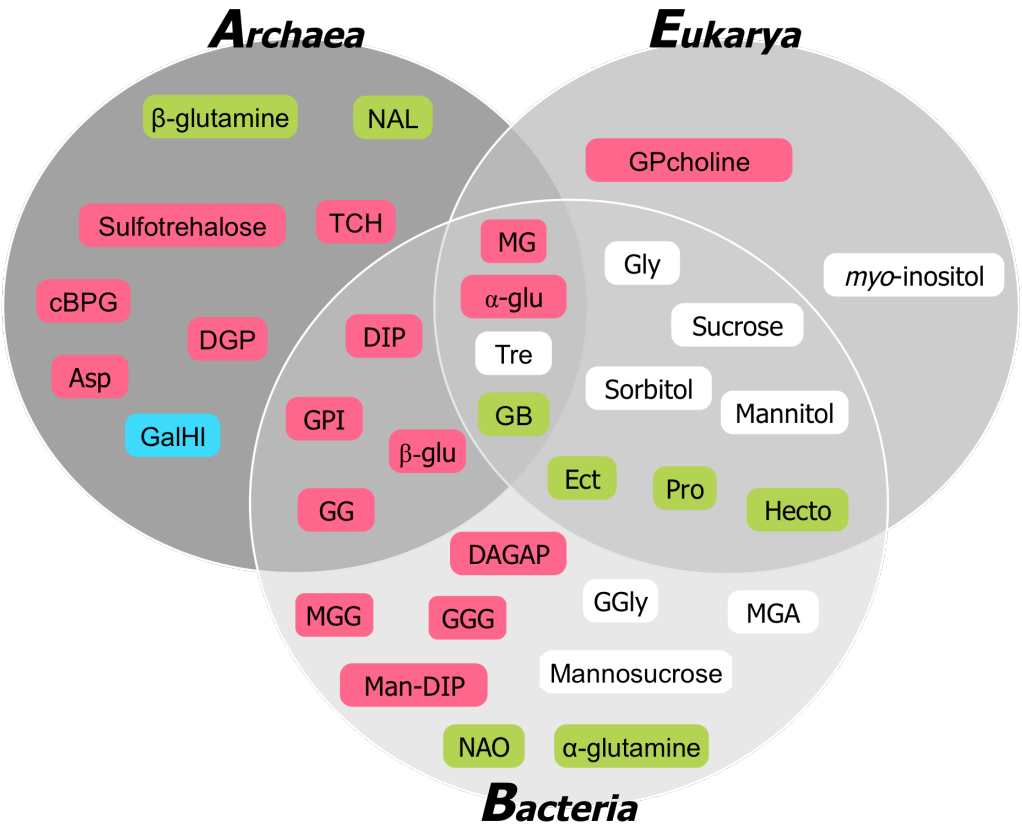


Figure 1.2. Distribution of organic solutes in the three domains of Life. Abbreviations: cBPG, cyclic 2,3-bisphosphoglycerate; Tre, trehalose; MG, mannosylglycerate; MGA, mannosylglyceramide; DIP, di-*myo*-inositol phosphate; DGP, diglycerol phosphate; GPI, glycero-phospho-*myo*-inositol; Man-DIP, mannosylated-DIP; TCH, 1,3,4,6-tetracarboxyhexane; α-Glu, α-glutamate; β-Glu, β-glutamate; Asp, aspartate; Ect, ectoine; Hect, hydroxyectoine; GalHl, β-galactopyranosyl-5-hydroxylysine; GB, glycine betaine; Pro, proline; NAL, N-acetyl-β-lysine; NAO, N-acetyl-ornitine; GG, glucosylglycerate; GGG, glucosyl-glucosylglycerate; Gly, glycerol; GGly, glucosylglycerol; MGG, mannosyl-glucosylglycerate; DAGAP, di-N-acetyl-glucosamine phosphate; GPcholine, glycerophosphocholine. Pink shadow, negatively charged solutes; blue shadow, positively charged solutes; white shadow, neutral solutes; green shadow, zwitterionic solutes (da Costa et al. 1998, Roberts 2004, Empadinhas and da Costa 2008, Gallazzini and Burg 2009, and Santos et al. 2010).

Amino acids and derivatives

Glutamate, proline, and glutamine are amino acids that function as compatible solutes in many mesophilic bacteria (da Costa et al. 1998) (Fig. 1.3). Both α- and β-amino acids are used for osmoadaptation. β-amino acids add the advantage of being highly soluble, to the fact that they are not metabolized by the cell, in contrast to their α-counterparts, making them almost ideal compatible solutes (Roeβler and Müller 2001). The accumulation of α-glutamate generally occurs with the purpose of neutralizing the rapid influx of K^+ ions at the initial stages of low-level osmotic adjustement in many bacteria and archaea (Santos and da Costa 2001). β-glutamate on the other hand, is a rare compatible solute that protects a few bacteria and many methanogens from salt stress (Robertson et. al. 1990, Santos et al. 2001).

Proline is an important osmolyte that can accumulate to high concentrations, reaching almost 20% of the cell dry weight (da Costa et al. 1998). Proline is usually accumulated in halophilic/halotolerant bacteria, namely of the genus *Bacillus*. Proline is usually scavenged from the medium, but can also be synthesized *de novo* in an osmotically controlled manner (Bremer and Kramer 2000).

Usually, small amounts of α-glutamine accumulate for low level osmotic adjustement in Gram-positive bacteria (da Costa et. al. 1998); β-glutamine is far more soluble than α-glutamine, hence it can accumulate in the cell to higher levels acting as an efficient compatible solute, but apparently is restricted to halophilic methanogens (da Costa et al. 1998, Lai et al. 1991, Robertson and Roberts 1992, Roeβler and Müller 2001).

Glycine betaine is probably the most commonly used compatible solute, occurring in plants, mammals, algae, bacteria and archaea, where it is used to cope with salt stress. While the majority of microorganisms can

efficiently internalize glycine betaine from the medium, very few can synthesize it either by oxidation of choline or methylation of glycine (Bremer and Kramer 2000, Roeβler and Müller 2001).

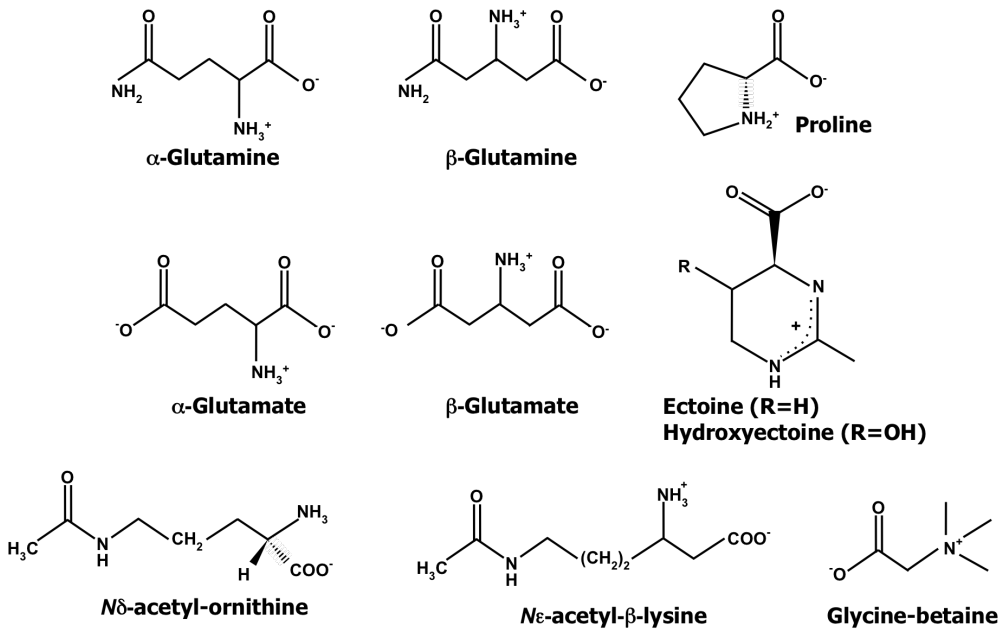


Figure 1.3. Structures of amino acids and derivatives used as compatible solutes.

The N-acetylation of the amino acids ornitine and β-lysine originates N-acetyl-ornitine and N-acetyl-β-lysine, respectively. The conversion of two positively charged amino acids into two neutral compounds is an advantageous strategy adopted by the cell allowing the accumulation of osmolytes that are more compatible with functionality in cells not adapted to high accumulation of ionic compounds. N-acetyl-ornitine is found in almost all *Bacillus* species while N-acetyl-β-lysine appears to be exclusive to methanogenic archaea where it accumulates only under salt stress (Wohlfarth et al. 1993, Lai et al. 1991, Sowers et al. 1990, Empadinhas and da Costa

2008). Ectoine and hydroxyectoine can be classified as cyclic forms of N-acetylated amino acids. Ectoine is widespread in heterotrophic halotolerant bacteria and it has not been found in archaea. Some halophilic bacteria have the ability to modify ectoine by hydroxylation (Ventosa et al. 1998). In addition to glycine betaine, ectoine is one of the most used compatible solutes in the microbial world (da Costa et al. 1998, Bremer and Kramer 2000). Bacteria unable to produce ectoine, like *Escherichia coli*, *Bacillus subtilis* or *Corynebacterium glutamicum*, have transport systems to sequester this compound from the medium for osmoprotection (Bremer and Kramer 2000).

Sugars and derivatives

Trehalose is a naturally occurring glucose disaccharide widespread in Nature, that can serve several functions including signaling, cell wall component, energy or carbon reserve, or protection against a variety of stresses (oxidation, dehydratation, heat, cold, desiccation) (Elbein et al 2003). In the domain *Eukarya*, trehalose has been found to protect plants, yeast and fungal spores from dehydration and proteins from extreme heat. In the domain *Bacteria* trehalose is found in many species ranging from the model organism *Escherichia coli*, to the soil bacteria *Rhizobium* spp., and to the thermophile *Thermus thermophilus* (Larsen et al. 1987, Streeter 1985, Nunes et al 1995, Elbein et al 2003). However, trehalose is also found in members of the domain *Archaea*, namely in the hyperthermophiles *Pyrococcus horikoshii* and *Thermococcus litoralis*, where it accumulates in response to osmotic stress (Empadinhas et al. 2001, Lamosa et al. 1998).

Other carbohydrates used as compatible solutes include sucrose, sulfotrehalose, glucosylglycerate, mannosylglycerate, or the unusual mannosucrose (da Costa et al. 1998) (Fig. 1. 4). Sucrose is mostly found in

non-halophilic cyanobacteria while mannosucrose was detected in *Agrobacterium tumefaciens* biotype I (Reed et al. 1984, Smith et al. 1990). Mannosylglycerate was originally identified in the red algae of the order *Ceramiales*, but today its presence is acknowledged in other red algae and in (hyper)thermophilic bacteria and archaea (Bouveng et al. 1955, Nunes et al. 1995, Martins and Santos 1995; Lamosa et al. 1998, Empadinhas et al. 2001, Gonçalves et al. 2003). In the red algae, mannosylglycerate does not behave as a compatible solute in response to osmotic stress, a role that is assumed by mannitol (Karsten et al. 1994).

Glucosylglycerate appears to be a rare compatible solute with limited distribution among mesophiles, originally found in the cyanobacterium *Agmenellum quadruplicatum* when cultivated in nitrogen limiting conditions, and later in the γ-proteobacterium *Erwinia chryssanthemi* grown in a combination of salt stress and nitrogen limitation (Goude et al. 2004, Kollman et al. 1979).

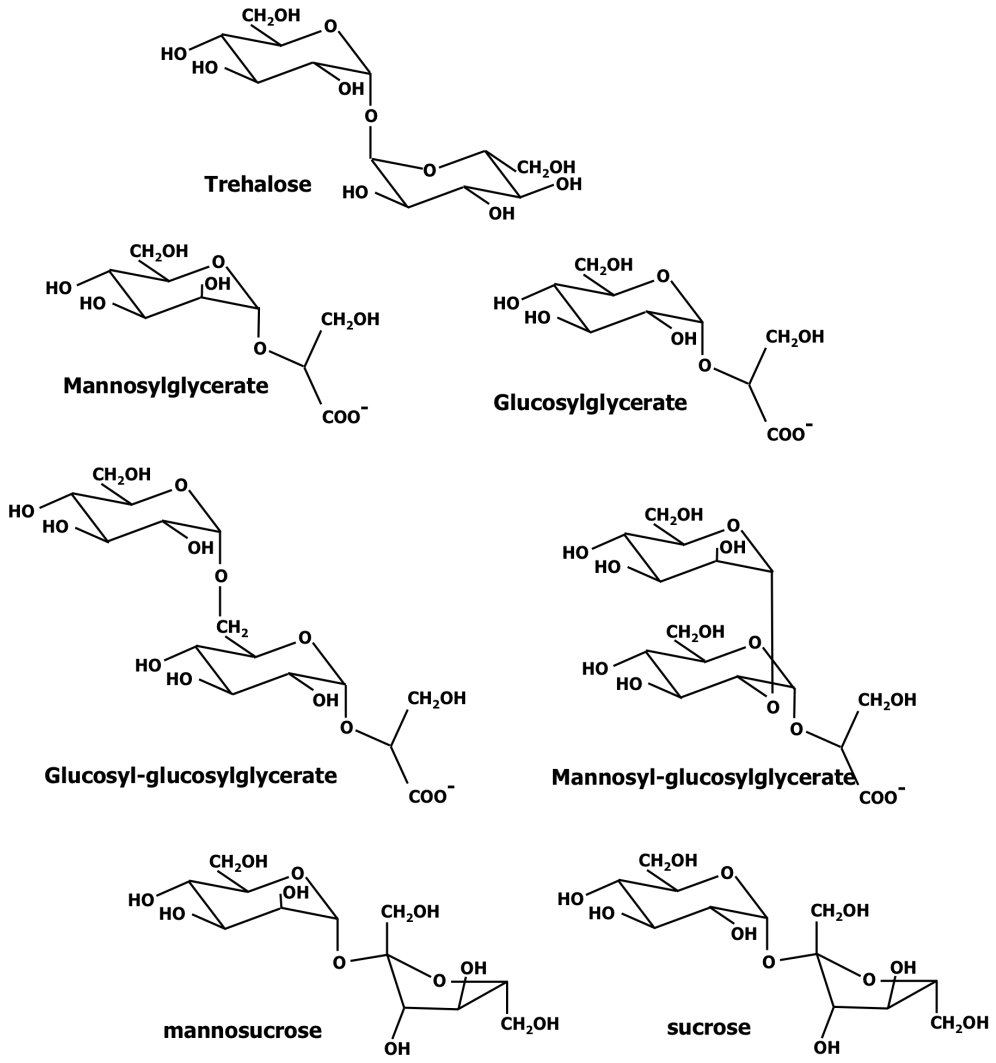


Figure 1.4. Structures of sugar-derived compatible solutes.

Polyols and derivatives

Sorbitol, *myo*-inositol, glycerol, arabitol, or mannitol, are compatible solutes widespread within the domain *Eukarya*, typically found in yeast, algae, plants or mammals (Fig. 1.5). The distribution of polyols in the domain *Bacteria* is

very restricted; the few reports include the accumulation of sorbitol by *Zygomonas mobilis* during growth on high levels of sucrose, and the accumulation of mannitol by *Pseudomonas putida*, in response to osmotic stress (Kets et al. 1996, da Costa et al. 1998). Glucosylglycerol is a rare compatible solute, used by moderately salt tolerant cyanobacteria, and a few other bacteria under osmotic stress (da Costa et al. 1998, Klahn and Hagemann 2011). Glycerophosphocholine acts as an osmoprotectant in yeast cells, and also protects renal medullary mammalian cells from high, physiological concentrations of NaCl and urea (Kiewietdejonge et al. 2006, Gallazzini and Burg 2009).

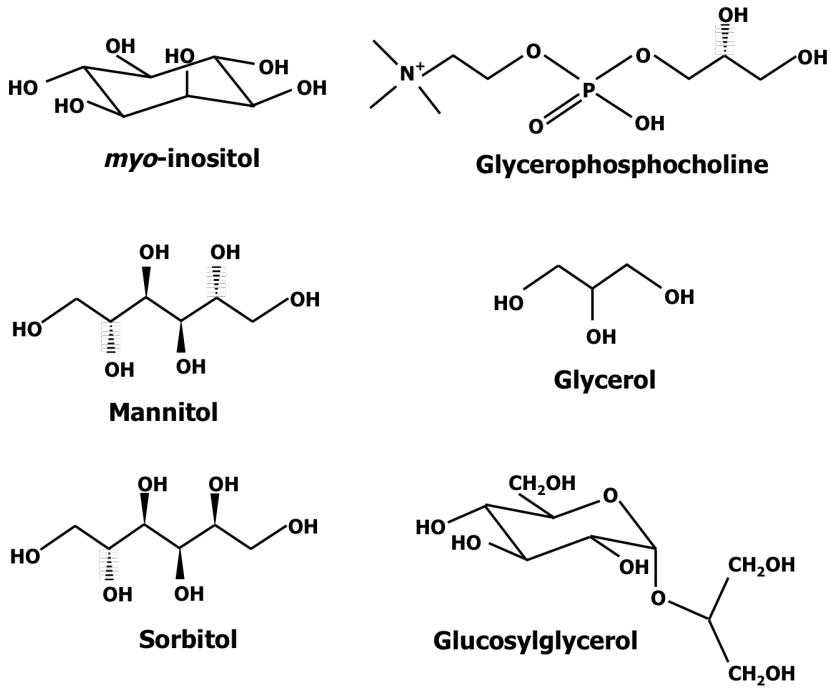


Figure 1.5. Structures of polyols and derivatives used as compatible solutes.

Compatible solutes of hyperthermophiles

In general, compatible solutes found in hyperthermophilic bacteria and archaea are negatively charged in contrast to what occurs in mesophiles, where neutral or positively charged solutes are predominant. The negative charge results from the addition of groups like phosphate, sulfate or carboxylate to carbohydrates or polyols. These solutes include sulfotrehalose, glucosylglycerate, mannosylglycerate, diglycerol phosphate or di-*myo*-inositol phosphate. Negatively charged solutes could serve to balance high intracellular K^+ or counteract osmotic pressure (Roberts 2005). In many organisms the accumulation of compatible solutes increases not only in response to an osmotic stress but also in response to an increase in the growth temperature. Furthermore, some of these compounds appear to be restricted to this type of organisms. These observations led to the view that compatible solutes play a role in thermoadpatation, hence the interest in the discovery of new compounds, elucidation of their biosynthesis and regulation.

Chemically, solutes encountered in hyperthermophiles belong to one of the following categories: hexose derivatives with the hydroxyl group of carbon 1 usually blocked in an α-configuration, or polyol phosphodiesters. Table 1.2 summarizes the distribution of compatible solutes commonly found in hyperthermophiles.

Hexose derivatives

α-Mannosylglycerate is currently one of the most widespread compatible solutes within thermophiles and hyperthermophiles. It has been found in the thermophilic bacteria *Rhodothermus marinus*, *Thermus thermophilus*, and *Rubrobacter xylanophilus* (Nunes et al. 1995, Empadinhas et al. 2007). Within

the domain *Archaea*, mannosylglycerate was found in members of the genera *Pyrococcus*, *Thermococcus*, *Palaeococcus*, *Aeropyrum*, *Methanothermus*, *Stetteria*, and *Archaeoglobus* (Martins et al. 1997, Martins and Santos 1995, Empadinhas et al. 2001, Neves et al. 2005, Lamosa et al. 1998, Gonçalves et al. 2003). Usually this solute accumulates in response to osmotic stress except in *Rubrobacter xylanophilus*, which constitutively accumulates mannosylglycerate, and in *Rhodothermus marinus* and *Palaeococcus ferrophilus* where its level increases also at supra-optimal temperatures. (Santos et al. 2010). The uncharged derivative of mannosylglycerate, mannosylglyceramide, is a rare compound encountered in *Rhodothermus marinus* only (Nunes et al. 1995).

The identification of glucosylglycerate, a compound distributed among halotolerant mesophilic bacteria and in the thermophilic bacterium *Persephonella marina*, further extended the scope of action of this solute. On the other hand, mannosylglucosylglycerate and glucosylglucosylglycerate are two rare sugar derivatives that have only been identified in *Petrotoga* spp. and *Persephonella marina*, respectively (Santos et al. 2010).

Table 1.2. Distribution of organic solutes in thermophiles and hyperthermophiles (adapted from Santos et al. 2010)

Family	Organisms	T_{Opt} (ºC)	Solutes							
			cBPG	Tre	MG	DIP	α-Glu	β-Glu	Asp	Other
Archaea										
Pyrodictiaceae	*Pyrolobus fumarii*	106				+				
	Pyrodictium occultum	105				+	+			
	Hyperthermus butylicus	99				+				
Thermoproteaceae	*Pyrobaculum aerophilum*	100		+						
	Pyrobaculum islandicum	100								
	Thermoproteus tenax	88		+						
Thermococcaceae	*Pyrococcus furiosus*	100			↑(S)	↑(T)	+			
	Pyrococcus horikoshii	98		+	↑(S)	+	+			
	Thermococcus stetteri	87			↑(S)	↑(T, S)	+		↑(S)	
	Thermococcus celer	87			↑(S)	↑(T)	+		↑(T)	
	Thermococcus litoralis	85		↑(S)	↑(S)	↑(T)	+		↑(S)	GalHl
	Thermococcus kodakarensis	85				+	+		+	
	Palaeococcus ferrophilus	83			↑(T)↑(S)		↑(T)		↑(S)	
	Thermococcus zilligii	75								
Methanopyraceae	*Methanopyrus kandleri*	98	+				+			
Desulfurococcaceae	*Stetteria hydrogenophila*	95		+	+	+				
	Aeropyrum pernix	90			+	+				
Methanocaldococcaceae	*Methanotorris igneus*	88				↑(T)	↑(S)	↑(S)		
	Methanocaldococcus jannaschii	85						↑(S)		
Archaeoglobaceae	*Archaeoglobus fulgidus* VC-16	83				↑(T)	+			DGP ↑(S) GPI ↑(T, S)
	Archaeoglobus profundus	83			+	+	+			
	Archaeoglobus veneficus	75			+	+	+			DGP
Methanothermaceae	*Methanothermus fervidus*	83	+				+			
Sulfolobaceae	*Acidianus ambivalens*	80		+						
	Sulfolobus solfataricus	75		+						
	Metallosphaera sedula	75		+						
Methanobacteriaceae	*Methanothermobacter thermoautotrophicus*	70	+				+			TCH

Table 1.2. Continuation.

Family	Organisms	$T_{opt.}$ (°C)	Solutes							
			cBPG	Tre	MG	DIP	α-Glu	β-Glu	Asp	Other
Archaea										
	Methanothermobacter marburgensis	65	+				+			TCH
Methanococcaceae	*Methanothermococcus okinawensis*	70					+		+	
	Methanothermococcus thermolithotrophicus	65					+	+	+	NAL
Thermoplasmataceae	*Thermoplasma acidophilum*	60	+							
Bacteria										
Aquificaceae	*Aquifex pyrophilus*	85				↑(T)		↑(S)		GPI ↑(T, S) Man-DIP ↑(S)
Thermotogaceae	*Thermotoga maritima*	80				↑(S)		↑(S)		Man-DIP ↑(T)
	Thermotoga neapolitana	80				↑(S)	+	↑(S)		Man-DIP
	Thermosipho africanus	75					+			Pro
	Thermotoga thermarum	70								
	Marinitoga piezophila	70					+			Pro
	Fervidobacterium islandicum	70								
	Petrotoga mobilis	60					↑(S)	↑(S)		MGG ↑(S)(T), GB
	Petrotoga miotherma	55					+			MGG, Pro
Hydrogenothermaceae	*Persephonella marina*	70				+		+		GG,GGG
Rhodothermaceae	*Rhodothermus marinus*	65		+	↑(S, T) ↑(T)		+			MGA ↑(S)
Rubrobacteraceae	*Rubrobacter xylanophilus*	60		+	+	↑(T)	+			GB, DAGAP
Thermacaea	*Thermus thermophilus*	70		↑(S)	↑(S)		+			GB

+, indicates the presence of the solute in cases for which the response to environmental conditions has not been reported. (S) and (T) indicate that the intracellular level of the solute increases in response to osmotic or heat stress, respectively. cBPG, cyclic 2,3-bisphosphoglycerate; Tre, trehalose; MG, mannosylglycerate; MGA, mannosylglyceramide; DIP, di-*myo*-inositol phosphate; Man-DIP,mannosylated-DIP; DGP, diglycerol phosphate; GPI, glycero-phospho-*myo*-inositol; TCH, 1,3,4,6-tetracarboxyhexane; α-Glu, α-glutamate; β-Glu, β-glutamate; Asp, aspartate; GalHl, β-galactopyranosyl-5-hydroxylysine; GB, glycine betaine; Pro, proline; NAL, N-acetyl-β-lysine; GG, glucosylglycerate; GGG, glucosyl-glucosylglycerate; MGG, mannosyl-glucosylglycerate; DAGAP, di-*N*-acetyl-glucosamine phosphate.

Polyol phosphodiesters

Within this category, di-*myo*-inositol phosphate (DIP), is the most notable example, since it has not been found in organisms thriving below 60ºC. DIP is the most widespread compatible solute among hyperthermophiles; its presence has been reported in one of the most hyperthermophilic organisms, *Pyrolobus fumarii*, but also in archaea belonging to the genera *Pyrodictium*, *Pyrococcus*, *Thermococcus*, *Methanotorris*, *Aeropyrum* and *Archaeoglobus*, and in the most hyperthermophilic bacteria known to date, *Aquifex* spp. and *Thermotoga* spp. (Gonçalves et al. 2008, Martins et al. 1997, Martins and Santos 1995, Empadinhas et al. 2001, Neves et al. 2005, Santos and da Costa 2001, Ciulla et al. 1994, Robertson et al. 1990, Lamosa et al. 1998, Gonçalves et al. 2003). DIP has been associated with thermophily not only because of its wide distribution among the heat-adapted organisms, but also because the accumulation of this solute usually increases in response to supra-optimal growth temperatures.

Other polyol phosphodiesters encountered in hyperthermophiles include mannosylated DIP-derivatives, diglycerol phosphate and glycero-phospho-*myo*-inositol. The mannosylated derivatives of DIP are known to accumulate in the hyperthermophilic bacteria *Aquifex pyrophilus*, *Thermotoga maritima*, and *Thermotoga neapolitana* (Martins et al. 1996, Lamosa et al. 2006). Diglycerol phosphate is a rare solute, which is restricted to members of the genus *Archaeoglobus*, namely *Archaeoglobus fulgidus*, where it acts as the major compatible solute under salt stress conditions (Martins et al. 1997). Glycero-phospho-*myo*-inositol resembles a structural chimera between DIP and diglycerol phosphate and was found only in *Archaeoglobus* spp. and in bacteria of the genus *Aquifex*, accumulating primarily in response to combined osmotic and temperature stresses (Lamosa et al. 2006).

Other solutes

Other solutes encountered in hyperthermophiles, although not restricted to this group of organisms, include: α- and β-glutamate, aspartate, proline, cyclic-2,3-bisphosphoglycerate, 1,3,4,6-tetracarboxy-hexane, galactosyl-hydroxylysine, and N-acetyl-β-lysine. 1,3,4,6-tetracarboxy-hexane and galactosyl-hydroxylysine are rare compatible solutes occurring only in a few hyperthermophiles like *Methanothermobacter* spp. and *Thermococcus litoralis*, respectively (Robertson et al. 1990, Robertson et al. 1992, Lamosa et al. 1998). N-acetyl-β-lysine is a very common solute among mesophiles, but it has also been observed in *Methanothermococcus thermolitotrophicus*. β-Glutamate and cyclic-2,3-bisphosphoglycerate, which were originally detected in methanogens with a broad range of growth temperature, have been found later in several hyperthermophiles. Aspartate, a negatively charged amino acid, is used as compatible solute by *Palaeococcus ferrophilus* and *Thermococcus* spp. (Santos et al. 2010).

Accumulation, role and transport

Compatible solutes not only are important for maintenance of cytoplasmatic osmotic pressure but also play an important role in the maintenance of protein structure, stability and solubility (RoeBler and Müller 2001). Many hyperthermophiles accumulate compatible solutes in response to supra-optimal temperatures, which suggests a role in thermoadaptation. Several *in vitro* experiments proved that charged solutes are generally more effective in the stabilization of protein structure against thermal denaturation and to prevent protein aggregation. For example, Faria and co-workers showed that mannosylglycerate and di-*myo*-inositol phosphate produced an increase of

8.5ºC and 11ºC, respectively, in the melting temperature of pig heart malate dehydrogenase, while the same concentration of glycerol increases only by 1ºC the melting temperature of that enzyme (Faria et al. 2008). However, the protective effect is often dependent on the pair solute/enzyme and this is particularly true in the protection of enzymes against heat inactivation, taking in consideration the irregular behavior observed for DIP (Faria et al. 2008).

Another interesting mark is the differential pattern of solute accumulation in response to different stress elements. In general, hyperthermophiles accumulate charged amino acids like aspartate or glutamate, and solutes like diglycerol phosphate or mannosylglycerate, in response to osmotic stress, while di-*myo*-inositol phosphate and its derivatives accumulate in response to heat stress. Of course, there are several exceptions to this trend, including *Palaeococcus ferrophilus* that does not synthesize DIP but accumulates mannosylglycerate and glutamate in response to heat stress (Santos et al. 2010).

Osmolytes are preferentially taken up from the medium through high affinity transporters rather than synthesized *de novo* by energetically expensive processes (Oren 1999). The solutes can be released upon death of an organism, rendering them accessible to others, which will utilize them for cellular protection or as carbon source. Several transport systems for proline, glycine betaine or ectoine are well studied in the mesophiles *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Methanohalophilus portucalensis* or *Methanosarcina* spp. (Lai et al. 2000, Roeβler et al. 2002, Peter et al. 1998, von Blohn et al. 1997, Roeβler and Müller 2001, Bremer and Kramer 2000). The same strategy is likely to occur in hyperthermophiles, however, only scarce information is available on the compatible solute transport systems used by these organisms, particularly for the solutes highly restricted to hot environments. The transport of trehalose has been

characterized in members of the *Thermococcales* and a homolog for the glycine betaine uptake system has been identified in *Archaeoglobus fulgidus* (Xavier et al. 1996, Horlacher et al. 1998, Pflüger and Müller 2004, Galinski et al. 1994).

The composition of the growth medium is intimately connected to the pattern of accumulation of compatible solutes and within the hyperthermophilic world some examples can be found: *Thermococcus litoralis* only accumulates trehalose when it is present in the growth medium and the same was observed for the solutes hydroxyproline and β-galactopyranosyl-5-hydroxylysine (Lamosa et al. 1998); *Archaeoglobus fulgidus* strain 7324 responds to osmotic stress accumulating preferentially diglycerol phosphate if lactate is used as carbon source or exclusively mannosylglycerate if starch is used instead (Gonçalves 2008). The capacity to scavenge metabolites from the medium can sometimes determine the survival of an organism that is exposed to stressful conditions. However, some organisms are deprived of such transport systems and the *de novo* synthesis may represent the only strategy to cope with fluctuations of the environmental conditions.

Biosynthesis and regulation

In the last years, several pathways for the synthesis of compatible solutes in thermophiles and hyperthermophiles were elucidated: cyclic-2,3-bisphosphoglycerate (Lehmacher et al. 1990), mannosylglycerate (Martins et al. 1999), glucosylglycerate (Costa et al. 2006; Fernandes et al. 2007), mannosylglucosylglycerate (Fernandes et al. 2010), and di-*myo*-inositol phosphate (Chen et al. 1998; Scholz et al. 1998). Generally, these solutes are synthesized in a two-step reaction, involving the formation of a

phosphorylated intermediate that is ultimately dephosphorylated, originating the final product (Fig. 1.6). Therefore, the synthesis proceeds via an irreversible process, which may result from an efficient evolutionary strategy to allow the intracellular accumulation of high levels of solutes. In some organisms an additional route may existe, in which the synthesis proceeds in a single-step without the formation of an intermediate. Furthermore, the genes for the synthesis of mannosylglycerate, glucosylglycerate, and mannosylglucosylglycerate have been identified (Empadinhas et al. 2001, Borges et al. 2004, Costa et al. 2006, Fernandes et al. 2007, Fernandes et al. 2010).

The slow progress on the knowledge of the genes involved in the synthesis of compatible solutes and the inexistence of genetic tools to manipulate these organisms are major bottlenecks in the process of understanding the physiological role of compatible solute accumulation and respective regulatory mechanisms. Despite these difficulties, work has been done in this area and particularly it was found that in *Rhodothermus marinus* the two pathways are differently regulated at the translational level: while the single step pathway is induced in response to heat stress, the two-step pathway is induced during osmotic stress (Borges et al. 2004).

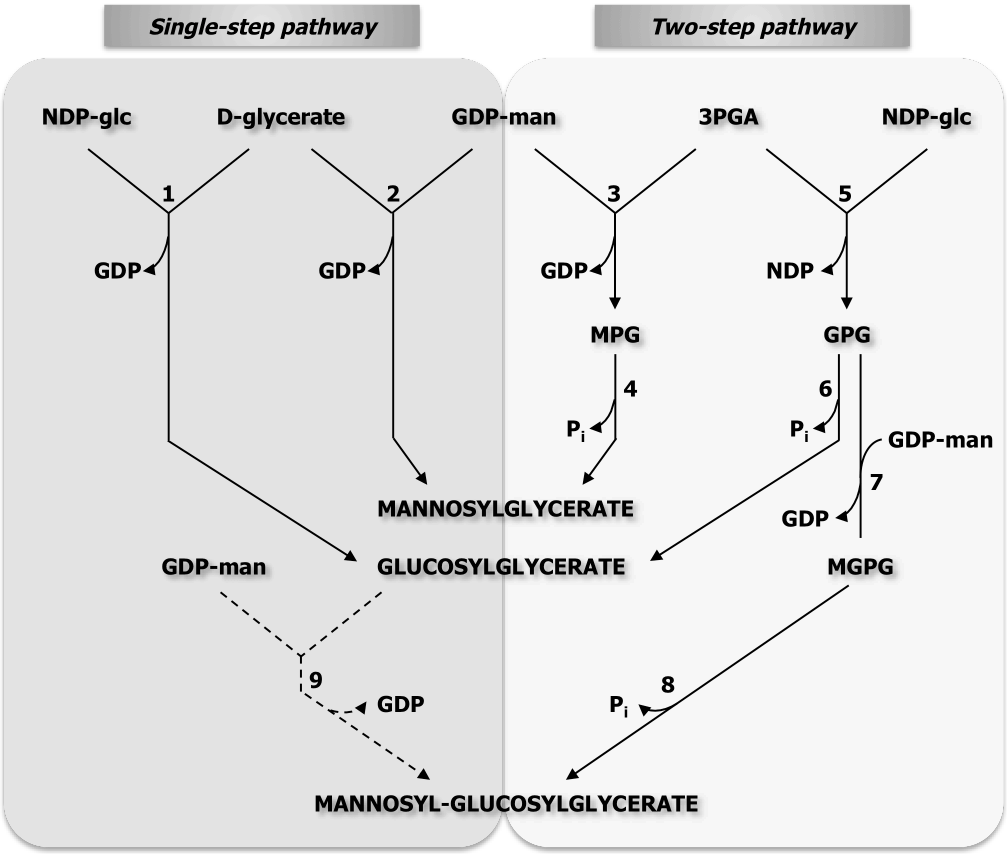


Figure 1.6. Biosynthetic pathways of mannosylglycerate, glucosylglycerate, and mannosyl-glucosylglycerate. Enzymes: (1) glucosylglycerate synthase, (2) mannosylglycerate synthase, (3) mannosyl 3-phosphoglycerate synthase, (4) mannosyl 3-phosphoglycerate phosphatase; (5) glucosyl 3-phosphoglycerate synthase, (6) glucosyl 3-phosphoglycerate phosphatase (7) mannosyl-glucosyl-3-phosphoglycerate synthase, (8) mannosyl-glucosyl 3-phosphoglycerate phosphatase, and (9) mannosyl-glucosylglycerate synthase. The discontinuous line indicates that this reaction has not been firmly established. NDP-glc, NDP-glucose, GDP-man, GDP-mannose, MPG, mannosyl-3-phosphoglycerate, 3-PGA, 3-phosphoglycerate, GPG, glucosyl 3-phosphoglycerate, GG, glucosylglycerate, MG, mannosylglycerate, MGG, mannosyl-glucosylglycerate MPG, mannosyl 3-phosphoglycerate, and MGPG, mannosyl-glucosyl-3-phosphoglycerate. (Martins et al. 1999, Costa et al. 2006, Fernandes et al. 2010).

Accumulation of β-amino acids can also be found in hyperthermophiles, but their synthesis was primarily studied in halophilic

organisms. The synthesis of N-acetyl-β-lysine is well established and involves two enzymes: first the action of lysine aminomutase isomerizes α-lysine to β-lysine, and second an acyltransferase acetylates the ε-amino group. In *Methanosarcina mazei* the two genes are organized in an operon and proved to be essential for growth under salt stress (Roberts et al. 1992, Pflüger et al. 2003). On the other hand, the synthesis of β-glutamate is more controversial and the enzymes involved in the biosynthetic pathway are not well known (Roberts 2005).

Biosynthesis of DIP

The initial reports on the biosynthesis of DIP suggested the existence of distinct pathways for the synthesis of this solute in prokaryotes (Fig. 1.7). Chen and co-workers investigated the synthesis of DIP in crude extracts of *Methanotorris igneus* and proposed that L-*myo*-inositol 1-phosphate was activated with CTP and the resulting CDP-inositol would react with *myo*-inositol to yield DIP. *myo*-Inositol results from the action of a phosphatase on L-*myo*-inositol 1-phosphate. These authors proposed the presence of two intermediates, CDP-inositol and *myo*-inositol, but failed to prove the existence of CDP-inositol (Chen et al. 1998).

At the same time, Scholz and colleagues investigated the synthesis of DIP in *Pyrococcus woesei* and suggested that the synthesis of DIP resulted from the condensation of two molecules of L-*myo*-inositol 1-phosphate with an NTP (Scholz et al. 1998). Nevertheless, both reports confirmed the activity of L-*myo*-inositol 1-phosphate synthase (MIPS), the first enzyme of the pathway, which is responsible for the conversion of glucose 6-phosphate into L-*myo*-inositol 1-phosphate. The two teams agreed on the stereochemistry of DIP, *i.e.*, the synthesis of DIP occurred under stereochemical conservation making the final product the L,L stereoisomer.

These proposals do not account for the formation of phosphorylated intermediates and hence disagree with the general trend observed for mannosylglycerate, glucosylglycerate, mannosylglucosylglycerate or even trehalose, as was mentioned earlier. When this thesis work started only the enzymes *myo*-inositol-1 phosphate and inositol monophosphatase had been characterized, and nothing was known about the regulation of DIP synthesis.

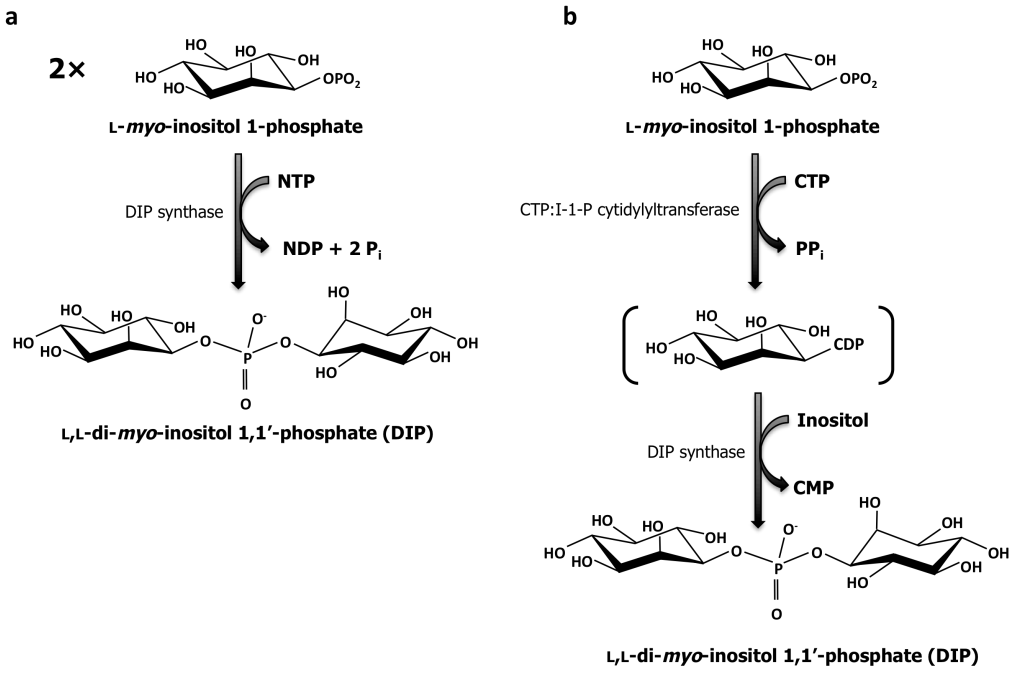


Figure. 1.7. Proposed pathways for the biosynthesis of DIP by Scholz et al. 1998 (a), and Chen et al. 1998 (b).

Inositol and derivatives

Inositol is a cyclitol with empirical formula $C_6H_{12}O_6$, composed by a cyclohexane ring substituted on each carbon by one hydroxyl group. Arrangement of the six hydroxyl groups around the inositol ring can originate

nine stereoisomers: *scyllo*-, *myo*-, *epi*-, D-*chiro*-, L-*chiro*-, *neo*-, *allo*-, *cis*- and *muco*-inositol (Fig. 1.8), but thus far only six forms, *myo*-, *scyllo*-, D-*chiro*-, L-*chiro-, neo*-, and *muco*-inositol, have been found in Nature (Murthy 2006, Michell 2008). *myo*-Inositol was isolated in 1850 from muscle extracts by Scherer who named the compound after the Greek word for muscle, inositol (Murthy 2006). *myo*-Inositol is the oldest known form of inositol and is the predominant form in Nature. It has been suggested that the other biological forms of inositol may derive from *myo*-inositol by simple interconversion (epimerization) of the hydroxyl groups, but further evidence is needed to substantiate this idea (Murthy 2006).

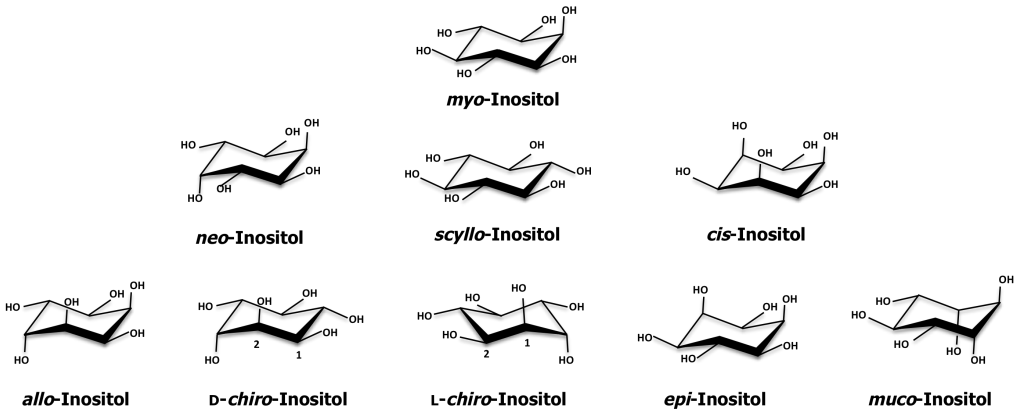


Figure 1.8. Stereoisomers of inositol.

myo-Inositol is a pro-chiral molecule that can be divided into two mirror halves (Fig. 1.9). Disrupting the symmetry of the molecule, for example by phosphorylation at carbons C_1, C_3, C_4 or C_6, will render a chiral molecule. The product of substitution at C_1 will be an enantiomer of the product of the reaction at C_3, and the same holds between carbons C_4 and C_6.

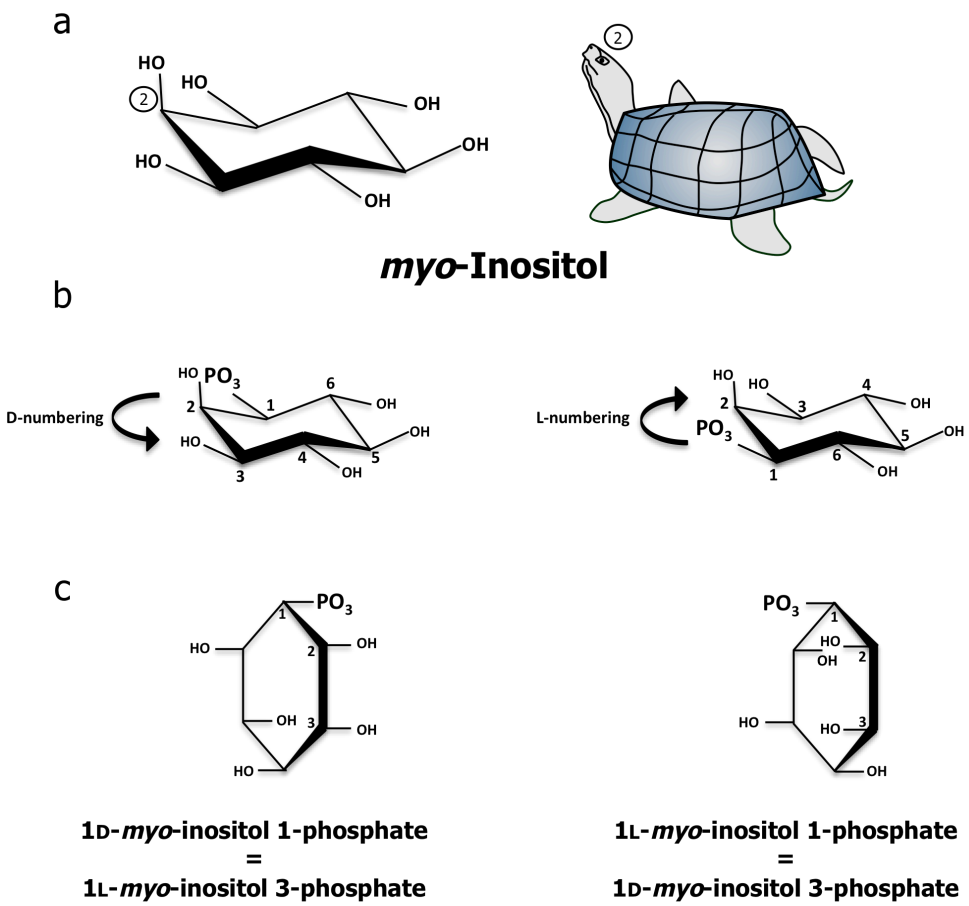


Figure 1.9. Nomenclature of inositol phosphates. (a) Chair conformation of *myo*-inositol and turtle representation used for mnemonic purposes. (b) Chair conformations of D-*myo*-inositol 1-phosphate (left) and L-*myo*-inositol 1-phosphate (right). (c) Fisher-Tollens projections of D-*myo*-inositol 1-phosphate (left) and L-*myo*-inositol 1-phosphate (right).

The nomenclature of *myo*-inositol has been a source of confusion for many years, but after several attempts, in 1989 a set of rules was established and these have been generally followed ever since (Nomenclature Commitee of the International Union of Biochemistry 1989). This is particularly important to distinguish between the L-enantiomer of *myo*-inositol 1-phosphate which is

directly formed from glucose 6-phosphate, and the other enantiomer, D-*myo*-inositol 1-phosphate, which results from the metabolism of lipids (Parthasarathy and Eisenberg 1986). Hence, it was proposed that whenever it is necessary to highlight structural relationships, the substituent groups are not necessarily numbered so that the smallest possible locant is used.

Commonly, *myo*-inositol is represented in an horizontal chair conformation and the absolute configuration of this molecule can be deciphered as follows: if the substituent on the lowest asymmetric carbon (in this case the axial carbon) is above the plane of the ring, the direction of the numbering dictates the stereoconfiguration, *i.e.*, D-configuration is assigned if the numbering goes counterclockwise, and the L-configuration is assigned if the numbering goes clockwise (Fig. 1.9b). To remember the numbering of *myo*-inositol, Agranoff pointed the resemblance between the chair conformation of *myo*-inositol and a turtle and proposed a three-dimensional visual mnemonic: the axial hydroxyl group is the turtle head, and the five equatorial hydroxyls serve as limbs and the tail (Fig. 1.9a) (Agranoff 1978). The head represents carbon 2, and if the turtle is facing left, its right hand flipper is designated D-1, and the numbering goes counterclockwise. Another way to decipher the absolute configuration of *myo*-inositol substituted at C_1 or C_3 is by vertically projecting the molecule (Fisher-Tollens projection) with the substituted carbon C_1 at the top, and carbons C_2 and C_3 on the front edge of the ring (Fig. 1.9c). If the carbons are projected to the right, the D-configuration is assigned, while if the carbons are projected to the left the L-configuration is assigned instead (Murthy 2006).

myo-Inositol has a fundamental position in cellular metabolism given that the inositol moiety is used for the synthesis of many biological compounds and structures, such as inositol phosphates, phosphoinositides (or phophatidylinositides), glycosylphosphatidylinositols, compatible solutes, or

inositol ethers/esters, which mediate several biological processes (Michell 2008).

The distribution and usage of inositol derivatives varies among the three domains of Life: inositol derivatives are found in all eukaryotes and in most archaea, but very few bacteria use them (Table 1.3). In eukaryotes, free *myo*-inositol occurs and accumulates, acting as an osmolyte, while different forms of phosphorylated inositol can be found in membrane lipids, playing important roles in signal transduction or even acting as phosphate reserve (Michel 2008). Inositol hexakisphosphate, also known as phytic acid, was the first inositol phosphate discovered. Its main role is to act as a phosphate reserve in seeds (Irvine and Schell 2001). Sixty years later, inositol 1,4,5-triphosphate was recognized as a Ca^{2+} mobilizing second messenger and in a quick succession other inositol phosphates were discovered. There are 63 possible phosphomonoesters of *myo*-inositol and at least half have been identified in eukaryotes thus far (Agranoff 2009, Irvine and Schell 2001, Michel 2008). They are important second messengers participating for example in intracellular signal transduction, calcium mobilization, vesicular movement, membrane dynamics, or channeling events (Lykidis et al. 1997, Irvine and Schell 2001). Furthermore, other inositol derivatives, such as phosphoinositides and glycosylphosphatidylinositols, also play important functions in the cell, namely signaling and structural functions. Phosphoinositides are compounds in which a phosphodiester linkage bridges inositol phosphate and diacylglycerol; phosphatidyl-*myo*-inositol (phosphatidylinositol) is the most abundant phosphatidylinositide in cells, accounting to about 90% of the inositol lipids. *myo*-Inositol is the predominant isomer of most inositol-containing lipids, although *scyllo*- and *chiro*-inositol have been found in phosphoinositide lipids of plants and animal cells, respectively (Murthy 2006). Glycosylated forms of phosphatidylinositol,

glycosylphosphatidylinositol, anchor several cell surface proteins to cell membranes by forming covalent bonds to the C-terminus of proteins (Murthy 2006, Roberts 2006, Michell 2008).

Table 1.3. Distribution and utilization of inositol in the three domains of Life. (Compiled from Irvine and Schell 2001, Murthy 2006, Roberts 2006, Michell 2008, 2011, Morita et al. 2010, Santos et al. 2010).

Domain	Cell components	Function
Archaea	Archaetidylinositol [a]	Membrane phospholipids
	DIP, GPI [b]	Compatible solutes
Bacteria	DIP, MDIP, GPI [b]	Compatible solutes
	Phosphatidylinositol [c]	Membrane phospholipids
	Mycothiol [c]	Protection
	Phosphatidylinositol 3-phosphate [c]	Signalling?
	Phosphatidylmannosides (lipomannan and lipoarabinomann) [c]	Pathogenicity
Eukarya	*myo*-Inositol, methylinositol	Compatible solutes
	Phosphatidylinositides (Phosphatidylinositol 4,5-diphosphate; Phosphatidylinositol 3,5-diphosphate; Phosphatidylinositol 3-phosphate; Phosphatidylinositol 3,4,5-triphosphate)	Signalling, cytoskeletal, secretory, membrane trafficking,
	Glycosylphosphatidylinositol	GPI anchor
	Inositol polyphosphates (inositol tetra-, penta-, or hexakisphosphate, InosP_7 or Inos P_8)	Signalling, storage of organic phosphate

a, except in extremely halophilic archaea; b, in some thermophiles and hyperthermophiles; c, in Actinobacteria. DIP, di-*myo*-inositol phosphate. GPI, glycero-phospho-*myo*-inositol.

Few bacteria synthesize and use inositol and inositol-containing compounds (Michell 2008). The high G+C Gram positive bacteria actinomycetes (*e.g.*, species of the genera *Mycobacterium*, *Corynobacterium*, or *Streptomyces*) constitute a notable exception, since they make and use a wide variety of inositol derivatives. These bacteria can import *myo*-inositol

from the environment, resorting to H^+ driven antiporters, but can also synthesize it using enzymes that probably were recruited from archaea by lateral gene transfer events during evolution (Michell 2008). Inositol is principally used for the generation of phosphatidylinositols, which are versatile structures that can be altered to anchor proteins or complex carbohydrates to the cellular membrane (Roberts 2006). In these bacteria, *myo*- and *scyllo*-inositol are precursor of many important molecules, including complex mannosylated derivatives of phosphatidylinositol, inositol lipid glycans or mycothiol, a molecule similar to glutathione that occurs in mycobacteria. Inositol can be found in the phosphatidylmannosides, namely lipoarabinomannan, present in mycobacterial cell envelopes, structures that mediate adhesion of the pathogen to the lung (Roberts 2006, Michell 2008). A recent study revealed that a *mips* (*myo*-inositol 1-phosphate synthase gene) deficient mutant of *Mycobacterium tuberculosis* was only viable if *myo*-inositol was supplied in the growth media (Movahedzadeh et al. 2004). Interestingly, bacteria are equipped with phytases, which are specialized phosphatases that degrade phytate, and provide inorganic phosphate to cells. Phytate, however, is absent in bacteria.

Although it is known for a long time that phosphoinositides play key roles in signalling processes in eukaryotic cells, only recently a lipid based signalling pathway was detected in bacteria. Phosphatidylinositol 3-phosphate, which is an important eukaryotic second messenger that mediates several cellular processes, can be transiently synthesized by mycobacteria in response to salt stress (Morita et al. 2010).

Inositol containing polyol phosphodiesters can also be found among deeply rooted hyperthermophilic bacteria, contributing to the osmotic balance and thermoadapatation of these organisms. Di-*myo*-inositol phosphate was identified in *Thermotoga* spp., *Aquifex* spp., *Persephonella marina* and

Rubrobacter xylanophilus (Santos et al. 2010). Interestingly, *Thermotoga* spp. and *Aquifex* spp., add mannose residues to DIP, forming a mannosylated derivative of this inositol-containing solute (Martins et al. 1996, Roberts 2006).

Except for the Halobacteriales, the Methanococcales and the Methanomicrobiales, most archaea use *myo*-inositol as the major headgroup of their membrane phospholipids (Morii et al. 2009). The archaeal phospholipid backbone (named archaetidyl) is a mirror image of the phospholipid backbone found in bacteria and eukaryotes (named phosphatidyl): archaeal lipids are *sn*-2,3-glycerol ethers instead of the bacterial/eukaryal *sn*-1,2-glycerol esters. Nevertheless, despite the reverse configuration of the archaeal phospholipids, the inositol phosphate headgroup is linked through the 1D-hydroxyl in all the inositol-containing phospholipids of *Bacteria*, *Archaea* and *Eukarya* (Roberts 2006, Michel 2008). *myo*-Inositol, the most used substrate for the synthesis of phosphatidylinositol, can be produced by dephosphorylation of L-*myo*-inositol 1-phosphate or D-*myo*-inositol 1-phosphate by an inositol monophosphatase (IMP) (Fig. 1.10) (Parthasarathy and Eisenberg 1986).

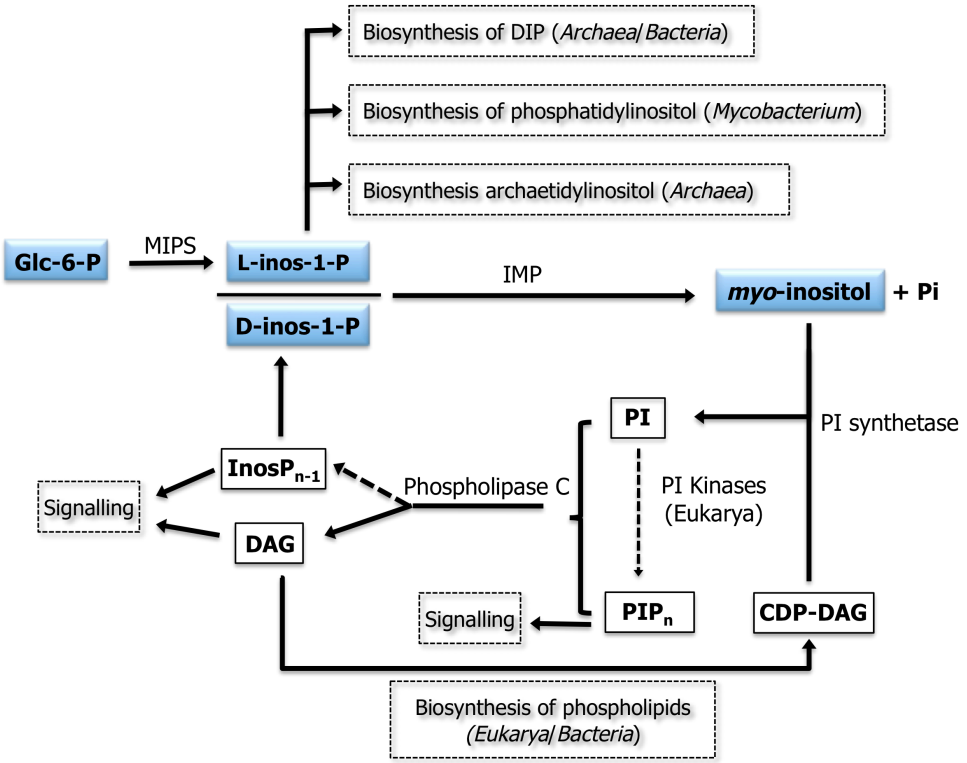


Figure 1.10. Schematic representation of the *de novo* biosynthesis of *myo*-inositol and derivatives. The two enantiomers of inositol phosphate do not have a common precursor: D-*myo*-inositol 1-phosphate originates from the cyclic synthesis and hydrolysis of phosphatidylinositol, while L-*myo*-inositol 1-phosphate originates directly from glucose 6-phosphate by the action of a NAD^+ dependent *myo*-inositol 1-phosphate synthase in an irreversible reaction, but *myo*-inositol can be produced by the dephosphorylation of either enantiomer. Glc-6-P, glucose 6-phosphate; L-inos-1-P, L-*myo*-inositol 1-phosphate; D-inos-1-P, D-*myo*-inositol 1-phosphate; Pi, inorganic phosphate; PI, phosphatidylinositol; PI synthetase, phosphatidylinositol synthetase; PIP_n, phosphorylated phosphatidylinositol; CDP-DAG, CDP-diacylglycerol; DAG, diacylglycerol; $InosP_{n-1}$, phosphorylated inositol phosphate; PI, phosphatidylinositol synthetase, PI Kinases, phosphatidylinositol kinases; MIPS, *myo*-inositol 1-phosphate synthase; IMP, inositol monophosphatase. (Adapted from Parthasarathy and Eisenberg 1986, and Michell 2011).

Some organisms, namely some archaea that live in hostile conditions, cannot import free *myo*-inositol from the environment, hence they must synthesize it. MIPS is an extremely important enzyme as it catalyzes the sole

route for the *de novo* production of *myo*-inositol in cells (Parthasarathy and Eisenberg 1986, Roberts 2006). MIPS is present in almost all eukaryotes and archaea, including the recently emerging phyla, the *Thaumarchaeota* and *Korarchaeota*. However, extremely halophilic archaea lack *mips* and *imp* genes. Probably these genes were lost by their ancestors since they became unnecessary given that archaeal inositol containing phospholipds do not form stable bilayers in saturated salt solutions (Michell 2008). The phospholipids of extreme halophiles are denominated archaetidylglycerolmethylphosphate instead of archaetidylinositolphosphate. MIPS is present only in a few bacteria, namely of the genus *Mycobacterium*, *Streptomyces* and in hyperthermophiles *Aquifex* spp. and *Thermotoga* spp..

CHAPTER 2

Pathways for the synthesis of two inositol phosphodiesters, di-*myo*-inositol phosphate and glycero-phospho-*myo*-inositol, used during stress adaptation by *Archaeoglobus fulgidus*

Part of this Chapter is published in:

Borges N, Gonçalves LG, Rodrigues MV, Siopa F, Ventura R, Maycock C, Lamosa P, Santos H (2006). The biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile *Archaeoglobus fulgidus* in stress adaptation. J Bacteriol **188**:8128-35.

Chapter 2 – Contents

Summary

Archaeoglobus fulgidus accumulates di-*myo*-inositol phosphate (DIP) and diglycerol phosphate (DGP) in response to heat and osmotic stresses, respectively, but when both stresses are combined, this hyperthermophile accumulates preferentially glycero-phospho-*myo*-inositol (GPI). In this work, the pathways for the biosynthesis of the compatible solutes DIP and GPI were established based on the detection of the relevant enzymatic activities. The synthesis of DIP proceeds from glucose 6-phosphate via four steps: (1) glucose 6-phosphate was converted into L-*myo*-inositol 1-phosphate by L-*myo*-inositol 1-phosphate synthase; (2) L-*myo*-inositol 1-phosphate was activated to CDP-inositol at the expense of CTP; this is the first demonstration of CDP-inositol synthesis in a biological system; (3) CDP-inositol was coupled with L-*myo*-inositol 1-phosphate originating di-*myo*-inositol phosphate phosphate (DIPP); (4) finally, DIPP was dephosphorylated to yield DIP by the action of a phosphatase. The synthesis of GPI proceeds via the condensation of CDP-glycerol with L-*myo*-inositol 1-phosphate, yielding the respective phosphorylated intermediate, glycero-phospho-*myo*-inositol phosphate, which is subsequently dephosphorylated to form the final product, GPI. The results reported here represent an important step towards the elucidation of the regulatory mechanisms underlying the differential accumulation of these compounds in response to heat and osmotic stresses.

Introduction

Archaeoglobus fulgidus is a hyperthermophilic archaeon first isolated from marine hydrothermal vents (Achenbach-Richter et al. 1987). In addition to the type strain, designated VC-16, a few others belonging to the same species were isolated from hot marine sediments (strain Z) (Zellner et al. 1989) and oil field water (strain 7324) (Beeder et al. 1994). The type strain has an optimal growth temperature around 83°C and like other marine hyperthermophiles, is slightly halophilic displaying optimal growth in medium containing 1.9% (wt/vol) NaCl (Stetter, 1988). Osmoregulation mechanisms in marine organisms may involve the accumulation of organic solutes, some of them very unusual (da Costa et al. 1998). The solute pool of this archaeon comprises diglycerol phosphate (DGP), di-*myo*-inositol phosphate (DIP), minor amounts of glutamate, and the newly discovered glycero-phospho-*myo*-inositol (GPI) (Lamosa et al. 2006, Martins et al. 1997). The most striking feature is the occurrence of DIP, GPI and DGP, which are polyol-phosphodiesters, a class of compatible solutes encountered in organisms thriving in hot environments but very rare in mesophiles (Santos et al. 2007, Gonçalves et al. 2003, Lamosa et al. 2006). The curious structural relationships between the three polyol phosphodiesters accumulated by *Archaeoglobus fulgidus* are also reflected in their pattern of accumulation. While DIP increases consistently in response to elevated temperature and DGP to osmotic stress, the level of GPI, which structurally resembles a chimera between DIP and DGP, seems to respond to a combination of both stresses (Gonçalves et al. 2003, Lamosa et al. 2006). Thus, the synthesis and accumulation of these three solutes seems intertwined by a regulatory mechanism linked to osmotic pressure and heat, which can only be unraveled once the biosynthetic routes of these solutes are known.

DIP was the first of these compounds to be discovered in members of the genus *Pyrococcus* (Scholz et al. 1992). Since then, the accumulation of DIP has been reported in several hyperthermophiles both from the domains *Archaea* and *Bacteria*. Presently, DIP is regarded as the canonical solute at high temperatures given that it has never been encountered in organisms with optimal growth temperature below 60ºC (Roberts et al. 2005, Santos et al. 2007). DGP is restricted to members of the genus *Archaeoglobus*, and GPI was, until now, only identified in *Archaeoglobus fulgidus* and in the two bacterial species of the genus *Aquifex*. Hence, a putative thermo-protective function of cell components in vivo was ascribed to these polyolphosphodiesters, and the ability to stabilize proteins *in vitro* has been demonstrated at least for DIP and DGP (Lamosa et al. 2000, Ramakrishnan et al. 1997, Santos et al. 2002, Scholz et al. 1992).

Nothing is known about the biosynthesis of DGP or GPI, but the biosynthesis of DIP has been studied in *Methanotorris igneus* and *Pyrococcus woesei* (Chen et al. 1998, Scholz et al. 1998), and two distinct synthetic routes were proposed. In *Methanotorris igneus*, Chen and co-workers (Chen et al. 1998) proposed that L-*myo*-inositol 1-phosphate is synthesized from glucose 6-phosphate by L-*myo*-inositol-1-phosphate synthase. Part of the L-*myo*-inositol 1-phosphate is dephosphorylated into *myo*-inositol while another part is presumably activated to CDP-inositol, both molecules then being condensed to yield DIP; biosynthesis of CDP-inositol was not observed. In *Pyrococcus woesei*, it was proposed that two molecules of L-*myo*-inositol 1-phosphate are condensed to yield DIP with the consumption of NTP (Scholz et al. 1998). The question concerning the contribution of compatible solutes from hyperthermophiles to the strategies of osmo- and thermo-adaptation can only be answered once the pathways for the synthesis of these solutes are known in detail. With this goal in mind we decided to study the pathways for the

synthesis of DIP and GPI. Herein we report the new biosynthetic pathways based on the activities of relevant enzymes detected in cell extracts of *Archaeoglobus fulgidus* and the NMR structural characterization of intermediate metabolites. Concurrently, the biosynthesis of DGP was established (Borges et al. 2006).

Materials and methods

Materials

L-glycerol 3-phosphate, DL-glycerol 3-phosphate, glucose 6-phosphate, NAD^+, CTP, ATP, UTP, GTP, CDP-glycerol, and *myo*-inositol were purchased from Sigma-Aldrich (St. Louis, USA). NaF was obtained from Pfizer (New York, USA). DL-CDP-inositol, L-*myo*-inositol 1-phosphate and D-*myo*-inositol 1-phosphate were obtained by chemical synthesis (this work). DGP and DIP were provided by bitop AG (Witten, Germany): DIP was extracted from *Pyrococcus woesei* and DGP was obtained by chemical synthesis. GPI was isolated from *Archaeoglobus fulgidus* biomass (Lamosa et al. 2006).

Organisms and growth conditions

Archaeoglobus fulgidus strain 7324 (Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany) was used in this study. The pool of compatible solutes in *Archaeoglobus fulgidus* strain 7324 was very similar to what has been described for strain VC-16 and VC-16 S (Gonçalves et al. 2003). Furthermore, compared with the type strain VC-16 and with its variant, strain VC-16S, which we used in previous studies

(Lamosa et al. 2000, Martins et al. 1997, Gonçalves et al. 2003), strain 7324 displayed greater biomass yield and different growths produced highly reproducible results insofar as enzymatic activities were concerned. *Archaeoglobus fulgidus* strain 7324 was grown anaerobically in 2-liter static vessels with a gas phase composed of N_2 and CO_2 (80:20, vol/vol) on the medium described by Labes and Schönheit (2001) with modifications. The basal medium contained (per liter) 0.5 g of yeast extract, 10 mmol of L-lactate, 3 g of PIPES, 30 g of NaCl, 2 mg of $(NH_4)_2Fe(SO_4)_2 \cdot 7H_2O$, 100 mL of salt solution (without NaCl) and 10 mL of a modified trace element solution. The salt solution contained (per liter) 74 g of $MgSO_4 \cdot 7H_2O$, 3.4 g of KCl, 27.5 g of $MgCl_2 \cdot 6H_2O$, 2.5 g of NH_4Cl, 1.4 g of $CaCl_2 \cdot 2H_2O$, and 1.4 g of K_2HPO_4. The modified trace element solution contained (per liter) 1.5 g of nitriloacetic acid, 0.05 g of $CoSO_4 \cdot 6H_2O$, 0.1 g of $CaCl_2 \cdot 2H_2O$, 0.1 g of $ZnSO_4 \cdot 7H_2O$, 0.01 g of $KAl(SO_4)_2 \cdot 12H_2O$, 0.1 g of H_3BO_3, 0.01 g of $Na_2WO_4.2H_2O$, 0.01 g of $Na_2MoO_4 \cdot 2H_2O$, 0.3 g of Na_2SeO_3, and 0.1 g of $NiCl_2 \cdot 6H_2O$. To investigate the biosynthesis of GPI and DIP, *Archaeoglobus fulgidus* cells were cultured under a combination of heat and osmotic stress (87ºC, 3% (wt/vol) NaCl). Cell growth was assessed by measuring optical density at 600 nm.

Preparation of *Archaeoglobus fulgidus* cell extracts

Cells were harvested by centrifugation (1000 × *g*, 10 min) during the late exponential growth phase (optical density at 600 nm of 0.35), and washed twice with MOPS buffer (50 mM, pH 7.6) containing 3% (wt/vol) NaCl in anaerobic conditions. The cell pellet was suspended in MOPS (50 mM, pH 7.6) containing 3% (wt/vol) NaCl. Cells were disrupted in a French press, and the cell debris was removed by centrifugation (30,000 × *g*, 4ºC, 30 min). Low

molecular mass compounds were removed in a PD-10 column (Amersham Biosciences) equilibrated with MOPS (50 mM, pH 8.1) containing 5 mM $MgCl_2$. Protein content was estimated by the BCA Protein Assay Kit (Pierce Protein Research Products, Thermo Scientific).

Enzyme assays

Enzymatic assays were performed at 80ºC in a total volume of 0.5 ml and no effort was taken to avoid oxygen. All the reaction mixtures contained 50 mM MOPS pH 8.1, 5 mM $MgCl_2$, and 5 mM of putative substrates for each investigated pathway. The reactions were initiated by addition of 50 mg (total protein) of *Archaeoglobus fulgidus* cell extract. After 1 h incubation, the reaction mixtures were centrifuged; 5 mM EDTA (pH 8.0) and 50 µl of ^2H_2O were added to the supernatant, and the pH was adjusted to 7.6. The reaction products were analyzed by ^{31}P-NMR spectroscopy.

DIP synthesis was investigated using the following substrates: CDP-inositol and L-*myo*-inositol 1-phosphate (or D-*myo*-inositol 1-phosphate or *myo*-inositol); or glucose 6-phosphate, NAD^+ and CTP. To study the specificity of the enzymes regarding nucleotide usage, all the reactions requiring CTP were also carried out in the presence of ATP, UTP, or GTP. To examine the presence of phosphorylated intermediate compounds a phosphatase inhibitor (NaF) was added to the reaction mixtures to a final concentration of 5 to 50 mM. The biosynthetic pathway of GPI was examined using the substrates: CDP-glycerol and L-*myo*-inositol 1-phosphate (or D-*myo*-inositol 1-phosphate or *myo*-inositol) or CDP-inositol and DL-glycerol 3-phosphate (or glycerol).

NMR spectroscopy

The identification of the phosphorylated intermediates in the synthesis of DIP and GPI was accomplished using ^{1}H, ^{13}C, and ^{31}P-NMR. One-dimensional spectra were recorded on a Bruker DRX500 spectrometer using standard Bruker pulse programs (Bruker, Rheinstetten, Germany) following a strategy previously described (Lamosa et al. 1998, Silva et al. 1999). Proton, carbon, and phosphorus chemical shifts are relative to 3 (trimethylsilyl)propanesulfonic acid (sodium salt) (at 0 ppm), external methanol (at 49.3 ppm) or external 85% H_3PO_4 (at 0 ppm), respectively. Assignment of resonances to DIP, GPI, DGP, CDP-glycerol, and CDP-inositol were confirmed by addition of small amounts of the authentic compounds.

Chemical synthesis

DL-*myo*-inositol 1-phosphate was obtained through a significant modification of the established procedures according to Figure 2.1 (Massy and Wyss 1990, Watanabe et al. 1997, Yu and Fraser-Reid 1988). Enantiomerically pure D- and L-*myo*-inositol 1-phosphate were prepared by optical resolution of DL-2,3,4,5,6-penta-*O*-benzyl-*myo*-inositol via separation of the diastereoisomeric camphanates (Dreef et al. 1991). After hydrolysis, 1D- and 1L-2,3,4,5,6-penta-*O*-benzyl-*myo*-inositol were obtained separately, and converted into optically pure D- and L-*myo*-inositol 1-phosphate as for the racemic compound (Fig. 2.1). The synthesis of CDP-inositol was accomplished according to the established procedures (Moffatt and Khorana 1961, Roseman et al. 1961).

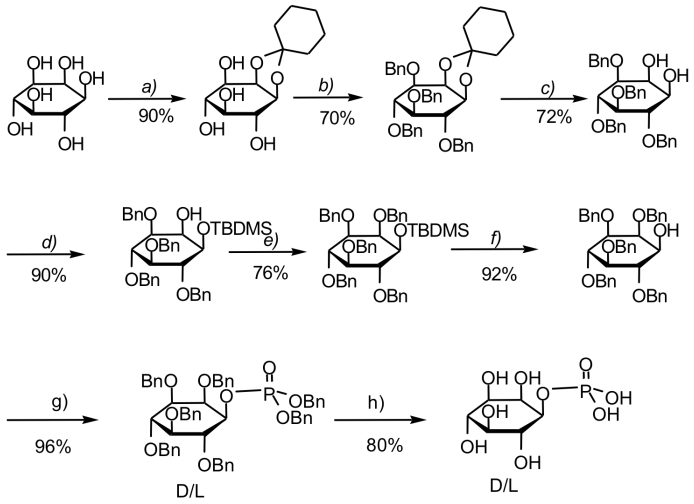


Figure 2.1. Scheme for the chemical synthesis of DL-*myo*-inositol 1-phosphate from *myo*-inositol. **a**) 1. Cyclohexanone (16eq), DMF/Toluene, *p*-TsOH, 6h, 150°C; 2. EtOH, *p*-TsOH 1h, r.t. **b**) BnBr (12eq), DMF, 16h, r.t. **c**) TFA (7eq), MeOH (1.2eq), DCM **d**) TBDMSOTf (1.2eq), $(i\text{-}Pr)_2Net$ (1.2eq), DCM **e**) BnBr (4eq), NaH (4eq), DMF, 2h, r.t. **f)** TBAF 1M (1.5eq), THF, 2h, r.t. **g**) 1. $(BnO)_2PN(iPr)_2$ (1.5eq), *1H*-Tetrazole (1.5eq), DCM, 2h30m, r.t. 2. MCPBA (2eq), DCM, 1h30m, -40°C-0°C **h**)Pd/C, 20 psi H_2, EtOH, 2h, r.t. The yields of the individual steps are indicated as percentages.

Results

Biosynthesis of DIP

To investigate DIP biosynthesis, glucose 6-phosphate (the substrate of inositol 1-phosphate synthase), plus NAD^+ and CTP were incubated with a cell extract of *Archaeoglobus fulgidus*. DIP and CDP-inositol were identified by ^{31}P-NMR as final products in the reaction mixture, and the assignments were confirmed by spiking with the pure compounds. DIP and CDP-inositol were also formed when CTP and L-*myo*-inositol 1-phosphate were provided as substrates (Fig. 2.2a). We verified that ATP, UTP or GTP did not replace CTP in the synthesis of DIP or CDP-inositol. D-*myo*-inositol 1-phosphate and *myo*-inositol were also

examined as possible substrates of this cytidylyltransferase, but none of them was used. Therefore, this new enzyme was specific for L-*myo*-inositol 1-phosphate and CTP. By analogy with other cytidylyltransferases, pyrophosphate was expected as a product of the reaction, but was not detected in the ^{31}P-NMR spectra of the reaction mixtures; this is explained by a very active pyrophosphatase present in *Archaeoglobus fulgidus* extracts (data not shown).

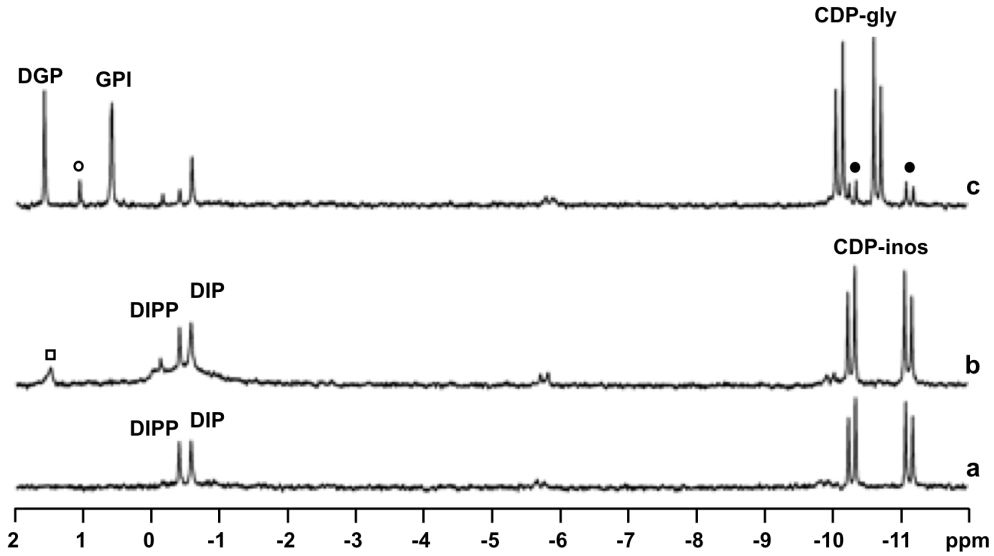


Figure 2.2. ^{31}P-NMR spectra of the final products from 1-hour incubation at 80ºC of reaction mixtures containing *Archaeoglobus fulgidus* cell extract (50 mg total protein) in 50 mM MOPS buffer (pH 8.1), 5 mM $MgCl_2$, and 5 mM of the following substrates: (a) CTP plus L-*myo*-inositol 1-phosphate; (b) CDP-inositol plus L-*myo*-inositol 1-phosphate; and (c) CDP-glycerol plus L-*myo*-inositol 1-phosphate. Resonances due to CDP-inositol, CDP-glycerol, DIP, DIPP, DGP and GPI are labeled. The spectra shown are the difference between the final spectra of the reaction mixtures with the substrates and the spectrum of the control (cell extract with no additions after 1h incubation). The resonances labeled with (○) and (□) are due to contaminants in the CDP-glycerol and CDP-inositol preparations, respectively; the resonances labeled with (•) are from CDP-inositol formed during the reaction (see text for details). DGP, diglycerol phosphate; GPI, glycero-phospho-*myo*-inositol; DIP, di-*myo*-inositol phosphate; DIPP, di-*myo*-inositol phosphate phosphate; CDP-inos, CDP-inositol; CDP-glycerol, CDP-glycerol.

These results definitely show that the synthesis of DIP proceeds from L-*myo*-inositol 1-phosphate and CTP via the formation of the intermediate CDP-inositol. To our knowledge, this is the first demonstration of the biosynthesis of CDP-inositol. The subsequent reactions in the synthesis of DIP were further studied using CDP-inositol obtained by chemical synthesis and each of the following putative substrates: L-*myo*-inositol 1-phosphate, D-*myo*-inositol-1-phospahte, and *myo*-inositol. DIP synthesis was not detected from CDP-inositol and D-*myo*-inositol 1-phosphate or *myo*-inositol, but was clearly apparent when CDP-inositol and L-*myo*-inositol 1-phosphate were used as substrates (Fig. 2.2b). Curiously, two resonances (-0.57 and -0.38 ppm) were detected in the phosphodiester region of the ^{31}P-NMR spectrum of the reaction mixture. The resonance at -0.57 ppm was due to DIP, but the other resonance could not be immediately assigned. The hypothesis that it was due to a phosphorylated form of DIP seemed like a reasonable one. To confirm this hypothesis the reaction was repeated with addition of 7 mM NaF, a general phosphatase inhibitor. As a result, the intensity of the resonance at -0.38 ppm increased relatively to the spectrum of the reaction mixture without fluoride. Moreover, after treatment with alkaline phosphatase, the new phosphodiester resonance disappeared, with the concomitant increase in intensity of the DIP resonance (Fig. 2.3). The putative phosphorylated intermediate, was purified by anion exchange chromatography and its structure established by NMR as di-*myo*-inositol phosphate phosphate (DIPP) (Chapter 4).

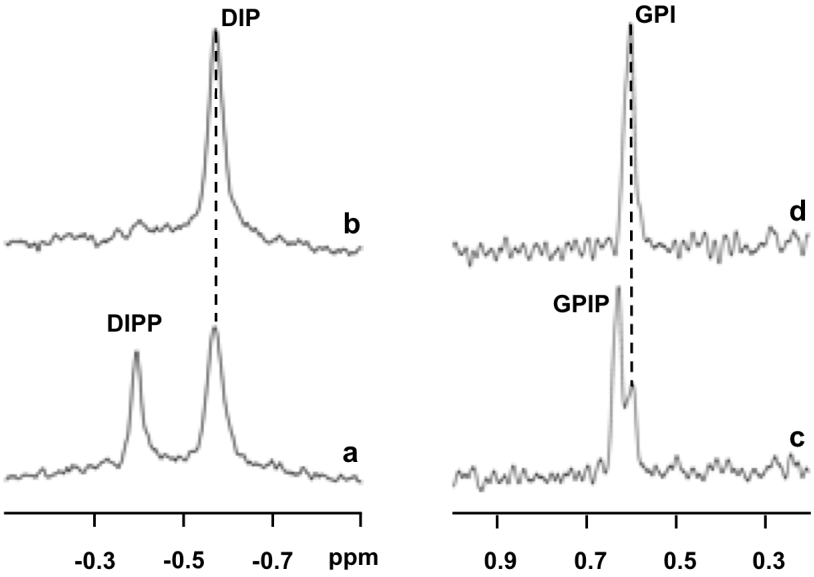


Figure 2.3. ^{31}P-NMR spectra showing the conversion of the phosphorylated intermediates into the final products (GPI and DIP) by the action of alkaline phosphatase. In the bottom are the spectra of the reaction mixtures containing: cell extract, CTP, NaF and (a) L-*myo*-inositol 1-phosphate, and (c) L-*myo*-inositol 1-phosphate plus DL-glycerol 3-phosphate. After spectral acquisition, samples (a) and (c) were treated with alkaline phosphatase, originating spectra (b) and (d), respectively. The phosphodiester resonances due to DIP (-0.57 ppm), DIPP (-0.38 ppm), GPI (0.60 ppm) and GPIP (0.63 ppm) are represented. GPI, glycero-phospho-*myo*-inositol; GPIP, glycero-phospho-*myo*-inositol phosphate; DIP, di-*myo*-inositol phosphate; and DIPP, di-*myo*-inositol phosphate phosphate;

Biosynthesis of GPI

An array of experiments was designed to investigate the synthesis of GPI. Given the chimerical structure of GPI containing glycerol and inositol moieties, CDP-glycerol as well as CDP-inositol were examined as putative polyol donors. CDP-glycerol was combined with *myo*-inositol, L-*myo*-inositol 1-phosphate, or D-*myo*-inositol 1-phosphate as potential polyol acceptors; CDP-inositol was

examined in combination with glycerol or DL-glycerol 3-phosphate. Maximal formation of GPI (resonance at 0.60 ppm) was observed from CDP-glycerol and L-*myo*-inositol 1-phosphate (Fig. 2.2c). Vestigial amounts of GPI were also detected when CDP-inositol and DL-glycerol 3-phosphate were used (less than 10% of that for CDP-glycerol and L-*myo*-inositol 1-phosphate) showing that the synthase is not absolutely specific for CDP-glycerol and L-*myo*-inositol 1-phosphate. All the other assays were negative for GPI synthesis. As for DIP, CTP was the only nucleoside triphosphate that supported GPI formation.

DGP was synthesized whenever CDP-glycerol was supplied, a result explained by the generation of the additional substrate required for DGP synthesis, glycerol phosphate, during the thermal degradation of CDP-glycerol. The addition of CDP-glycerol and L-*myo*-inositol 1-phosphate led to the synthesis of GPI and DGP as major products, but minor amounts of DIP and CDP-inositol were also detected (Fig. 2.2c). This could be explained by the exchange between the glycerol and inositol moieties arising from the reversibility of the reaction catalyzed by "GPI synthase" in conjunction with the observed minor affinity for the alternative substrates (CDP-inositol and glycerol 3-phosphate).

The presence of an intermediate phosphorylated compound was also confirmed in the synthesis of GPI: incubation of cell extracts with CDP-glycerol, L-*myo*-inositol 1-phosphate and 40 mM NaF led to the appearance of a new phosphorus resonance at 0.63 ppm that disappeared upon treatment with alkaline phosphatase (Fig. 2.3). The nature of the phosphorylated intermediate was established by NMR analysis after partial purification (Chapter 4).

Discussion

The biosynthetic pathways of DIP and GPI in *Archaeoglobus fulgidus* strain 7324 were established based on the detection of the relevant enzymatic activities in cell extracts and the identification of the intermediates and final products by NMR (Fig. 2.4). The synthesis of DGP, the major polyolphosphodiester accumulating in *Archaeoglobus fulgidus* was elucidated concomitantly (Borges et al. 2006). The routes for the synthesis of DIP, GPI and DGP have several features in common: i) in all cases CTP is the only nucleoside triphosphate used to activate the polyol, either glycerol or inositol; thus, CDP-glycerol or CDP-inositol are the polyol donors; ii) the polyol acceptor is always in the phosphorylated form, *i.e.*, L-glycerol 3-phosphate or L-*myo*-inositol 1-phosphate; iii) the central biosynthetic step involves the condensation of CDP-polyol with the phosphorylated polyol to form a phosphorylated intermediate; iv) finally, the dephosphorylation of this intermediate leads to the final product, DGP, DIP or GPI.

This work provides firm evidence for the involvement of phosphorylated intermediates in the synthesis of these compatible solutes (see Chapters 3 and 4). There are no earlier reports on the synthesis of GPI, but the synthesis of DIP has been studied by other teams, who put forward two different pathways (Chen et al. 1998, Scholz et al. 1998). None of the proposed pathways contemplates the occurrence of a phosphorylated intermediate. Chen and co-workers (Chen et al. 1998) proposed that in *Methanotorris igneus* DIP is formed in a single step from CDP-inositol and *myo*-inositol. On the other hand, the pathway proposed for *Pyrococcus woesei* involves the condensation of two molecules of L-*myo*-inositol 1-phosphate with the consumption of NTP, but no evidence for the formation of NDP-inositol was given (Scholz et al. 1998).

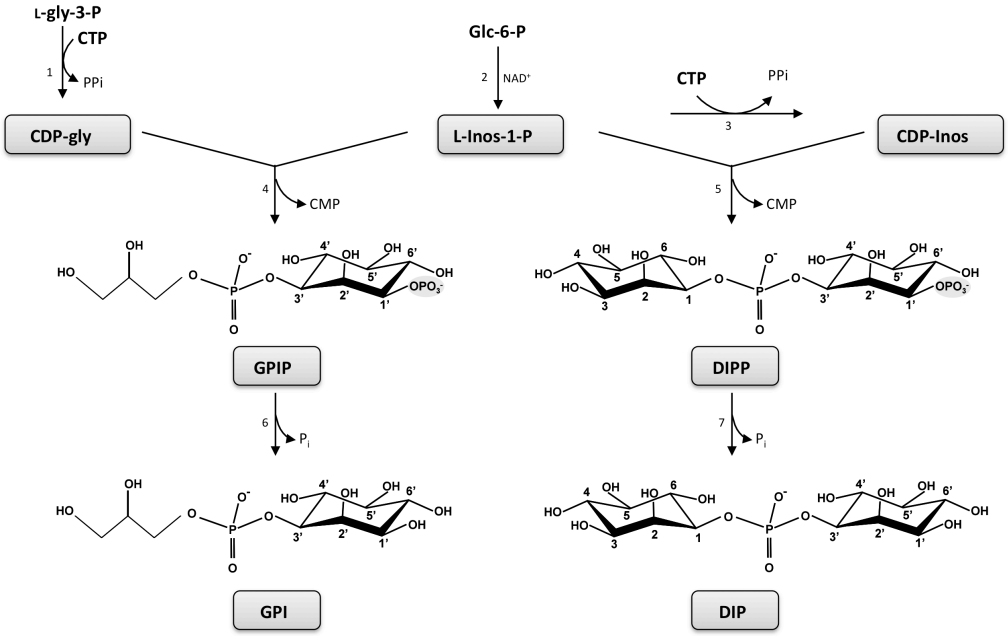


Figure 2.4. Proposed pathways for the synthesis of DIP and GPI in *Archaeoglobus fulgidus*. Enzymes: (1) CTP:glycerol 3-phosphate cytidylyltransferase, (2) L-*myo*-inositol 1-phosphate synthase, (3) CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase, (4) GPIP synthase, (5) DIPP synthase, (6) GPIP phosphatase, and (7) DIPP phosphatase. DIP, di-*myo*-inositol phosphate, DIPP, di-*myo*-inositol phosphate phosphate (DIPP); GPI, glycero-phospho-*myo*-inositol; GPIP, glycero-phospho-*myo*-inositol phosphate; CDP-gly, CDP-glycerol; CDP-Inos, CDP-inositol; Glc-6-P, glucose 6-phosphate; L-Inos1-P, L-*myo*-inositol 1-phosphate; gly, glycerol; and L-gly-3-P, L-glycerol 3-phosphate.

The pathway for DIP synthesis proposed here for *Archaeoglobus fulgidus* is different from that found in *Methanotorris igneus* insofar as L-*myo*-inositol 1-phosphate, instead of *myo*-inositol, is used by the synthase and the product of this reaction is a phosphorylated form of DIP instead of DIP itself. Our proposal for DIP synthesis in *Archaeoglobus fulgidus* is also different from that put forward a few years ago for *Pyrococcus woesei* (Scholz et al. 1998), but the coincidence of the input reagents in the two routes is a curious detail:

two molecules of L-*myo*-inositol 1-phosphate and one molecule of NTP. We demonstrated, however, that CDP-inositol and the phosphorylated form of DIP are intermediate metabolites in the pathway. To determine whether these discrepancies were associated with the distant phylogenetic relationship between the two organisms studied, the biosynthesis of DIP was also examined in *Pyrococcus furiosus*, an archaeon closely related to *Pyrococcus woesei*. Like in *Archaeoglobus fulgidus*, we observed the formation of CDP-inositol and the phosphorylated intermediate (DIPP) in cell extracts of *Pyrococcus furiosus* when L-*myo*-inositol 1-phosphate and CTP were provided as substrates (data not shown). Therefore, the biosynthetic routes proposed in Figure 2.4 are supported by solid experimental evidence. It is intriguing, however, that this reaction scheme points towards a DL stereochemical configuration of the inositol moieties in DIP, rather than the LL configuration found in DIP isolated from *Pyrococcus* spp. (van Leeuwen et al. 1994). The solution for this apparent paradox requires the identification of the gene encoding DIPP synthase and the characterization of the catalytic reaction (see Chapters 3 and 4).

To our knowledge, this is the first time that the synthesis of CDP-inositol is demonstrated in a biological system. Although Chen et al. (1998) have shown that CDP-inositol is used for DIP synthesis in cell extracts of *Methanotorris igneus*, these authors failed to detect formation of this activated compound. Remarkably, searches in the literature confirmed that CDP-inositol has not been found in any metabolic pathway other than DIP biosynthesis. This is somewhat surprising since the inositol moiety is found in a variety of biological molecules, *e.g.*, as a constituent of phospholipid polar heads and glycolipid protein anchors (Ikezawa 2001, Kent et al. 1991). On the other hand, this finding strengthens the view that hyperthermophiles have developed unique metabolic strategies to synthesize unique molecules that

are likely to play a role in the adaptation of these organisms to extremely hot environments.

In contrast with CDP-inositol, the synthesis of CDP-glycerol is widely distributed and has been studied in detail particularly in *Bacillus subtilis* (Park et al. 1993, Pattridge et al. 2003). CTP:L-glycerol 3-phosphate cytidylyltransferase (GCT) catalyses the reversible formation of CDP-glycerol and pyrophosphate from CTP and L-glycerol 3-phosphate. A search in the *Archaeoglobus fulgidus* genome revealed an ORF (AF1418) sharing 36% amino acid identity with the GCT of *Bacillus subtilis*, and most likely encodes the activity responsible for CDP-glycerol synthesis in the archaeon. CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase remains unknown, but it should be related with GCT, both enzymes belonging to the cytidylyltransferase family.

GPI is a rare solute so far only encountered in hyperthermophiles of the genera *Archaeoglobus* and *Aquifex*. The occurrence of GPI as a cell metabolite has not been reported elsewhere, however the unit glycero-phospho-inositol is found in phosphatidylinositol, an important constituent of lipid membranes. Interestingly, in eukaryotes, the synthesis proceeds in a single step involving the condensation of CDP-diacylglycerol with *myo*-inositol to yield the final product, phosphatidylinositol (Kent et al. 1991), while in archaea *myo*-inositol 1-phosphate is used instead of *myo*-inositol (Morii et al. 2009).

The preponderance of phosphodiester compounds in the solute pool of the hyperthermophile *Archaeoglobus fulgidus* is noteworthy, especially because this class of compounds is rarely found in connection with stress adaptation of mesophilic organisms. In fact, the only case reported refers to glycerophosphocholine that acts as an osmoprotectant in yeast cells, and also protects renal medullary mammalian cells from high, physiological

concentrations of NaCl and urea (Kiewietdejonge et al. 2006, Gallazzini and Burg 2009). In contrast, sugars, polyols, amino acids, and non-charged derivatives are compatible solutes typical of mesophiles.

We have shown that the synthesis of DIP, DGP, and GPI proceeds in two steps with the involvement of a phosphorylated intermediate that is subsequently dephosphorylated. Interestingly, this is the most common strategy for the synthesis of sugars or sugar derivatives for the purpose of osmoprotection or thermoprotection, such as mannosylglycerate, glucosylglycerate, glucosylglycerol, trehalose, and sucrose (Costa et al. 2006, Giaever et al. 1988, Hagemann et al. 2001, Martins et al. 1997, Porchia and Salerno 1996). It seems as though the concerted action of a synthase and a phosphatase, resulting in the irreversible synthesis of the final product, was selected throughout evolution as the most efficient strategy to allow accumulation of compatible solutes to high levels, which can reach the molar range in the cytoplasm.

In conclusion, the work presented here and the knowledge on the biosynthesis of DGP (Borges et al. 2006) uncovered the routes for the synthesis of three major compatible solutes in *Archaeoglobus fulgidus*. These pathways must be differentially regulated by salt and temperature as the pattern of accumulation of DIP, DGP, and GPI depends clearly on the type of stress imposed. This study adds an important step towards the identification of the genes and enzymes implicated in these novel biosynthetic routes.

Acknowledgments and work contributions

F. Siopa and R. Ventura synthesized DL-CDP-inositol, L-*myo*-inositol 1-phosphate and D-*myo*-inositol 1-phosphate. C. P. Almeida and L. G.

Gonçalves cultivated *Archaeoglobus fulgidus* strain 7324 and performed the analysis of the compatible solute pool. P. Lamosa acquired the NMR spectra for the characterization of phosphorylated intermediates. The NMR spectrometers at CERMAX are part of the National NMR Network and were acquired with funds from FCT and FEDER. This work was funded by the European Commission, contract COOP-CT-2003-508644, and Fundação para a Ciência e a Tecnologia/FEDER, and POCTI Portugal, Projects A004/2005 Action V.5.1. M. V. Rodrigues, L. G. Gonçalves and N. Borges acknowledge FCT for research fellowships.

CHAPTER 3

A bifunctional enzyme catalyzes the synthesis of di-*myo*-inositol phosphate in several (hyper)thermophiles

Part of this Chapter is published in:

Rodrigues MV, Borges N, Henriques M, Lamosa P, Ventura R, Fernandes C, Empadinhas N, Maycock C, da Costa MS, Santos H (2007). Bifunctional CTP:inositol 1-phosphate cytidylyltransferase/CDP-inositol:inositol 1-phosphate transferase, the key enzyme for di-*myo*-inositol phosphate synthesis in several (hyper)thermophiles. J Bacteriol **189**:5405-12.

Chapter 3 – Contents

Summary

A genomic approach was used to identify the genes involved in the synthesis of di-*myo*-inositol phosphate (DIP). Cloning and expression in *Escherichia coli* of the putative genes for CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) and CDP-inositol:inositol 1-phosphate transferase (DIPPS or DIPP synthase) from several (hyper)thermophiles (*Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus* and *Rubrobacter xylanophilus*) confirmed the presence of those activities in the gene products. The DIPP synthase activity was found in a bifunctional enzyme that catalyzed the condensation of CTP and L-*myo*-inositol 1-phosphate into CDP-L-*myo*-inositol as well as the synthesis of DIPP from CDP-L-*myo*-inositol and L-*myo*-inositol 1-phosphate. The IPCT was absolutely specific for CTP and L-*myo*-inositol 1-phosphate; the DIPP synthase used only L-*myo*-inositol 1-phosphate as alcohol acceptor, but CDP-glycerol as well as CDP-L-*myo*-inositol, and CDP-D-*myo*-inositol were recognized as alcohol donors. Genome analysis showed homologous genes in all organisms known to accumulate DIP and for which genome sequences were available. In most cases, the two activities (IPCT and DIPP synthase) were fused in a single gene product, but separate genes were predicted in *Aeropyrum pernix*, *Thermotoga maritima*, and *Hyperthermus butylicus*. This work uncovered the genes involved in the synthesis of DIP, a solute exclusive to (hyper)thermophiles.

Introduction

Di-*myo*-inositol phosphate (DIP) is the most widespread organic solute in microorganisms adapted to hot environments (Santos et al. 2007). This compound was first identified by Hensel and coworkers in the archaeon *Pyrococcus woesei* (Scholz et al. 1992), and later encountered in many other hyperthermophiles, including members of the genera *Methanotorris*, *Thermococcus*, *Thermotoga*, *Aquifex*, *Pyrodictium*, *Aeropyrum*, *Archaeoglobus*, *Stetteria*, and *Pyrolobus* (Ciulla et al. 1994, Gonçalves et al. 2003, Lamosa et al. 1998, Lamosa et al. 2006, Martins et al. 1996, Martins et al. 1997, Santos et al. 2007). Additionally, DIP occurs as a minor solute in the thermophilic bacteria *Rubrobacter xylanophilus* and *Persephonella marina* (Empadinhas et al. 2007, Santos et al. 2007). Thus far, DIP has never been encountered in organisms with optimal growth temperature below 60ºC, hence the assumption that it plays a role in the thermoprotection of cellular components *in vivo* is often stated (Ramakrishnan et al. 1997, Santos and da Costa 2002, Scholz et al. 1992).

The question concerning the contribution of compatible solutes from hyperthermophiles to the mechanisms of osmo- and thermo-adaptation can only be answered once the pathways for the synthesis of these solutes are known in detail, *i.e.*, the genes, enzymes, substrates and reaction products are fully characterized. With this goal in mind our team has elucidated the biosynthetic pathways of mannosylglycerate, and characterized the respective genes and enzymes (Borges et al. 2004, Empadinhas et al. 2001, Martins et al. 1999). Recently, we have successfully identified the routes for the synthesis of diglycerol phosphate, di-*myo*-inositol phosphate and glycero-phospho-*myo*-inositol in *Archaeoglobus fulgidus* (Chapter 2). The sequence of reactions leading to DIP synthesis was established on the basis of the

detection of the relevant enzymatic activities in cell extracts of *Archaeoglobus fulgidus* and the structural characterization of intermediates and final products by NMR (Chapter 2). L-*myo*-inositol 1-phosphate is activated to CDP-inositol via the activity of CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase; subsequent condensation of CDP-inositol and L-*myo*-inositol 1-phosphate yields a phosphorylated form of DIP designated DIPP. Finally, DIPP is dephosphorylated to yield DIP. Even though DIP was discovered more than fifteen years ago, the genes and enzymes for the synthesis of this solute remain elusive. In this context, the main objective of the present work was to identify the genes implicated in the synthesis of DIP, a solute exclusive to (hyper)thermophiles. For this purpose, a genomic approach was used. The genes encoding CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT) and CDP-inositol:inositol 1-phosphate transferase (DIPPS or DIPP synthase) in several (hyper)thermophiles were firmly identified by functional expression in *Escherichia coli*.

Materials and methods

Materials

DL-glycerol 3-phosphate, glucose 6-phosphate, NAD^+, CTP, ATP, UTP, GTP, CDP-glycerol, glycerol, *myo*-inositol, SDS, deoxycholic acid, sodium cholate hydrate, LDAO and CHAPS were purchased from Sigma-Aldrich (St. Louis, MO, USA). Fos-choline-12 was purchased from Anatrace (Maumee, OH, USA). Mega-9 and DDM were acquired from GLYCON Biochemicals GmbH (Luckenwalde, Germany). Triton X-100 was purchased from BDH-Chemicals. CDP-D-*myo*-inositol was obtained by chemical synthesis (this work). CDP-L-*myo*-inositol was synthesized enzymatically (this work). DIP from *Pyrococcus*

woesei was obtained from bitop AG (Witten, Germany) and GPI was isolated from *Archaeoglobus fulgidus* biomass (Lamosa et al. 2006).

Organisms and growth conditions

Pyrococcus furiosus strain 3638^T, *Archaeoglobus fulgidus* strain 7324, *Rubrobacter xylanophilus* strain 9941^T and *Thermotoga maritima* strain 3109^T were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig, Germany. *Thermococcus kodakarensis* strain KOD1, and biomass of *Aquifex aeolicus* strain VF5 were kindly provided by T. Imanaka (Kyoto University, Japan), and H. Huber (University of Regensburg, Germany), respectively. *Archaeoglobus fulgidus, Pyrococcus furiosus* and *Rubrobacter xylanophilus* were cultivated as previously described in Chapter 2, Martins and Santos (1995) and Empadinhas et al. (2007), respectively. *Thermotoga maritima* was grown in medium 343 as described in the Deutsche Sammlung von Mikroorganismen und Zellkulturen catalogue.

Partial purification of native *Archaeoglobus fulgidus* IPCT/DIPPS

Cells of *Archaeoglobus fulgidus* were harvested during late exponential phase of growth (optical density at 600 nm of 0.35), and suspended in Tris-HCl (20 mM, pH 7.6), containing 5 mM $MgCl_2$ and DNase I (10 µg/ml). After cell disruption and removal of cell debris by centrifugation, the supernatant was dialyzed against Tris-HCl (20 mM, pH 7.6) prior to fractionation by fast protein liquid chromatography (Amersham Biosciences). The cell extract was applied to a Q-Sepharose column (20 mM Tris-HCl, pH 7.6), which was eluted with a linear gradient of NaCl in the same buffer. Activities of IPCT and DIPP

synthase were detected in the flow-through. Active samples were loaded onto a second Q-Sepharose column, equilibrated at pH 8.1, and again, the two activities were not adsorbed. Subsequently, the active fractions were applied to a SP-Sepharose column (20 mM Tris-HCl, pH 7.0) and both activities were detected in the flow-through. The active fractions were pooled and dialyzed against Tris-HCl (50 mM, pH 7.6) containing 1 M ammonium sulfate. The sample was applied to a Phenyl-Sepharose column, and elution carried out with a linear gradient of ammonium sulfate (1 M – 0 M). The two activities co-eluted at 0.1 M of $(NH_4)_2SO_4$.

Cloning and expression of putative *ipct/dipps* genes

Chromosomal DNA from *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Rubrobacter xylanophilus* and *Aquifex aeolicus* was isolated according to Ramakrishnan and Adams (Ramakrishnan and Adams 1995). The *ipct/dipps* gene from *Pyrococcus furiosus* (PF1058), *Archaeoglobus fulgidus* (AF0263), *Rubrobacter xylanophilus* (Rxyl_1212), *Thermococcus kodakarensis* (TK2279) and *Aquifex aeolicus* (aq_1367) was amplified by PCR using *Pfu* DNA polymerase (Fermentas). Many sequencing errors were found in the gene Rxyl_1212 (www.jgi.doe.gov), where around 10% of the residues are incorrectly determined. Therefore, the correct sequence determined by PCR using a DNA polymerase with proofreading activity, *Pfu* DNA polymerase, was submitted to GenBank with the Accession Number EF523341. The bifunctional IPCT/DIPPS from *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus* and *Rubrobacter xylanophilus* have between 428 to 436 amino acids, while the DIPP synthase from *Archaeoglobus fulgidus* is annotated with 490 amino acids. The extension of

approximately 50 amino acids in the N-terminal region of DIPP synthase from *Archaeoglobus fulgidus* is presented in Figure 3.1.

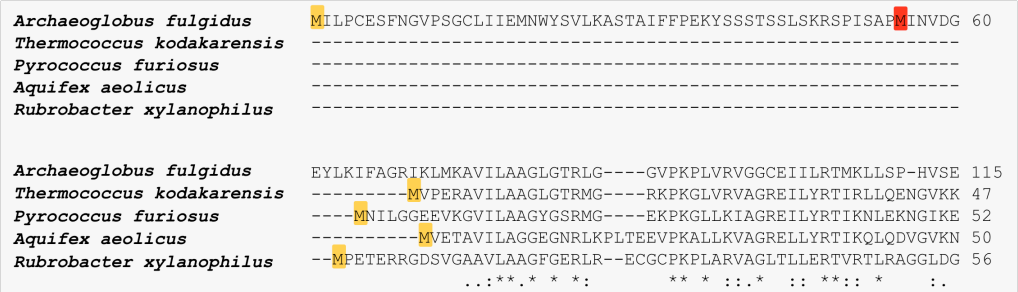


Figure 3.1. Multiple sequence alignment of the N-terminal amino acid sequences of putative IPCT/DIPPS from several (hyper)thermophiles: *Thermococcus kodakarensis* (*ipct/dipps* gene TK2279), *Pyrococcus furiosus* (*ipct/dipps* gene PF1058), *Aquifex aeolicus* (*ipct/dipps* gene aq_1367), *Archaeoglobus fulgidus* (*ipct/dipps* gene AF0263), and *Rubrobacter xylanophilus* (*ipct/dipps* gene Rxyl_1212). The alignment was generated with the software package ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/). Orange boxes highlight the initial methionine of the IPCT/DIPPS as annotated in the respective genomes. Red box highlights the initial methionine used for cloning of the smaller IPCT/DIPPS from *Archaeoglobus fulgidus*.

Hence, the gene AF0263 lacking the N-terminal extension (*i.e.* lacking the first 162 nucleotides) as well as the other genes mentioned above were cloned in different plasmids (Table 3.1) following standard protocols (Sambrook et al. 1989). In order to select the best condition for the production of large amounts of recombinant IPCT/DIPPS, the plasmids were transformed into different *E. coli* strains (Table 3.1) and several expression conditions were tested, such as variable IPTG concentration (0.1 - 1 mM), induction times (3 - 16 h) and incubation temperatures (18ºC - 37ºC). Constructions pET23a-Af, pET23a-Pf, pET23a-Aa, pTRC99a-Tk, and pET30a-Rx (Table 3.1) were used for the experiments concerning the identification of

the putative genes encoding IPCT/DIPPS activity. *E. coli* BL21(DE3) cells, bearing the constructs, were grown at 37ºC in LB medium with ampicillin (100 µg/ml) to an optical density at 600 nm of 0.5, and induced with 1 mM IPTG for 4 h. Kanamycin (30 µg/ml) was used for the selection of *E. coli* BL21(DE3) bearing the *Rubrobacter xylanophilus* gene. The cells were harvested, suspended in Tris-HCl (20 mM, pH 7.6) containing 10 mM $MgCl_2$ and disrupted in a French press; cell debris was removed by centrifugation (30,000 × *g*, 4ºC, 1 h). Total protein was estimated by the Bradford method (Bradford 1976).

Construction pET23a-tAf (Table 3.1) was used in the attempts to solubilize the recombinant IPCT/DIPPS with detergents. Cells were harvested and suspended in MOPS buffer (50 mM, pH 8.1) containing 5 mM $MgCl_2$. In a subsequent experiment constructions pET19b-Af and pET19b-Tk (Table 3.1) were used; the recombinant enzyme from *Archaeoglobus fulgidus* was produced in *E. coli* C41(DE), while the recombinant enzyme from *Thermococcus kodakarensis* was produced in *E. coli* BL21(DE3). Cells were harvested and suspended in Tris-HCl (20 mM, pH 7.6), containing 10 mM $MgCl_2$. In all the cases, the cell suspension was treated as described above.

Assays for recombinant enzymes

The activity of IPCT/DIPPS was determined in cell extracts of *E. coli*, harboring plasmids with the *ipct/dipps* genes. The reaction mixtures, in a total volume of 0.5 ml, containing 20 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM of the putative substrates, and cell extract (around 12 mg of total protein), were incubated for 1 h at 80ºC (or 45ºC in the case of the *Rubrobacter*

xylanophilus enzyme). The reaction products were analyzed by ^{31}P-NMR spectroscopy.

Table 3.1. List of constructions of *ipct*/*dipps* genes used in this study.

Gene	Protein length (nº of aa)	Vector	Restriction enzimes	Histidine tag	Construction name	*E. coli* strains
AF0263	490	pET23a	NdeI; BamHI	No	pET23a-Af	BL21(DE3) Rosetta2(DE3) C41(DE3) C43(DE3)
		pET19b	NdeI; BamHI	N-terminus	pET19b-Af	BL21(DE3) Rosetta2(DE3)
AF0263	436*	pET23a	NdeI; BamHI	No	pET23a-tAf	BL21(DE3) BL21(DE3)pLysS Rosetta2(DE3) C41(DE3)
aq_1367	428	pET23a	NdeI; XhoI	No	pET23a-Aa	BL21(DE3) Rosetta2(DE3)
TK2279	432	pTRC99a	NcoI; BamHI	No	pTRC99a-Tk	BL21(DE3) Rosetta2(DE)
		pET19b	NdeI; XhoI	N-terminus	pET19b-Tk	BL21(DE3) Rosetta2(DE) XL1-Blue
PF1058	430	pET19b	NdeI; XhoI	N-terminus	pET19b-Pf	BL21(DE3) Rosetta2(DE)
		pET23a	NdeI; HindIII	No	pET23a-Pf	BL21(DE3) Rosetta2(DE)
Rxyl1212	436	pET30a	**	No	pET30a-Rx	BL21(DE3) Rosetta2(DE)

* The first 162 nucleotides (corresponding to 54 amino acids) were not cloned, *i.e.*, the gene AF0263 was cloned from the ATG at position 163. ** This construction was provided by Prof. Milton Costa, University of Coimbra, Portugal. aa, amino acids.

Solubilization of recombinant IPCT/DIPPS with detergents

Attempts to solubilize the bifunctional IPCT/DIPPS were performed according to established protocols (Niegowski et al. 2006, Gutmann et al. 2007). A cell-free extract of *E. coli*, harboring pET23a-tAf, was ultra-centrifuged (100,000 x *g*, 4ºC, 1.5 h) for membrane separation. The supernatant was removed and the membranes (pellet) were solubilized in MOPS buffer (50 mM, pH 8.1), containing 5 mM $MgCl_2$ (Sample A); the supernatant and the solubilized pellet were examined for activity. Immediately after, 1-ml aliquots of the solubilized membrane fraction (around 6.8 mg of total protein) were added into eppendorfs containing detergents (DDM, Triton X-100, Mega-9, LDAO, Fos-choline-12, CHAPS, sodium cholate, or deoxycholic acid) in a final concentration of 1% and 2% (wt/vol). The samples were homogenized with strong agitation for 3 h at 4ºC and then heated for 10 min at 60ºC to denature thermolabile proteins. Centrifugation to remove precipitated material was performed (30,000 x *g*, 4ºC, 15 min) and the supernatant was directly tested for IPCT/DIPPS activity without any effort to remove the detergents. Protein purity was assessed by SDS-PAGE and the gel was stained with Coomassie Blue.

Promising results were obtained with Mega-9 and DDM, hence these detergents were selected for further solubilization experiments. As a strategy to optimize the purification process of IPCT/DIPPS, supernatants containing the recombinant enzyme with histidine tags from *Thermococcus kodakarensis* and *Archaeoglobus fulgidus* (constructions pET19b-Af and pET19b-Tk) were ultracentrifuged (80,000 x *g*, 4ºC, 1.5 h) and the resulting pellets solubilized with Mega-9 and DDM. The solubilization buffer was composed of Tris-HCl (20 mM), glycerol 20% (vol/vol), 20 mM imidazole, 300 mM NaCl and 1% (wt/vol) DDM or Mega-9. For each sample the volume was adjusted to 2 ml with

solubilization buffer. Homogenization was performed with strong agitation (200 rpm) during 45 min on ice. Afterwards the samples were ultracentrifuged (80,000 x *g*, 4ºC, 1.5 h). The supernatants were examined for IPCT/DIPPS and then loaded onto a HisTrap HP column (GE Healthcare). The purification was performed according to the manufacture instructions; the binding buffer contained 20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, glycerol 20% (wt/vol) and 1% DDM or Mega-9, pH 7.6; the elution buffer contained 20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole, 20% glycerol (wt/vol) and 1% DDM or Mega-9 (wt/vol), pH 7.6. Possible active fractions were combined, concentrated and dialyzed using an Amicon Centricon Centrifugal filter device (Millipore Corporation, Billerica, MA, USA) and Tris-HCl (20 mM, pH 7.6) supplemented with 10 mM $MgCl_2$. After dialysis each sample was examined for IPCT/DIPPS activity and protein purity was judged by SDS-PAGE on a gel stained with Coomassie Blue. Protein content was estimated by the Bradford method (Bradford 1976).

Assays for recombinant enzymes solubilized with detergents

The activity of IPCT/DIPPS from *Archaeoglobus fulgidus* (construction pET23a-tAf) was examined in reaction mixtures of 0.4 ml, containing MOPS (50 mM, pH 8.1), 5 mM $MgCl_2$, 3.5 mM of the putative substrates, and protein; the total amount of protein varied according to the detergent as follows: 0.2 to 0.6 mg for Mega-9, Triton X-100, sodium cholate, deoxycholic acid, DDM, LDAO and CHAPS; and 1.1 to 1.3 mg for Fos-choline-12. The differences in the protein content were due to the differentiated precipitation of thermolabile proteins (see above). Reaction mixtures were incubated for 1 h at 80ºC and the products were analyzed by ^{31}P-NMR spectroscopy.

The activity of IPCT/DIPPS from *Thermococcus kodakarensis* and *Archaeoglobus fulgidus* (constructions pET19b-Af and pET19b-Tk) was investigated in 0.4 ml reaction mixtures, containing Tris-HCl (20 mM, pH 7.6), 10 mM $MgCl_2$, 2 - 4 mM of substrates. Reactions were incubated for 1 h at 80ºC (when the protein content in the sample was 2.1 mg; 4 mM of substrates were used) or 2 h at 80ºC (when the amount of protein was 0.15 mg; 2 mM of substrates were used). The reaction products were analyzed by ^{31}P-NMR spectroscopy.

Characterization of recombinant enzymes

The substrate specificity of CTP:inositol 1-phosphate cytidylyltransferase (IPCT) was studied in assays using CTP and each of the following alcohol donors: *myo*-inositol, L-*myo*-inositol 1-phosphate, D-*myo*-inositol 1-phosphate, glycerol, and DL-glycerol 3-phosphate; additionally, CTP, ATP, GTP, or UTP were examined as putative nucleotidyl donors in combination with L-*myo*-inositol 1-phosphate. The substrate specificity of CDP-inositol:inositol 1-phosphate transferase (DIPPS or DIPP synthase) was investigated using *myo*-inositol, L-*myo*-inositol 1-phosphate, D-*myo*-inositol 1-phosphate, glycerol, and DL-glycerol 3-phosphate as alcohol acceptors and CDP-L-*myo*-inositol, CDP-D-*myo*-inositol, and CDP-glycerol as alcohol donors.

Preparation of CDP-L-*myo*-inositol

CDP-L-*myo*-inositol was produced using cell extracts of *E. coli* BL21(DE3) harboring a plasmid with the gene (EF523341) (*Rubrobacter xylanophilus*). The cell extract was incubated at 45ºC with CTP and L-*myo*-inositol

1-phosphate. After 1 h, the reaction mixtures were treated with alkaline phosphatase (30 min at 37ºC) to dephosphorylate residual *myo*-inositol 1-phosphate. The sample was lyophilized, the residue dissolved in 1-propanol/25% ammonia (1:1.5, vol/vol), and applied to a Silica gel S column. The same solvent system was used for elution. CDP-L-*myo*-inositol eluted first and was followed by *myo*-inositol. The fractions containing CDP-L-*myo*-inositol were pooled after analysis by ^{1}H- and ^{31}P-NMR for purity assessment.

NMR spectroscopy

The identification of metabolites involved in the synthesis of DIP and GPI was accomplished by using ^{1}H-, ^{31}P-, and ^{13}C-NMR on a Bruker DRX500 (Bruker, Rheinstetten, Germany). Proton and phosphorus chemical shifts are relative to 3-(trimethylsilyl)propanesulfonic acid (at 0 ppm) and external 85% H_3PO_4 (at 0 ppm), respectively.

Chemical synthesis

D- and L-*myo*-inositol 1-phosphate were obtained through a significant modification and improvement of the established procedures, as described in Chapter 2. Namely, the selective protection of the 1-OH of 3,4,5,6-tetra-*O*-benzyl-*myo*-inositol was carried out using TBDMSOTf in the presence of diisopropylethylamine with 90% yield of the desired product. Benzylation of the remaining free 2-OH group followed by fluorolysis of the silyl ether afforded racemic1,2,3,4,5,6-penta-*O*-benzyl-*myo*-inositol (Chapter 2, Fig. 2.1). Enantiomerically pure D- and L-*myo*-inositol 1-phosphate were prepared by

resolution of racemic 1,2,3,4,5,6-penta-*O*-benzyl-*myo*-inositol via separation of the diastereoisomeric camphanates (Dreef et al. 1991). After hydrolysis, 1D- and 1L-2,3,4,5,6-penta-*O*-benzyl-*myo*-inositol were obtained separately (L$[a]_D^{20}$ +5.96 (c=0.805, CH_2Cl_2), Literature (Dreef et al. 1991). $[a]_D^{20}$ +9.2 (c=1, $CHCl_3$); D$[a]_D^{20}$ -5.39 (c=0.715, CH_2Cl_2), Literature (Dreef et al. 1991) $[a]_D^{20}$ -9.0 (c=1, $CHCl_3$)), and converted into optically pure D- and L-*myo*-inositol 1-phosphate as for the racemic mixture. The coupling to form CDP-D-*myo*-inositol was accomplished according to established procedures (Moffatt et al. 1961, Roseman et al. 1961).

Results

Identification of genes for DIP synthesis in *Archaeoglobus fulgidus*

In a first approach, the purification of the DIPP synthase activity from *Archaeoglobus fulgidus* cell extracts was carried out to obtain amino acid sequence information that would lead to the identification of the respective gene in the genome sequence of this organism. We failed to reach a satisfactory purification yield, but noticed that DIPP synthase activity was not separated from the CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase activity even after four chromatographic steps (Q-Sepharose, pH values 7.6 and 8.1, SP-Sepharose, and Phenyl-Sepharose). The observed co-purification of the two activities led to the hint that they could be present in a single polypeptide chain. Further evidence in support of this hypothesis came from the analysis of the *Archaeoglobus fulgidus* genome (www.tigr.org). We found five genes (AF0263, AF1143, AF1744, AF2044, and AF2299) whose predicted products belong to the family of proteins containing a domain characteristic of

CDP-alcohol phosphatidyltransferases (www.sanger.ac.uk//cgi-bin/Pfam). Three of them (AF1143, AF1744, and AF2044) are annotated as coding for enzymes implicated in the synthesis of phospholipids, leaving two genes (AF0263 and AF2299) as good candidates to encode the DIPP synthase. Gene AF0263, encoding a bifunctional protein with homology to nucleotidyltransferases (N-terminal domain) and to CDP-alcohol phosphatidyltransferases (C-terminal domain), appeared specially promising. Meaningfully, genes homologous to AF0263 were found in all the genomes of organisms known to accumulate DIP (*Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Thermococcus kodakarensis*, *Aquifex aeolicus*, *Aeropyrum pernix*, *Thermotoga maritima* and *Rubrobacter xylanophilus*). Furthermore, genes encoding L-*myo*-inositol 1-phosphate synthase, the enzyme catalyzing the first reaction in DIP synthesis (see Table 3.3), were found in the immediate flanking region in *Thermococcus kodakarensis*, *Rubrobacter xylanophilus*, *Thermotoga maritima*, *Persephonella marina* and *Aeropyrum pernix*, substantiating the suspected connection between AF0263 and DIP synthesis. Therefore, we proceeded to clone and express this gene in *E. coli* with the objective of validating the predicted functions.

Cell extracts of *E. coli* BL21(DE3) harboring the plasmid containing the gene AF0263 were examined for the presence of CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase/CDP-inositol:inositol 1-phosphate transferase activities using CTP and L-*myo*-inositol 1-phosphate as substrates. DIPP and vestigial amounts of CDP-inositol were identified by ^{31}P-NMR as products in the reaction mixture (Fig. 3.2). We verified that these products were not formed in extracts of *E. coli* BL21(DE3) cells harboring the empty plasmid. These results showed that AF0263 encodes a bifunctional enzyme able to catalize the synthesis of DIPP from CTP and L-*myo*-inositol 1-phosphate, via CDP-inositol.

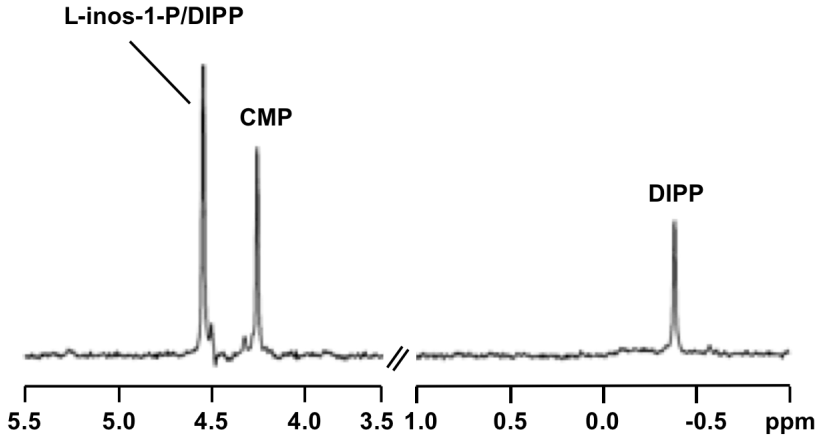


Figure 3.2. ^{31}P-NMR spectra of the final products resulting from incubation at 80ºC of an *E. coli* extract (12 mg total protein), harboring the *ipct/dipps* gene from *Archaeoglobus fulgidus*, with 5 mM CTP plus 5 mM L-*myo*-inositol 1-phosphate. The spectrum is the difference between the spectrum of the reaction mixture with addition of substrates, and the spectrum of the control (reaction mixture without substrate addition). Spectra were run at a sample temperature of 25ºC. Symbols: L-inos-1-P, L-*myo*-inositol 1-phosphate; CMP, cytidine 5'-monophosphate; DIPP, phosphorylated form of di-*myo*-inositol phosphate.

Characterization of recombinant enzymes

Research effort was directed to overproduce and purify the recombinant IPCT/DIPPS from *Archaeoglobus fulgidus*. Different plasmids and expression conditions were tested, but in all cases we failed to obtain large amounts of the recombinant enzyme. Furthermore, all purification strategies (heat-treatment and/or liquid chromatography) applied to extracts of *E. coli* BL21(DE3) harboring the construction with the gene AF0263 led to loss of the DIPP synthase activity, *i.e.*, only the cytidylyltransferase activity (IPCT) was detected after heat treatment or chromatography. The gene was also expressed in *E. coli* Rosetta2(DE), which carries a plasmid containing the tRNA genes for codons rarely used in *E. coli*, with equally poor results. Expression of

homologs from *Thermococcus kodakarensis*, *Pyrococcus furiosus* and *Aquifex aeolicus* led to similar results. The apparent instability of the DIPP synthase activity led us to suspect that we might be in the presence of a membrane protein. Using informatics tools to predict transmembrane helices in proteins, namely TMHMM Server v. 2.0, from the Center for Biological Sequence analysis (Technical University of Denmark DTU; http://www.cbs.dtu.dk/), the complete DIPP synthase sequence showed high probability of occurrence of transmembrane helices in the C-terminal domain, *i.e.*, in the DIPP synthase domain. Hence, we tried to purify this protein in a hydrophobic environment using several detergents: anionic (deoxycholic acid, sodium cholate); zwitterionic (CHAPS, fos-chline-12, LDAO); and non-ionic (Mega-9, triton X-100, DDM). Analysis of the SDS-PAGE gels revealed that there was no significant improvement in the purification of IPCT/DIPPS, *i.e.*, there were still many contaminating proteins. DIPP synthase activity was detected in fractions that were solubilized with sodium cholate, Mega-9, DDM and triton X-100, even after heat treatment (Fig. 3.3). Thus, in face of these results we decided to attempt the purification of IPCT/DIPPS in the presence of DDM and Mega-9 and to ease the purification procedure we used recombinant IPCT/DIPPS (from *Thermococcus kodakarensis* and *Archaeoglobus fulgidus*) with a histidine tag in the N-terminus. Despite our efforts, the DIPP synthase activity was lost after the HisTrap column.

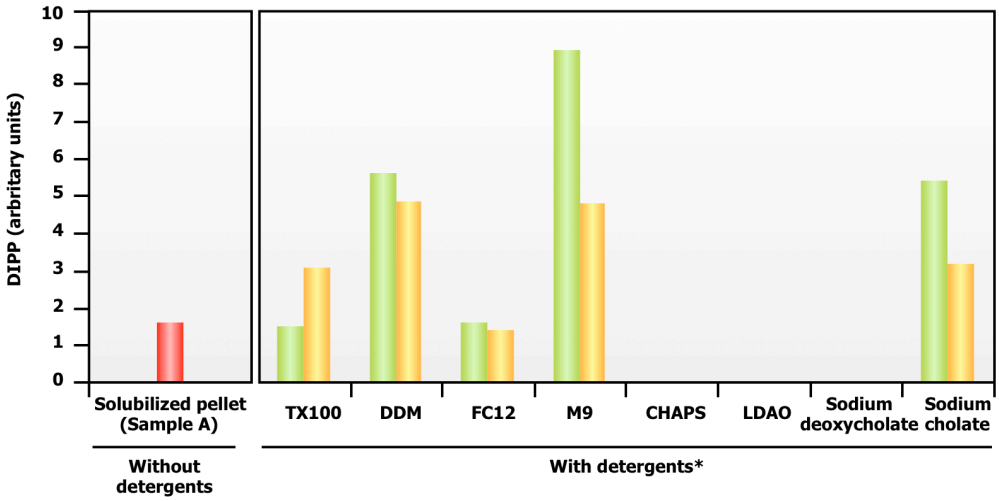


Figure 3.3: Amount of DIPP formed by IPCT/DIPPS at different stages of the purification process. DIPP was measured by ^{31}P-NMR; the area of the NMR signal is proportional to the concentration of the product; DIPP is represented in arbitrary units, *i.e.*, the area of the NMR signal·min^{-1}·mg protein^{-1} (total amount of protein). The area of the DIPP signal in sample A was considered to be 100; the intensity of the DIPP signals in the spectra of the other samples are relative to that of sample A. Sample A: pellet after ultracentrifugation solubilized in 50 mM MOPS, 5 mM $MgCl_2$ (pH 8.2). *Pellets after ultracentrifugation were homogenized in 50 mM MOPS, 5 mM $MgCl_2$ (pH 8.2) containing 1% detergent (green bars) or 2% detergent (orange bars), and heat treated for 10 min at 60°C before assaying IPCT/DIPPS activity (see Materials and methods for details). TX100, Triton X-100; FC12, Fos-choline-12; M9, Mega-9.

In view of the apparent instability of the second activity and the failure to overproduce the respective protein, the subsequent characterization of the enzyme activities was performed directly in cell extracts of *E. coli* Bl21(DE3) harboring the relevant plasmid. The recombinant enzyme from *Rubrobacter xylanophilus* showed a distinctive behavior insofar as the DIPP synthase activity was never detected; only the cytidylyltransferase activity was observed. The enzyme synthesized CDP-L-*myo*-inositol from CTP and L-*myo*-inositol 1-phosphate with high yield and this property was exploited to produce CDP-L-*myo*-inositol required for the substrate specificity assays (see below). It was confirmed that *myo*-inositol was not a substrate for this

enzyme. The substrate specificity of the bifunctional IPCT/DIPPS from *Archaeoglobus fulgidus* was investigated in regard to both cytidylyltransferase and DIPP synthase activities. The cytidylyltransferase was absolutely specific for CTP and L-*myo*-inositol 1-phosphate. Other nucleotide donors (ATP, GTP, and UTP) and other alcohol acceptors (*myo*-inositol, D-*myo*-inositol 1-phosphate, glycerol, and DL-glycerol 3-phosphate) were not substrates for the enzyme. The DIPP synthase recognized CDP-L-*myo*-inositol, CDP-D-*myo*-inositol and CDP-glycerol as alcohol donors, but L-*myo*-inositol 1-phosphate was the only alcohol acceptor used by the enzyme. The other potential substrates examined were *myo*-inositol, D-*myo*-inositol 1-phosphate, glycerol, and DL-glycerol 3-phosphate (Table 3.2).

Table 3.2: Substrate specificity of the bifunctional enzyme IPCT/DIPPS.

Cytidylyltransferase activity (N-terminal domain)		
Nucleotide donors	**Alcohol acceptors**	**Product**
CTP*	*myo*-inositol L-inos-1-P* D-inos-1-P Glycerol DL -gly-3-P	CDP- L-*myo*-inositol
ATP GTP UTP	L-inos-1-P	
DIPP synthase activity (C-terminal domain)		
Nucleotide donors	**Alcohol acceptors**	**Product**
CDP- L-*myo*-inositol* CDP- D-*myo*-inositol*	*myo*-inositol L-inos-1-P* D-inos-1-P Glycerol DL-gly-3-P	DIPP
CDP-glycerol*	*myo*-inositol L-inos-1-P* D-inos-1-P Glycerol DL-gly-3-P	GPIP

DL-gly-3-P, DL-glycerol 3-phosphate; L-inos-1-P, L-*myo*-inositol 1-phosphate; D-inos-1-P, D-*myo*-inositol 1-phosphate; GPIP, phosphorylated form of glycero-phospho-*myo*-inositol; DIPP, phosphorylated form of di-*myo*-inositol phosphate. * designates the substrates used by the enzyme.

It was verified by ^{31}P-NMR analysis that glycero-phospho-*myo*-inositol phosphate (GPIP) was the final product when CDP-glycerol was provided as alcohol donor (Fig. 3.4). We checked that the recombinant, bifunctional IPCT/DIPPS from *Pyrococcus furiosus*, *Thermococcus kodakarensis*, and *Aquifex aeolicus* used L-*myo*-inositol 1-phosphate and CTP as substrates for the synthesis of CDP-L-*myo*-inositol and DIPP. All recombinant enzymes were able to catalyze also the synthesis of GPIP when CDP-glycerol was provided in combination with L-*myo*-inositol 1-phosphate.

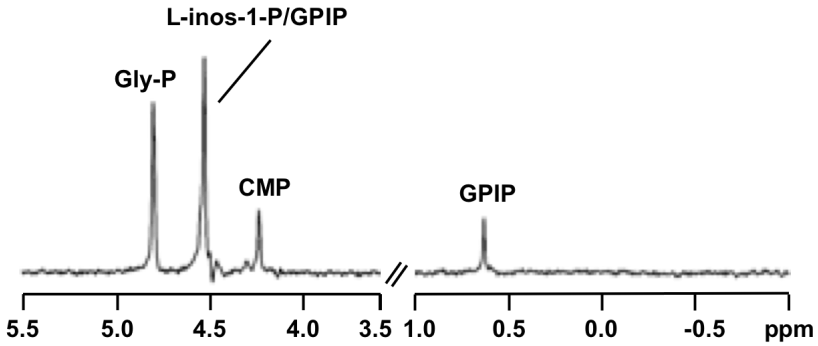


Figure 3.4. ^{31}P-NMR spectrum of the final products resulting from incubation at 80ºC of an *E. coli* extract (12 mg total protein), harboring the *ipct*/*dipps* gene from *Archaeoglobus fulgidus*, with 5 mM CDP-glycerol plus 5 mM L-*myo*-inositol 1-phosphate. The trace is the difference between the spectrum of the reaction mixture with addition of substrates, and the spectrum of the control (reaction mixture with no substrate addition). Spectra run at a sample temperature of 25ºC. Symbols: Gly-P, DL-glycerol 3-phosphate; L-inos-1-P, L-*myo*-inositol 1-phosphate; GPIP, glycero-phospho-*myo*-inositol phosphate; CMP, cytidine 5'-monophosphate.

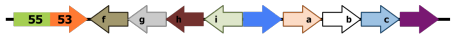
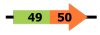
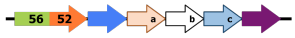
Discussion

In the previous chapter we established the biosynthetic pathway of DIP in *Archaeoglobus fulgidus* and showed that CDP-inositol and the phosphorylated form of DIP are intermediate metabolites in DIP synthesis. Furthermore, the

presence of CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT), CDP-inositol:inositol 1-phosphate transferase (DIPPS or DIPP synthase), and DIPP phosphatase activities was demonstrated in extracts of *Archaeoglobus fulgidus* and *Pyrococcus furiosus* (Chapter 2). As a logical extension of this work we set out to characterize the genes and enzymes ascribed to the biosynthetic pathway of DIP. A genomic approach combined with clues originating from our attempts to purify the relevant activities led us to propose the gene AF0263 as a strong candidate to encode the bifunctional IPCT/DIPPS. The final proof came from the functional expression of the gene in *E. coli*: the recombinant protein in cell extracts catalyzed the synthesis of DIPP from L-*myo*-inositol 1-phosphate and CTP, via CDP-inositol. Moreover, cloning and expression of the homologous genes from *Pyrococcus furiosus*, *Aquifex aeolicus*, and *Thermococcus kodakarensis* revealed identical function.

A BLAST search in public databases with the bifunctional IPCT/DIPPS from *Archaeoglobus fulgidus* resulted in the identification of homologous proteins in several cultured organisms, all of them known to thrive in hot environments (Table 3.3). The two activities, cytidylyltransferase (IPCT) and DIPP synthase (DIPPS), are encoded in a single polypeptide chain in *Archaeoglobus profundus*, *Pyrococcus* spp., *Thermococcus* spp., *Candidatus korarchaeon cryptofilum*, *Ferroglobus placidus*, *Methanocaldococcus infernus*, *Hydrogenivirga* strain 128-5-RI-1, *Thermobaculum terrenum*, *Rubrobacter xylanophilus*, *Persephonella marina*, and *Aquifex aeolicus*, while separate genes were found in *Hyperthermus butylicus, Aeropyrum pernix*, and *Thermotoga* spp.. Shortly before this work was submitted for publication, the demonstration of this function for the genes of *Thermotoga maritima* appeared in the literature (Rodionov et al. 2007).

Table 3.3. Genomic organization and flanking regions of IPCT/DIPPS from *Archaeoglobus fulgidus* and homologous proteins.

Organism	T_{opt} (°C)	Genomic organization	Accumulation of DIP
Archaea			
Pyrococcus abyssi	100		Predicted
Pyrococcus furiosus	100		Known[5]
Hyperthermus butylicus	99		Predicted
Pyrococcus horikoshii	95		Known[4]
Aeropyrum pernix	90		Known[6]
Thermococcus gammatolerans	88		Predicted
Archaeoglobus profundus	85		Known[2]
Candidatus korarchaeum cryptophilum	85		Predicted
Ferroglobus placidus	85		Predicted
Methanocaldococcus infernus	85		Predicted
Thermococcus barophilus	85		Predicted
Thermococcus kodakarensis	85		Known[3]
Archaeoglobus fulgidus	83		Known[1]
Thermococcus sp. AM4	82		Predicted
Thermococcus onnurineus	80		Predicted
Thermococcus sibiricus	78		Predicted
Bacteria			
Aquifex aeolicus	85		Known[7]
Thermotoga spp.	80-85		Known[10] /predicted
Hydrogenivirga sp 128-5RI-1	70-75		Predicted
Persephonella marina	70		Known[8]
Rubrobacter xylanophilus	50		Known[9]
Thermobaculum terrenum	67		Predicted

IPCT/DIPPS from *Archaeoglobus fulgidus* (NCBI accession number NP_069101) was considered as reference; *Archaeoglobus profundus* (YP_003400502.1); *Thermococcus gammatolerans* (YP_002960219.1); *Thermococcus* sp. AM4 (ZP_04879407.1); *Thermococcus kodakarensis* (YP_184692); *Thermococcus onnurineus* (YP_002307633.1); *Thermococcus sibiricus* (YP_002994415.1); *Thermococcus barophilus* (ZP_04876988.1); *Pyrococcus horikoshii* (NP_143114); *Pyrococcus abyssi* (NP_126714); *Pyrococcus furiosus* (NP_578787); *Ferroglobus placidus* (YP_003434610.1); *Methanocaldococcus infernus* (YP_003616088.1); *Hyperthermus butylicus* (YP_001012367.1 for IPCT and YP_001012368.1 for DIPPS); *Candidatus Korarchaeum cryptofilum* (YP_001736627.1); *Aeropyrum pernix* (NP_147991 for IPCT and NP_147993, for DIPPS); *Hydrogenivirga* sp. 128-5-R1-1 (ZP_02177680.1); *Aquifex aeolicus* (NP_213943); *Persephonella marina* (ref|YP_002731682.1); *Rubrobacter xylanophilus* (EF523341); *Thermobaculum terrenum* (YP_003322215.1); *Thermotoga* spp. (*Thermotoga maritima* (AE000512 for IPCT and DIPPS), *Thermotoga* sp. RQ2 (YP_001739339.1 for IPCT and YP_001739338.1 for DIPPS), *Thermotoga*

petrophila (YP_001244964.1 for IPCT and YP_001244965.1 for DIPPS), *Thermotoga neapolitana* (YP_002534616.1 for IPCT and YP_002534617.1 for DIPPS), *Thermotoga naphthophila* (YP_003346888.1 for IPCT and YP_003346889.1 for DIPPS)). Arrows represent identified or putative CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT, orange (■) arrow), CDP-inositol:L-*myo*-inositol 1-phosphate transferase (DIPPS, green (■) arrow), L-*myo*-inositol 1-phosphate synthase (MIPS, blue (■) arrow), glycerol 3-phosphate cytidylyltransferase (GCT, purple (■) arrow), phosphatase domain (yellow (■) arrow). The numbers indicate the percentages of amino acid identity between homologous sequences. a: hypothetical protein; b: orotidine 5'-phosphate decarboxylase; c: RNA binding protein; d: ABC transporter related protein; e: SufBD protein; f: malK-3; g: ABC type transport system permease; h: ABC transporter; i: Extracellular binding protein. **References**: 1, Martins et al. 1997; 2, Gonçalves et al. 2003; 3, Borges et al. 2010; 4, Empadinhas et al. 2001; 5, Martins and Santos 1995; 6, Rodionov et al. 2007; 7, Lamosa et al. 2006; 8, Santos et al. 2007; 9, Empadinhas et al. 2007; 10, Martins et al. 1996.

The N-terminal domain of the bifunctional IPCT/DIPPS showed high similarity with the nucleotidyltransferase family (pfam00483), comprising enzymes that catalyze the transfer of nucleotides onto phosphosugars, like UTP:glucose-1-phosphate uridylyltransferase, CTP:glucose-1-phosphate cytidylyltransferase or GTP:mannose-1-phosphate guanylyltransferase. The C-terminal domain of IPCT/DIPPS contains the conserved motif (D-G-2[X]-A-R-8[X]-G-3[X]-D-3[X]-D) that is characteristic of the CDP-alcohol phosphatidyltransferase family (pfam01066). These enzymes catalyze the displacement of CMP from CDP-alcohol (diacylglycerol is the most common alcohol) to a second alcohol with formation of a phosphodiester bond and concomitant hydrolysis of pyrophosphate. The phosphatidylinositol synthase from yeast is a well characterized representative of this family, having 23% identity with the amino acid sequence of *Archaeoglobus fulgidus* DIPP synthase. Hydropatic profiles predict the presence of three transmembrane segments in the yeast enzyme (Nikawa et al. 1987) and similar features are found in the DIPP synthase domains identified in our study: the stretches comprising residues 272-294, 334-356, and 393-435 have sufficient length

and hydrophobicity to span the membrane. The predicted membrane association of the DIPP synthase domain in the bifunctional IPCT/DIPPS provides a plausible explanation for our failure to purify the enzyme.

The reaction scheme established for DIP synthesis points to a configuration of the inositol units in disagreement with the stereochemical configuration earlier determined in DIP isolated from *Pyrococcus* spp. (van Leeuwen 1994). This intriguing inconsistency was definitely resolved in Chapter 4.

An unrooted phylogenetic tree based on the amino acid sequences of the bifunctional IPCT/DIPPS showed significant topological differences when compared with the 16S-rRNA tree (Woese et al. 1990). As more and more sequences of IPCT/DIPPS are available the comparison between them becomes very interesting and one could speculate about the evolution of this enzyme. It is interesting that two main clusters emerged in the protein-based tree (Fig. 3.5).

In the predominantly euryarchaeote cluster, the *Thermococcales* form the expected tight cluster, and the position of *Archaeoglobus fulgidus* suggests an early separation from the common ancestor of the genera *Thermococcus* and *Pyrococcus*. Curiously, the bacterial species *Aquifex aeolicus*, *Hydrogenivirga* and *Persephonella marina* also fall in this "euryarchaeote" cluster. The other cluster comprises the crenarchaeote members, *Aeropyrum pernix*, and *Hyperthermus butylicus*, but also the bacteria *Thermotoga maritima* and *Rubrobacter xylanophilus*. It is tempting to speculate that bacterial IPCT/DIPPS either have evolved from distinct ancestors, or the respective genes were acquired from archaea by lateral gene transfer. However, analysis of the GC content and the codon composition of bacterial *ipct*/*dipps* genes did not provide evidence for lateral gene transfer, since their GC content and codon usage were in line with the properties found

in the bulk genes of the respective genomes (L. G. Gonçalves, personal communication). It is noteworthy that the lack of unusual codon composition or GC content does not rule out the occurrence of lateral gene transfer, since it is conceivable that genes acquired by this means will adapt to the background genomic codon usage during evolution.

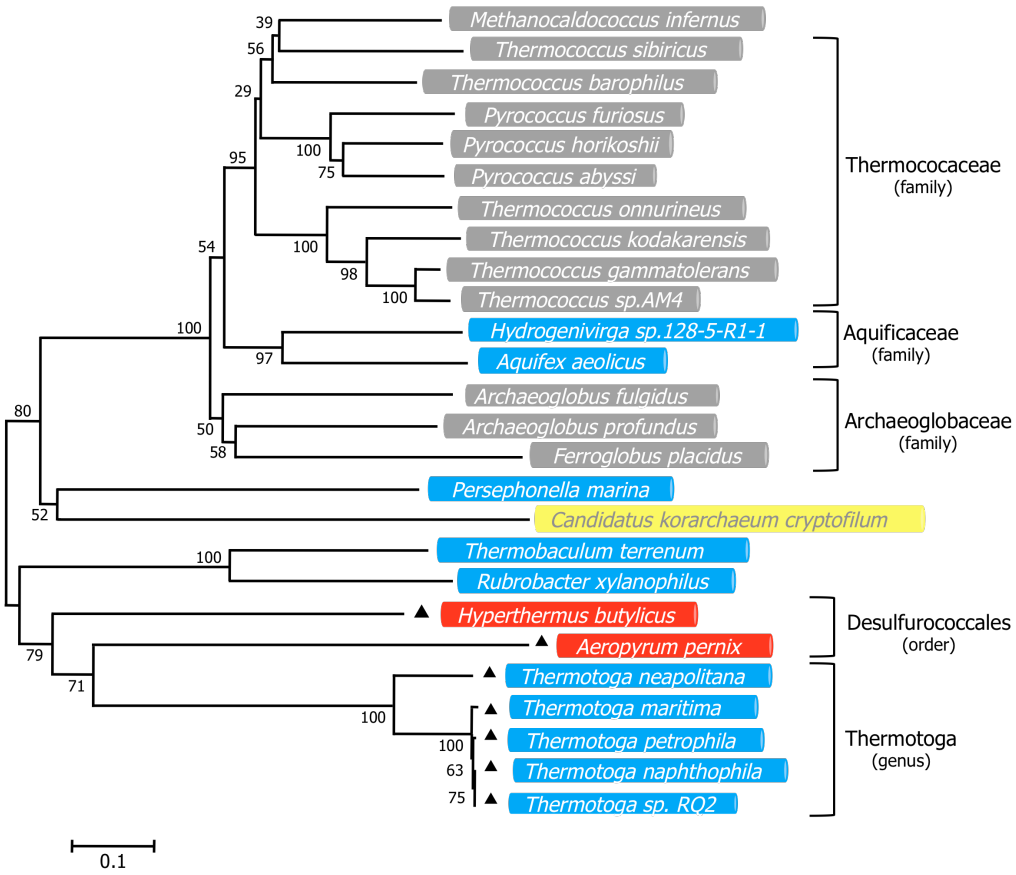


Figure 3.5. Unrooted phylogenetic tree based on available amino acid sequences of IPCT/DIPPS. The ClustalX program (Thompson et al. 1997) was used for sequence alignments. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007). The species and GenBank™

accession numbers are as follows: *Archaeoglobus fulgidus* (NCBI accession number NP_069101), *Archaeoglobus profundus* (YP_003400502.1); *Thermococcus gammatolerans* (YP_002960219.1); *Thermococcus* sp. AM4 (ZP_04879407.1); *Thermococcus kodakarensis* (YP_184692); *Thermococcus onnurineus* (YP_002307633.1); *Thermococcus sibiricus* (YP_002994415.1); *Thermococcus barophilus* (ZP_04876988.1); *Pyrococcus horikoshii* (NP_143114); *Pyrococcus abyssi* (NP_126714); *Pyrococcus furiosus* (NP_578787); *Ferroglobus placidus* (YP_003434610.1); *Methanocaldococcus infernus* (YP_003616088.1); *Hyperthermus butylicus* (YP_001012367.1 for IPCT and YP_001012368.1 for DIPPS); *Candidatus Korarchaeum cryptofilum* (YP_001736627.1); *Aeropyrum pernix* (NP_147991 for IPCT and NP_147993, for DIPPS); *Hydrogenivirga* sp. 128-5-R1-1 (ZP_02177680.1); *Aquifex aeolicus* (NP_213943); *Persephonella marina* (ref|YP_002731682.1); *Rubrobacter xylanophilus* (EF523341); *Thermobaculum terrenum* (YP_003322215.1); *Thermotoga maritima* (AE000512 for IPCT and DIPPS); *Thermotoga* sp. RQ2 (YP_001739339.1 for IPCT and YP_001739338.1 for DIPPS); *Thermotoga petrophila* (YP_001244964.1 for IPCT and YP_001244965.1 for DIPPS); *Thermotoga neapolitana* (YP_002534616.1 for IPCT and YP_002534617.1 for DIPPS); *Thermotoga naphthophila* (YP_003346888.1 for IPCT and YP_003346889.1 for DIPPS). The black triangle (▲) designates species where IPCT/DIPPS is the product of two separate genes. Grey, *Archaea*, *Euryarchaeota*; red, *Archaea*, *Crenarchaeota*; yellow, *Archaea*, *Korarchaeota*; and blue, *Bacteria*.

The cytidylyltransferase domain (N-terminus) of recombinant IPCT/DIPPS was highly specific for CTP and L-*myo*-inositol 1-phosphate, while the DIPP synthase domain showed some plasticity in regard to the alcohol activated donor. All the bifunctional IPCT/DIPPS examined (*i.e.*, from *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Aquifex aeolicus*, and *Thermococcus kodakarensis*) used both CDP-inositol and CDP-glycerol in combination with L-*myo*-inositol 1-phosphate. Therefore, either DIPP or GPIP can be end-products of this activity, depending on the specific CDP-alcohol provided (CDP-inositol or CDP-glycerol, respectively). Thus far, the compatible solute GPI has been found only in species of the genera *Archaeoglobus* and *Aquifex* (Lamosa et al. 2006). Thus, the observation that all DIPP synthases examined in this work are also able to catalyze GPI synthesis makes us wonder why GPI does not occur in members of other hyperthermophilic

genera, namely *Pyrococcus* or *Thermococcus*. A plausible explanation could be related with insufficient levels of CDP-glycerol, the substrate besides L-*myo*-inositol 1-phosphate, which is needed for the synthesis of GPI (Chapter 2). In *Aquifex aeolicus*, one of the two organisms known to accumulate GPI, the genes encoding glycerol 3-phosphate cytidylyltransferase and DIPP synthase are organized in the same operon-like structure, and probably under the control of the same promoter. In the other GPI-accumulating organism, *Archaeoglobus fulgidus*, this close spatial organization is not found, the gene encoding glycerol 3-phosphate cytidylyltransferase being located elsewhere. Nevertheless, CDP-glycerol is expected to be highly available in the cytoplasm of this organism, since CDP-glycerol is also a precursor for the synthesis of diglycerol phosphate, a prominent component of the solute pool of *Archaeoglobus fulgidus* (Gonçalves et al. 2003).

In conclusion, this work characterized the key genes and enzymes for the synthesis of DIP, the most widespread compatible solute in hyperthermophiles. Additionally, it was shown that the same genes are used in the synthesis of glycero-phospho-*myo*-inositol, a solute chemically related to DIP and also confined to (hyper)thermophiles. The gene encoding the phosphatase activity that dephosphorylates the intermediates, DIPP and GPIP, was not identified; however, the sequence information available for an uncultured organism GZfos13E1 revealed a gene encoding three remarkably expressive activities: cytidylyltransferase, DIPP synthase, and phosphatase. The hypothesis that the latter domain bears the activity that dephosphorylates DIPP appears highly plausible, and deserves further investigation. The results reported in the present work represent an important milestone towards the elucidation of the regulation of DIP biosynthesis and its role in thermoadaptation of hyperthermophiles.

Acknowledgments and work contributions

R. Ventura produced the chemically synthesized compounds used in this work. M. Gonçalves performed most of the cloning experiments of *ipct*/*dipps* genes. C. Fernandes and N. Empadinhas cloned the gene Rxyl_1212 in pET30a. We thank L. G. Gonçalves for valuable discussions and great interest in this work, and especially for the regular update of new sequences of IPCT/DIPPS. N. Borges and L. G. Gonçalves were involved in the phylogenetic analysis of IPCT/DIPPS. The NMR spectrometers at CERMAX are part of the National NMR Network and were acquired with funds from FCT and FEDER. This work was funded by Fundação para a Ciência e a Tecnologia, POCTI, Portugal, and FEDER Projects A004/2005 Action V.5.1., POCI/BIA-PRO/57263/2004 and POCI/BIA-MIC/56511/2004. M. V. Rodrigues, N. Borges, P. Lamosa, C. Fernandes and N. Empadinhas acknowledge FCT for research fellowships.

CHAPTER 4

Determination of the stereochemical configuration of di-*myo*-inositol phosphate and glycero-phospho-*myo*-inositol; structural characterization of intermediate phosphorylated metabolites

Part of this Chapter is published in:

Rodrigues MV, Borges N, Henriques M, Lamosa P, Ventura R, Fernandes C, Empadinhas N, Maycock C, da Costa MS, Santos H (2007). Bifunctional CTP:inositol 1-phosphate cytidylyltransferase/CDP-inositol:inositol 1-phosphate transferase, the key enzyme for di-*myo*-inositol phosphate synthesis in several (hyper)thermophiles. J Bacteriol **189**:5405-12.

Borges N, Gonçalves LG, Rodrigues MV, Siopa F, Ventura R, Maycock C, Lamosa P, Santos H (2006). The biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile *Archaeoglobus fulgidus* in stress adaptation. J Bacteriol **188**:8128-35.

Chapter 4 – Contents

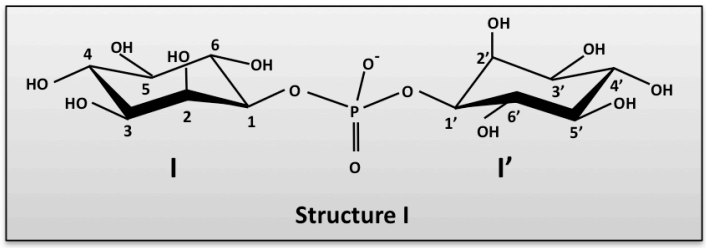
Summary

Di-*myo*-inositol phosphate (DIP) is synthesized in a two-step reaction: CTP and L-*myo*-inositol 1-phosphate originate CDP-inositol by the action of CTP:inositol 1-phosphate cytidylyltransferase (IPCT); subsequently CDP-inositol:inositol 1-phosphate transferase (DIPPS or DIPP synthase) catalyzes the reaction between CDP-inositol and L-*myo*-inositol 1-phosphate originating di-*myo*-inositol phosphate phosphate (DIPP). Finally, DIPP is further dephosphorylated by the action of a phosphatase yielding DIP. In this work, the phosphorylated intermediates, DIPP and glycero-phospho-*myo*-inositol phosphate (GPIP), were purified and characterized by NMR. Moreover, L-*myo*-inositol 1-phosphate, labeled on C_1 with carbon-13, was used to determine the stereochemical configuration of all the metabolites involved in DIP synthesis. It was shown that the two inositol moieties in DIP have different stereochemical configurations, in contradiction with previous reports. The use of the designation di-*myo*-inositol 1,3'-phosphate is recommended to facilitate tracing individual carbon atoms through metabolic pathways. The stereochemistry of the inositol group in glycero-phospho-*myo*-inositol (GPI) was also established. By using isotopic labeling and NMR we proved that in glycero-phospho-*myo*-inositol derived from both archaea and bacteria the inositol moiety has the phosphate group linked at position 3. Additionally, GPI derived from the membrane phospholipids of beef liver (Domain *Eukarya*) and GPI from bacteria gave rise to identical ^{31}P-NMR spectra, while GPI derived from archaea yielded a distinct ^{31}P-spectrum. Combining all the information available we postulate that the difference between archaeal and bacterial GPI arises from the different stereochemistry of the glycerol-phosphate moiety.

Introduction

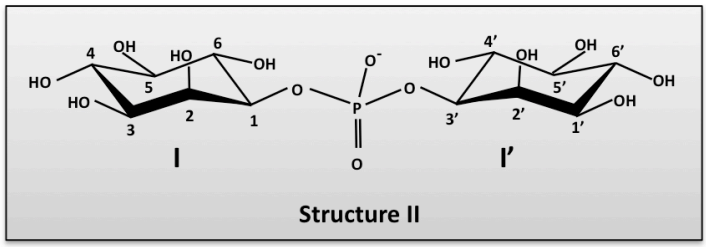
The genes and enzymes for the synthesis of di-*myo*-inositol phosphate (DIP) were recently identified (Chapter 3, Rodionov et al. 2007). In *Archaeoglobus fulgidus* DIP is synthesized in a phosphorylated form, di-*myo*-inositol phosphate phosphate (DIPP) through a two-step reaction catalyzed by a bifunctional enzyme, *i.e.*, CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase /CDP-inositol:inositol 1-phosphate transferase (IPCT/DIPPS). Subsequently, DIPP is dephosphorylated by the action of a phosphatase to yield DIP. However, the extra phosphorylation position in DIPP has not been assigned. Moreover, DIP synthesized via this reaction scheme was expected to exhibit the stereochemical configuration of Structure II (Fig. 4.1), in disagreement with the configuration (Structure I in Fig. 4.1), earlier established in *Pyrococcus woesei* (van Leeuwen et al. 1994).

The substrate specificity of the bifunctional enzyme IPCT/DIPPS was investigated: the cytidylyltransferase was specific for CTP and L-*myo*-inositol 1-phospahte; among the putative substrates examined, the DIPP synthase domain used only L-*myo*-inositol 1-phosphate as alcohol acceptor, while CDP-glycerol, CDP-L-*myo*-inositol, and CDP-D-*myo*-inositol were recognized as alcohol donors. Hence, the solute GPI is synthesized by the same enzyme, when using CDP-glycerol and L-*myo*-inositol 1-phosphate as substrates. A phosphorylated form of GPI is produced, glycero-phospho-*myo*-inositol phosphate phosphate (GPIP), which is ultimately dephosphorylated to GPI.



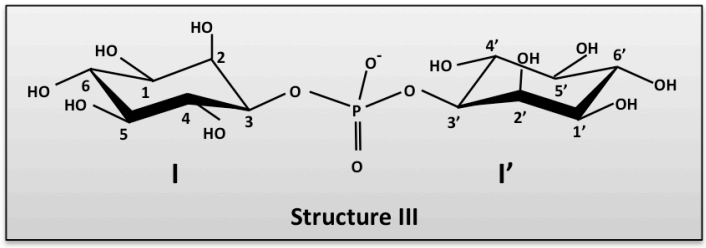
Di-*myo*-inositol 1,1'-phosphate

(Di-*myo*-inositol 3,3'-phosphate)



Di-*myo*-inositol 1,3'-phosphate

(Di-*myo*-inositol 3,1'-phosphate)



Di-*myo*-inositol 3,3'-phosphate

(Di-*myo*-inositol 1,1'-phosphate)

Figure 4.1. Structures of the three possible stereoisomers of DIP. Numbering of the inositol atoms is based on the L-numbering scheme. Hence, structure I is designated as L,L-di-*myo*-inositol 1,1'-phosphate and represents the configuration of DIP as reported by van Leeuwen et al. (1994). Likewise, structures II and III designate L,L-di-*myo*-inositol 1,3'-phosphate and L,L-di-*myo*-inositol 3,3'-phosphate, respectively. In parenthesis is the designation of DIP according to the 1D numbering scheme. Compounds I and III are enantiomers, while compound II is optically inactive.

The uncertainty about the stereochemistry of DIP prompted us to study the biochemistry of inositol. Throughout this thesis the stereospecific numbering of the inositol carbon atoms will be used based on the 1L-numbering scheme, *i.e.*, when the molecule is oriented so that the axial hydroxyl group is above the plane of the ring the carbon atom bearing the axial hydroxyl group is designated 2 and numbering of carbon atoms goes clockwise (see Fig. 4.2). The three possible configurations of di-*myo*-inositol phosphate are represented in Figure 4.1 for the sake of clarity and easy perception of the structural differences.

We resorted to ^{13}C specific labeling of suitable substrates to determine the stereochemistry of DIP and GPI. The compound L-*myo*-inositol 1-phosphate labeled with carbon-13 was synthesized from [6-^{13}C]glucose in a coupled, two-step reaction catalyzed by *Thermoproteus tenax* hexokinase and *Archaeoglobus fulgidus* L-*myo*-inositol 1-phosphate synthase. This L-*myo*-[1-^{13}C]inositol 1-phosphate was used as a substrate for recombinant or native enzymes involved in the synthesis of DIP. As a result of this strategy, the configuration of the inositol moieties of DIP and GPI was firmly established by NMR analysis of the reaction products. Additionally, the NMR structural characterization of the intermediate metabolites DIPP and GPIP was carried out.

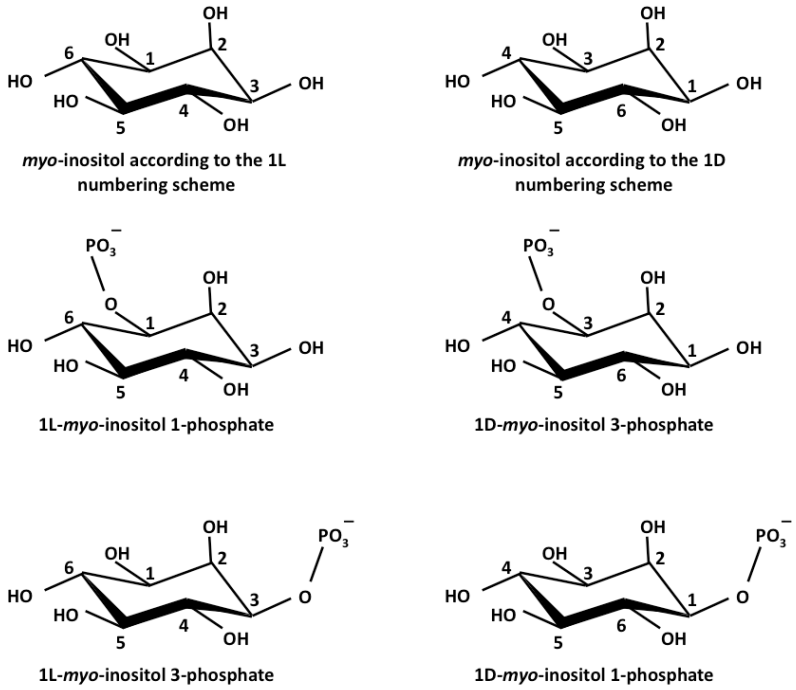


Figure 4.2. Molecular representations of *myo*-inositol and *myo*-inositol 1-phosphate showing the two possible numbering schemes. According to the 1L numbering scheme, when the axial hydroxyl group (which is always numbered as position 2) is above the plane of the ring, numbering is clockwise, while the 1D scheme is counterclockwise. If phosphorylation of *myo*-inositol takes place at position 1 according to the 1L numbering scheme we will have L-*myo*-inositol 1-phosphate; however, if phosphorylation occurs at position 3, the numbering scheme changes to 1D, to conform with the rule that "substituents must have the lowest possible locant" and the compound will be designated D-*myo*-inositol 1-phosphate instead of L-*myo*-inositol 3-phosphate, which is the same compound but with a different notation. It is noteworthy that in this context D- and L- do not necessarily mean the sense of rotation of polarized light, but indicate the numbering scheme. This change in atom numbering, although very useful to stress enantiomeric relationships, can seriously obscure metabolic relationships. To overcome this drawback, IUPAC recommends that the "lowest-locant rule" be relaxed whenever it is convenient to illustrate the tracing of individual atoms in biochemical work (Nomenclature Committee of the International Union of Biochemistry, 1989).

Materials and methods

Materials

NAD^+, CTP, ATP, CDP-glycerol, and *myo*-inositol were purchased from Sigma-Aldrich (St. Louis, USA). [6-^{13}C]glucose (99% enrichment) was obtained from Omicron Biochemicals, Inc. (South Bend, IN). L-*myo*-inositol 1-phosphate was produced from glucose 6-phosphate and NADH using recombinant *myo*-inositol phosphate synthase of *Archaeoglobus fulgidus* (Chapter 3). CDP- L-*myo*-inositol and L-*myo*-[1-^{13}C]inositol 1-phosphate were synthesized enzymatically (this work). DIP from *Pyrococcus woesei* was obtained from bitop AG (Witten, Germany). Glycero-phospho-*myo*-inositol (GPI) was partially purified from *Archaeoglobus fulgidus* cell extracts; an ethanolic extract of *Aquifex pyrophilus* containing GPI was used for comparison purposes (Lamosa et al. 2006).

Organisms and growth conditions

Pyrococcus furiosus strain 3638^T, *Archaeoglobus fulgidus* strain 7324, *Rubrobacter xylanophilus* strain 9941^T and *Thermotoga maritima* strain 3109^T were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen, Braunschweig (DSMZ), Germany. *Thermococcus kodakarensis* strain KOD1, and *Aquifex aeolicus* strain VF5 biomass were kindly provided by T. Imanaka (Kyoto University, Japan), and H. Huber (University of Regensburg, Germany), respectively. *Pyrococcus furiosus*, *Archaeoglobus fulgidus*, and *Rubrobacter xylanophilus* were cultivated as previously described by Martins et al. (1995), Chapter 2, and Empadinhas et al. (2007), respectively. *Thermotoga maritima* was grown in medium 343 as described in

the Deutsche Sammlung von Mikroorganismen und Zellkulturen. Cell growth was assessed by measuring optical density at 600 nm.

Purification of the phosphorylated intermediates

The reaction mixtures used to investigate the biosynthetic pathways of DIP and GPI (Chapter 2) containing each of the phosphorylated intermediates were subjected to anion exchange chromatography as follows: the sample was applied onto a QAE-Sephadex A-25 column previously equilibrated with 5 mM sodium bicarbonate (pH 9.8) and eluted with one bed volume of the same buffer, followed by a linear gradient of 5 mM to 1 M $NaHCO_3$. The eluted fractions were analyzed by ^{31}P-NMR and subsequently desalted by using an activated Dowex 50W-X8 resin and distilled water for elution. The active fractions were pooled, and the pH adjusted to 6 with KOH. The samples were lyophilized and dissolved in $^{2}H_2O$ prior to NMR analysis.

Cloning and expression of putative *ipct*/*dipps* genes

Experimental details about the cloning of putative *ipct*/*dipps* genes and expression of recombinant proteins are described in Chapter 3. *Escherichia coli* cells bearing the constructs for the expression of the recombinant IPCT/DIPPS were harvested, suspended in Tris-HCl (20 mM, pH 7.6) containing 10 mM $MgCl_2$ and disrupted in a French press; cell debris was removed by centrifugation (30,000 × *g*, 4ºC, 1 h). Total protein was estimated by the Bradford method (Bradford 1976).

Enzyme assays

Enzymatic assays were performed at 80ºC in a total volume of 0.5 ml and no effort was taken to avoid oxygen. All the reaction mixtures contained 50 mM MOPS pH 8.1, 5 mM $MgCl_2$, and 5 mM of substrates. The reactions were initiated by addition of 50 mg (total protein) of cell extract. After 1 h incubation, the reaction mixtures were centrifuged; 5 mM EDTA (pH 8.0) and 50 µl of ^2H_2O were added to the supernatant, and the pH was adjusted to 7.6. The reaction products were analyzed by ^{31}P-NMR spectroscopy.

The activity of IPCT/DIPPS was determined in cell extracts of *E. coli*, harboring *ipct/dipps* genes. The reaction mixtures, in a total volume of 0.5 ml, contained 20 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 5 mM of the putative substrates, and cell extract (around 12 mg of total protein) were incubated for 1 h at 80ºC (or 45ºC for the enzyme from *Rubrobacter xylanophilus*). The reaction products were analyzed by ^{31}P-NMR spectroscopy.

Determination of DIP stereochemistry

The stereochemical configuration of the inositol moieties in intermediate metabolites and end-products was determined by ^{31}P-NMR and ^{13}C-NMR spectroscopy using L-*myo*-[1-^{13}C]inositol 1-phosphate as substrate for the recombinant enzymes. The stereochemistry of DIP was also investigated in cell extracts of natural producers (*Archaeoglobus fulgidus*, *Pyrococcus furiosus*, and *Thermotoga maritima*).

Preparation of CDP-L-*myo*-inositol and L-[1-^{13}C]*myo*-inositol 1-phosphate

CDP-L-*myo*-inositol was produced by using cell extracts of *E. coli* BL21(DE3) harboring the gene (EF523341) (*Rubrobacter xylanophilus*). The cell extract was incubated at 45ºC with CTP and L-*myo*-inositol 1-phosphate. After 1 h, the reaction mixtures were treated with alkaline phosphatase (30 min at 37ºC) to dephosphorylate residual *myo*-inositol 1-phosphate. The sample was lyophilized, the residue dissolved in 1-propanol/25% ammonia (1:1.5, vol/vol), and applied to a Silica gel S column. The same solvent system was used for elution. CDP-L-*myo*-inositol eluted first and was followed by *myo*-inositol. The fractions containing CDP- L-*myo*-inositol were pooled after analysis by ^{1}H- and ^{31}P-NMR for purity assessment.

L-[1-^{13}C]*myo*-inositol 1-phosphate was produced from D-[6-^{13}C]glucose by coupling hexokinase (HK) from *Thermoproteus tenax* (Dorr et al. 2003) and *myo*-inositol 1-phosphate synthase (MIPS) from *Archaeoglobus fulgidus*. The *mips* gene was cloned into the pET23a; the pET-11c:hxk, containing the *hk* gene, was kindly provided by Dr. Bettina Siebers (University Duisburg-Essen, Germany). *E. coli* BL21(DE3) cells, bearing the constructs, were grown and induced as described above. Hexokinase and MIPS were partially purified by heat treatment; 30 min at 90ºC and 15 min at 60ºC, respectively. The synthesis of L-[1-^{13}C]*myo*-inositol 1-phosphate was performed in a reaction mixture containing 6 mM D-[6-^{13}C]glucose, 6 mM ATP, 6 mM NAD^{+}, 10 mM $MgCl_2$, 3 μg of hexokinase and 3 μg MIPS in 20 mM Tris-HCl (pH 7.6). After 1 h incubation at 70ºC, the labeled compound was purified by anionic exchange chromatography as described above for the phosphorylated intermediates.

Isolation of GPI from beef liver

Extraction and purification of GPI was performed as described in Hanahan et al. (1957) and in Brockerhoff et al. (1959) with slight modifications. A sample of beef liver (1.2 kg) was homogenized for 2 min with 95% ethanol, allowed to stand at room temperature for 4 h and centrifuged (2800 × *g*, 5 min, 25ºC). The supernatant was stored and the pellet was re-extracted with 95% ethanol:diethyl ether (3:1; vol/vol) for 3 h at room temperature. The mixture was centrifuged (2800 × *g,* 5 min, 25ºC) and the supernatant was combined with the previous one. The resulting sediment was treated a third time with the same solvent system. The three supernatants were combined and concentrated at 37°C in a rotary evaporator to remove the solvent. The aqueous solution resultant from the rotary evaporator was extracted with petroleum ether (1:1; vol/vol). The ether extract was washed with the same volume of water. The ether phase was removed by decantation. The process was repeated five to six times. Subsequently, the ether solution was concentrated using a rotary evaporator. To the concentrated ether solution 10 volumes of acetone were added. The mixture was then placed at -20°C for several hours to ensure a satisfactory separation of the phospholipids. The acetone-insoluble material (containing the majority of the phospholipids) was washed several times with acetone. The insoluble material was dissolved in ether and then precipitated with acetone. This procedure was repeated three times. The residue was treated with 95% ethanol, mixed as well as possible, and allowed to stand at room temperature for 3 to 4 h. The ethanol-soluble fraction was removed by decantation, and the residue re-extracted with 95% ethanol for an additional 4 h at room temperature. The ethanol-insoluble fraction (containing the phospholipids) was then dissolved in chloroform-methanol (4:1, vol/vol) and applied to a silica gel S column. Elution was

carried out with 1.6 volumes of chloroform-methanol (4:1, vol/vol) (Fraction 1); 2.8 volumes of chloroform-methanol (4:1, vol/vol) (Fraction 2); 2.5 volumes of chloroform-methanol (3:2, vol/vol) (Fraction 3); 6 volumes of chloroform-methanol (3:2, vol/vol) (Fraction 4); and 4 volumes of chloroform-methanol (1:4, vol/vol) (Fraction 5). According to the protocol, the phosphoinositides should be eluted with chloroform-methanol (3:2, vol/vol), hence should be present in Fraction 3. The solvent system of Fraction 3 was evaporated with a rotary evaporator at 37°C and the residue dissolved in water. To the water soluble material, methanolic sodium hydroxide was added in a ratio of 4:1 and the hydrolysis proceeded for 1.5 h at 37°C. The sample was centrifuged (39,000 × *g*, 15 min at 25°C) and the pellet was hydrolyzed again as described. Sodium hydroxide was removed from the samples with an activated Dowex 50W-X8 resin and water was used for elution. Afterwards each fraction was extracted with tetrachloromethane/chloroform/petroleum ether (1:1:1, vol/vol/vol). The aqueous layer was removed and neutralized with cyclohexylamine and the water removed by lyophilization. The residue was dissolved in water and an equal volume of acetone was added. After 2 h at 0°C, the solution was centrifuged (26,000 × *g*, 15 min at 4°C). The aqueous solution was lyophilized and analyzed by NMR spectroscopy. The NMR spectra revealed that GPI was contaminated with other compounds, so further purification was needed. The sample was applied to a QAE-Sephadex A-25 column previously equilibrated with 5 mM of sodium bicarbonate (pH 9.8) and elution was carried out with a step gradient of the same buffer. Sodium bicarbonate (pH 9.8) was used for elution in the following concentrations (4 bed volumes of each): 10 mM, 100 mM, 200 mM, 400 mM, 600 mM and 1 M. GPI was detected in fractions eluted with 10 mM and 100 mM of sodium bicarbonate (pH 9.8). The active fractions were desalted with an activated Dowex 50W-X8 using water for elution. The pH was raised to 6 with KOH.

This procedure was repeated until pure fractions of GPI were obtained. The samples were lyophilized and dissolved in $^{2}H_2O$ prior to NMR analysis. Approximately 39 mg of pure GPI were obtained.

NMR spectroscopy

^{1}H-, ^{31}P-, and ^{13}C-NMR spectra were run on a Bruker DRX500 (Bruker, Rheinstetten, Germany). One dimensional and two-dimensional spectra were recorded using standard Bruker pulse programs following a strategy described previously (Lamosa et al. 1998, Silva et al. 1999). ^{13}C-^{1}H correlation spectra were recorded using a delay of 3.44 ms for the evolution of the scalar couplings, while a delay of 62.5 ms was used for the ^{31}P-^{1}H correlation spectra. Assignment of resonances to DIP, GPI, CDP-glycerol, and CDP-inositol was confirmed by addition of small amounts of the authentic compounds. ^{1}H, ^{13}C and ^{31}P chemical shifts are relative to 4,4-dimethyl 4-silapentane sodium sulfonate (DSS) designated at 0 ppm, external methanol designated at 49.3 ppm and external 85% H_3PO_4 designated at 0 ppm, respectively.

Spiking experiments were performed as follows: a ^{31}P-NMR decoupled spectrum of GPI partially purified from an *Archaeoglobus fulgidus* extract was acquired; to this sample an aliquot of GPI from an *Aquifex pyrophilus* ethanolic extract was added, and a new ^{31}P-NMR spectrum was acquired; finally, an aliquot of GPI purified from beef liver was added and a ^{31}P-NMR spectrum was acquired.

GPIP produced with the recombinant IPCT/DIPPS from *Archaeoglobus fulgidus* was dephosphorylated with alkaline phosphatase and the sample analyzed by ^{31}P-NMR. This sample was spiked successively with GPI purified

from beef liver and GPI produced with recombinant IPCT/DIPPS from *Aquifex aeolicus*.

We found that the assignment of the ^{13}C-resonances of *myo*-inositol reported in the literature is incorrect (Breitmaier and Voelter 1989). Therefore, the assignment of the *myo*-inositol resonances was performed by running HMQC and COSY spectra. The carbon chemical shift values of *myo*-inositol are: C1/C3, 72.54 ppm; C2, 72.54 ppm; C4/C6, 71.50 ppm; and C5, 74.71 ppm.

Results

Determination of the stereochemical configuration of DIP

In Chapters 2 and 3 we demonstrated that the synthesis of DIP occurs via a phosphorylated intermediate, DIPP. Moreover, the proposed pathway would produce DIP with a stereochemical configuration different from that reported in the literature (van Leeuwen et al. 1994). Hence, we deemed it important to determine the stereo-configuration of the inositol moieties in DIP. For this purpose, L-*myo*-[1-^{13}C]inositol 1-phosphate was produced, purified, and analysed by ^{1}H-, ^{13}C- and ^{31}P-NMR. The C_1 resonance of L-*myo*-inositol 1-phosphate at 74.70 ppm was split due to coupling with phosphorus ($^{2}J_{C\text{-}P}$ = 5.2 Hz) (Fig. 4.3c) and, as expected, the same coupling constant was apparent in the phosphorus spectrum of the same sample (not shown). These results together with the information that only the L-stereoisomer is produced by the *Archaeoglobus fulgidus* L-*myo*-inositol 1-phosphate synthase (Chen et al. 2000), led to the firm identification of the product as L-*myo*-[1-^{13}C]inositol 1-phosphate.

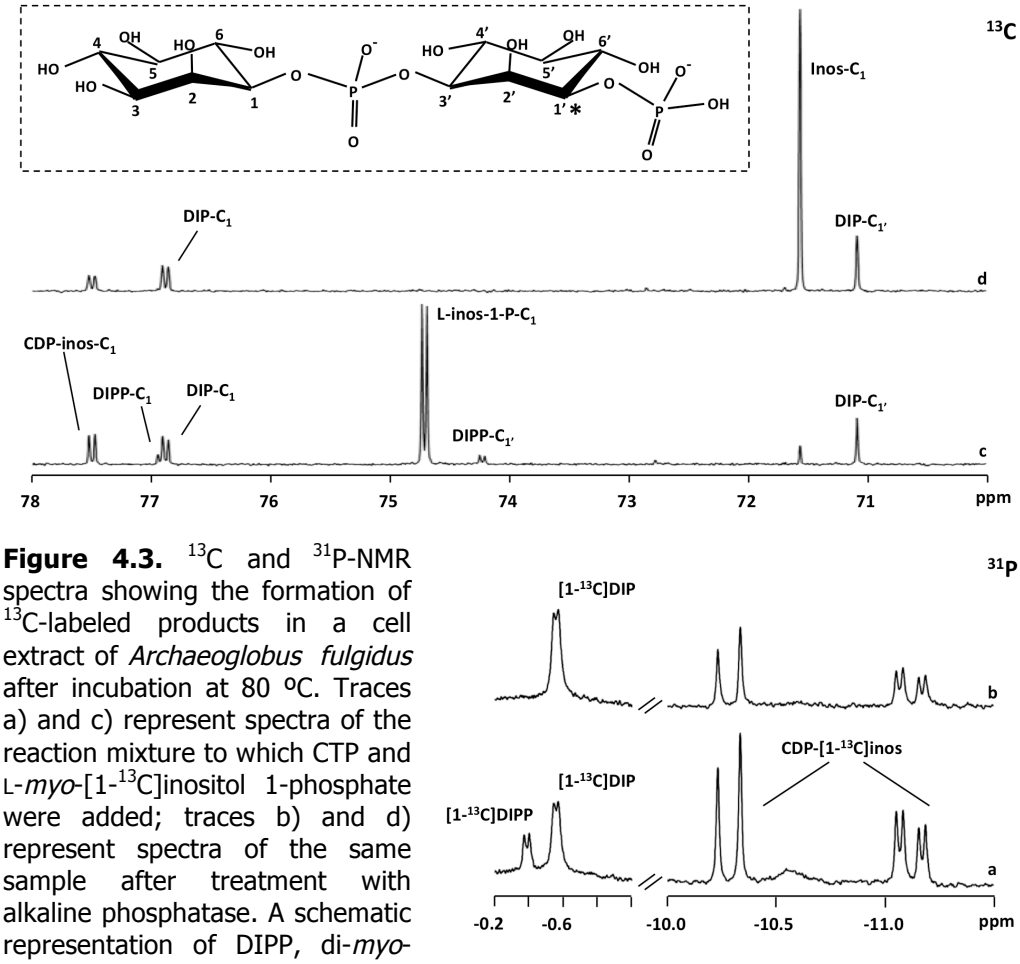


Figure 4.3. ^{13}C and ^{31}P-NMR spectra showing the formation of ^{13}C-labeled products in a cell extract of *Archaeoglobus fulgidus* after incubation at 80 ºC. Traces a) and c) represent spectra of the reaction mixture to which CTP and L-*myo*-[1-^{13}C]inositol 1-phosphate were added; traces b) and d) represent spectra of the same sample after treatment with alkaline phosphatase. A schematic representation of DIPP, di-*myo*-inositol 1,3'-phosphate 1'-phosphate is also shown in the dashed box. Spectra were run at a temperature of 25ºC. Symbols: inos-C_1, L-inos-1-P-C_1, DIP-C_1, DIPP-C_1, CDP-inos-C_1, designate resonances due to carbon 1 of the inositol moiety in *myo*-inositol, L-*myo*-inositol 1-phosphate, DIP, DIPP, and CDP-L-*myo*-inositol, respectively. The splitting of the carbon (or phosphorus) resonances denotes two-bond couplings with phosphorus (or carbon) nuclei.

The use of this labeled compound as substrate for DIP synthesis in cell extracts of the natural producers allowed us to determine the stereochemical configuration of inositol moieties in intermediates and end-products, simply by observing ^{31}P-^{13}C coupling patterns (Fig. 4.4).

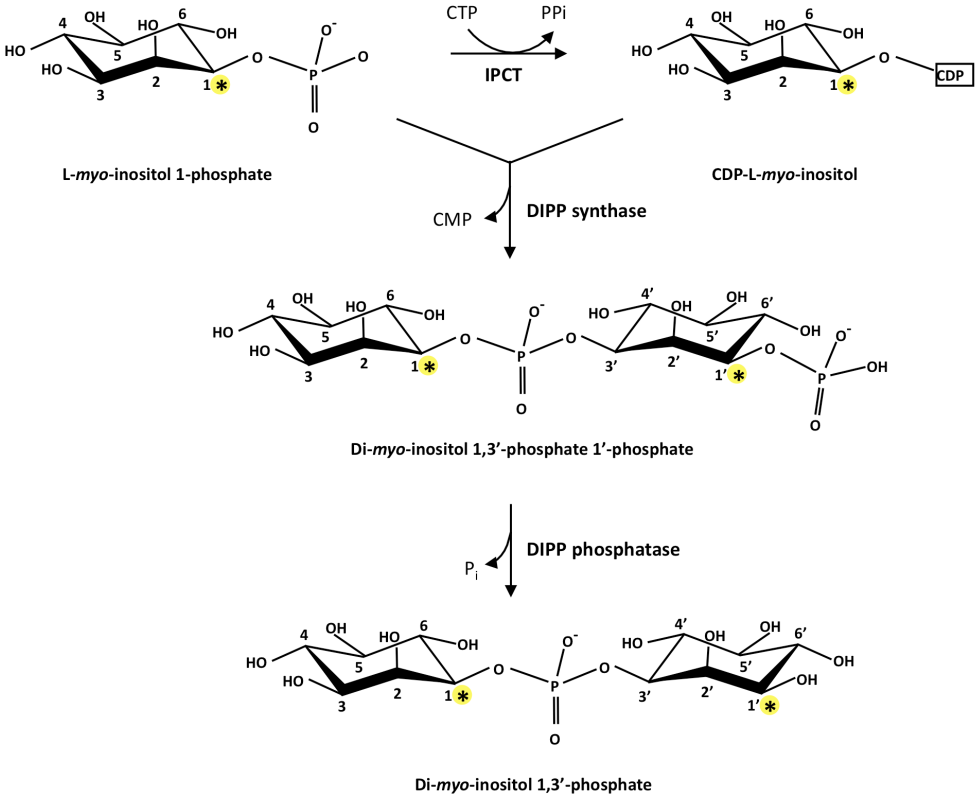


Figure 4.4. Pathway for DIP biosynthesis using L-*myo*-[1-^{13}C]inositol 1-phosphate as a labeled precursor. The asterisk highlights the positions of labeled carbon atoms in the several intermediate metabolites and end-products as established in this work from ^{13}C and ^{31}P-NMR analysis of reaction mixtures containing the bifunctional enzyme IPCT/DIPPS (see text for details). Dashes are used to designate the carbon atoms in the phosphorylated inositol moiety.

Briefly, the C_1 labeled with carbon-13 will originate splitting of the phosphorus NMR signal if the two atoms (carbon and phosphorus) are less than 2 bonds apart. As a result, if DIP was produced in the configuration of the inositol moieties displayed in Structure I (Fig. 4.1) its phosphorus signal

(in a spectrum with proton decoupling) should be a triplet, whereas Structures II and III would give rise to a doublet and a singlet, respectively.

The enzymatic assays were performed in *Archaeoglobus fulgidus* cell free extract using CTP and L-*myo*-[1-^{13}C]inositol 1-phosphate, and analyzed by NMR. The ^{31}P-NMR spectrum shows two doublets at -0.38 ppm ($J_{P\text{-}C}$ = 6.0 Hz), and at -0.56 ppm ($J_{P\text{-}C}$ = 5.6 Hz), corresponding to the phosphodiester resonances of DIPP and DIP, respectively (Fig. 4.3a), meaning that both molecules have configurations identical to Structure II in Figure 4.1. This conclusion is further supported by the ^{13}C-NMR spectra (Fig. 4.3c) where the resonances due to the two labeled carbons of DIPP (C_1 and $C_{1'}$) are clearly visible at 76.92 ppm and 74.22 ppm, respectively. Treatment of this sample with alkaline phosphatase led to the disappearance of these resonances and concomitant increase in the intensity of the resonances due to carbons 1 and 1′ of DIP (Fig. 4.3b,d). The identity of DIP was also confirmed by spiking with the pure natural compound.

The same strategy allowed us to determine the configuration of the inositol moiety in GPI, whose synthesis also proceeds via a phosphorylated intermediate, GPIP. Incubation of *Archaeoglobus fulgidus* extract with CDP-glycerol and L-*myo*-[1-^{13}C]inositol 1-phosphate (the substrates for GPI synthesis) did not originate splitting of the respective phosphorus resonances, either in the signals due to GPI or GPIP, thus establishing that the phosphate group is linked to carbon 3 of the inositol moiety in both these compounds (data not shown).

The stereochemical configurations of DIPP and GPIP were also confirmed using crude extracts of *E. coli* producing the recombinant enzymes. The results with the bifunctional, recombinant enzyme of *Pyrococcus furiosus* are illustrated in Figures 4.5 and 4.6. In Figure 4.5 the doublets assigned to C_1 and $C_{1'}$ in DIPP are clearly seen while in Figure 4.6 the doublet assigned to $C_{3'}$

of GPIP is visible. Furthermore, the ^{31}P-NMR spectra of Figure 4.6 shows that the phosphorus signal of GPIP is not split, proving that the carbon-13 labeled carbon atom of L-*myo*-inositol 1-phosphate is not involved in the ester bond with the phosphorus. Similar results were obtained in *E. coli* extracts producing the recombinant enzymes from *Archaeoglobus fulgidus*, *Thermococcus kodakarensis*, and *Aquifex aeolicus*.

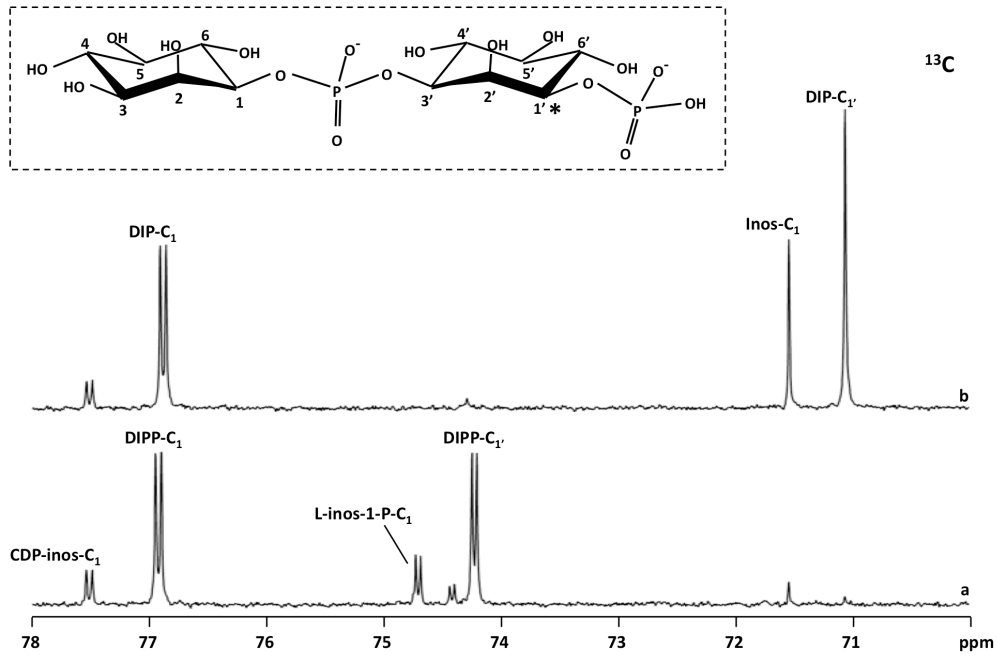


Figure 4.5. ^{13}C-NMR spectra showing the formation of ^{13}C-labeled products resulting from incubation at 80ºC of an extract of *E. coli*, bearing the *ipct/dipps* gene from *Pyrococcus furiosus*, with CTP and L-*myo*-[1-^{13}C]inositol 1-phosphate, before (a), and after (b) treatment of the final reaction mixture with alkaline phosphatase. The dashed box shows a schematic representation of di-*myo*-inositol 1,3'-phosphate 1'-phosphate. Spectra were run at a temperature of 25ºC. Symbols: inos-C_1, L-inos-1P-C_1, DIP-C_1, DIPP-C_1, CDP-inos-C_1, designate resonances due to carbon 1 of the inositol moiety in *myo*-inositol, L-*myo*-inositol 1-phosphate, DIP, DIPP, and CDP- L-*myo*-inositol. The splitting of the carbon (or phosphorus) resonances denotes two-bond couplings with phosphorus (or carbon) nuclei.

In summary, DIP comprises two stereochemically distinct inositol moieties, *i.e.*, in all cases examined, the phosphate bridges between position 1 of the inositol moiety I and position 3′ of the inositol moiety I′ (using the L-stereospecific numbering convention (Nomenclature Committee of the International Union of Biochemistry, 1989). Therefore, the correct structure of DIP is represented by Structure II in Figure 4.1, and the proposed name is di-*myo*-inositol 1,3′-phosphate. As for GPI, the phosphate is linked to the inositol moiety I′ (using the L-stereospecific numbering convention).

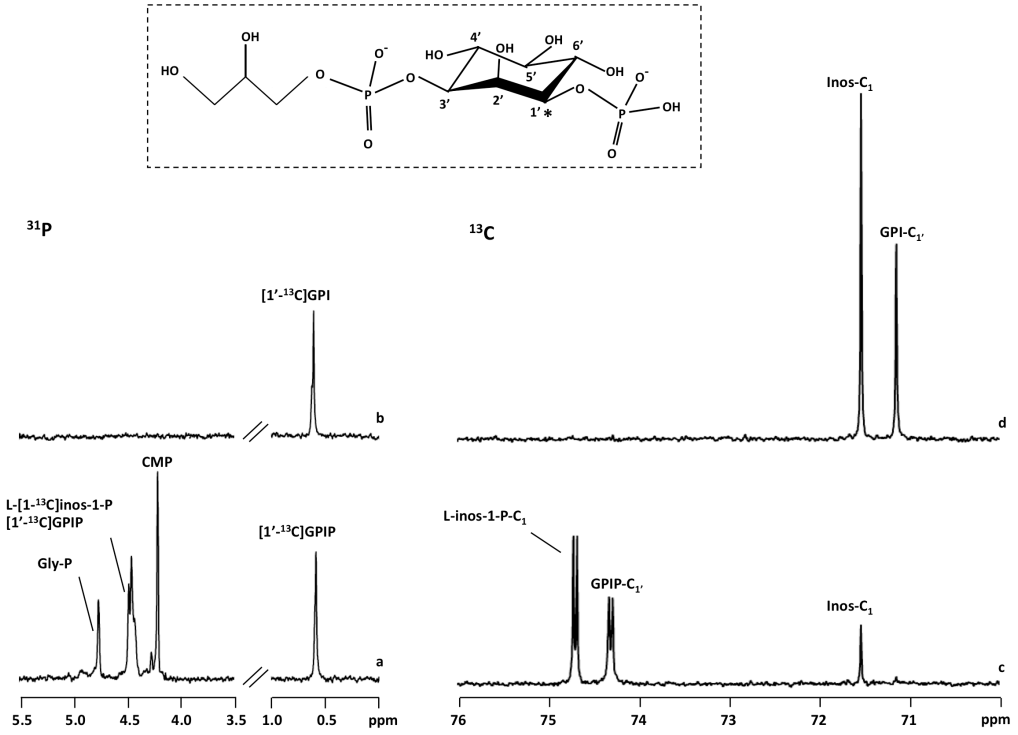


Figure 4.6. ^{31}P- and ^{13}C-NMR spectra showing the formation of ^{13}C-labeled products resulting from incubation at 80ºC of an extract of *E. coli*, bearing the *ipct/dipps* gene from *Pyrococcus furiosus*, with CDP-glycerol and L-*myo*-[1-^{13}C]inositol 1-phosphate, before (a) and (c), and after treatment of the final reaction mixture with alkaline phosphatase (b) and (d). The dashed box shows a schematic representation of glycero-phospho-*myo*-inositol. Spectra were run at a temperature of 25ºC. Symbols: inos-C_1, L-inos-1P-C_1, GPIP-$C_{1'}$, GPI-$C_{1'}$, designate resonances due to carbon 1 of the

inositol moiety in *myo*-inositol, L-*myo*-inositol 1-phosphate, GPIP and GPI; Gly-P, D,L-glycerol 3-phosphate. The splitting of the carbon resonances denotes two-bond couplings with phosphorus nuclei.

NMR characterization of the phosphorylated intermediates

The proton correlation spectrum (COSY) of the purified DIPP (unlabeled) allowed following sequentially all resonances of the two distinct moieties in the compound. Proton-proton coupling patterns revealed these moieties as *myo*-inositol residues. The combination of the ^{13}C-^{1}H HMQC spectrum and the COSY spectrum enabled the assignment of all proton and carbon signals (Table 4.1). The results revealed that the inositol moieties in DIP have different stereoconfigurations and that the extra phosphate group in DIPP is linked at position 1' (see Figure 4.4). The signals belonging to one of the inositol moieties overlapped those of DIP. On the proton dimension the signals of the other inositol moiety displayed shifts towards lower field except for proton 1 which shifted towards high field (Fig. 4.7). This led us to suspect that the additional phosphate group was linked at position 1 of inositol. In comparing the two moieties, the signal bearing the largest low field shift was that of proton 3, while protons 2 and 4 showed moderate shifts, and protons 5 and 6 displayed even smaller ones. The proton decoupled ^{31}P-NMR spectrum showed a resonance in the phosphomonoester region and another in the phosphodiester region (4.78 and 0.38 ppm, respectively). Irradiation of the phosphodiester resonance resulted in the collapsing of the multiplet due to proton 1 and 3' in the two inositol moieties, indicating that these moieties are linked via a phosphodiester bond. When the irradiation was performed at the monoester frequency, at 4.78 ppm, we observed changes only in the multiplet structure due to proton at position 1' of the moiety originating shifted resonances, thus confirming the view that this moiety was phosphorylated and

that the additional phosphate group was linked at position 1'. Therefore, the phosphorylated intermediate is di-*myo*-inositol 1,3'-phosphate 1'-phosphate.

Table 4.1. NMR parameters of di-*myo*-inositol 1,3'-phosphate 1'-phosphate (DIPP) and glycero(X)-phospho(3')-*myo*-inositol 1'-phosphate (GPIP).

Moiety	Chemical shift[a] (ppm)			
	DIPP		GPIP	
	^{13}C	^{1}H	^{13}C	^{1}H
Inositol				
C1	76.9	4.03		
C2	71.6	4.28		
C3	71.0	3.55		
C4	72.6	3.64		
C5	74.3	3.33		
C6	71.9	3.76		
Inositol-P				
C1'	74.2	3.92	74.7	3.92
C2'	71.2	4.41	71.3	4.39
C3'	76.6	4.08	76.5	4.03
C4'	72.5	3.77	72.3	3.78
C5'	74.3	3.39	74.5	3.39
C6'	71.8	3.78	71.3	3.78
Glycerol				
C1			67.3	4.00; 3.92
C2			71.2	3.90
C3			62.4	3.73; 3.67

[a] Referenced to TSPSA designated at 0.0 ppm. ^{31}P-NMR chemical shifts of the phosphodiester resonances of DIPP and GPIP were -0.38 and 0.63 ppm, respectively; the phosphomonoester resonances were at 4.78 and 4.81 ppm, respectively. The pH of the samples was 9.0. Inositol, *myo*-inositol; inositol-P, L-*myo*-inositol 1-phosphate; glycerol-P, glycerol phosphate.

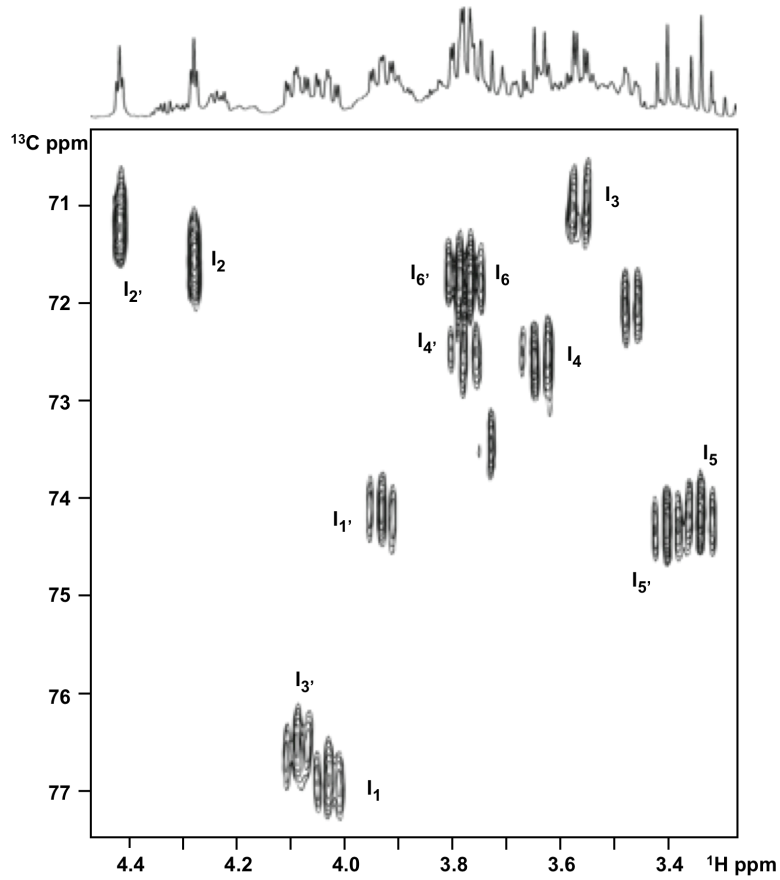


Figure 4.7. ^{13}C-^{1}H HMQC of di-*myo*-inositol 1,3'-phosphate 1'-phosphate (DIPP) partially purified from the reaction mixtures. Peaks due to the inositol and inositol 1'-phosphate moieties are labeled with I and I', respectively. The subscripts represent proton numbering. Cross peaks correspond to connectivities between proton and carbon atoms covalently bonded.

The same overall strategy was used to determine the structure of GPIP, the intermediate in the synthesis of GPI. Partial purification of the compound led to a proton decoupled ^{31}P-NMR spectrum bearing only two signals in the phosphomonoester and diester regions (4.81 and 0.63 ppm, respectively). When proton decoupling was removed, the phosphorus signal at 0.63 ppm revealed the complex structure expected for the phosphodiester

signal of GPI, while the monoester signal at 4.81 ppm split into a doublet, meaning that it is coupled to a single proton. Therefore, the additional phosphate group could only be linked to the inositol moiety. Accordingly, the ^{13}C-^{1}H HMQC spectrum presented a group of six resonances that completely overlaped those of the phosphorylated inositol moiety of DIPP. The signals belonging to the glycerol moiety were assigned by comparison with those from GPI. To confirm the assignment, proton spectra were also acquired with selective irradiation of each of the phosphorus signals. Irradiation of the diester signal at 0.63 ppm caused the collapse of multiplicities in the signals at position 3' of the inositol and position 1 or 3 of the glycerol, while irradiation at the frequency corresponding to 4.81 ppm only affected the proton signal assigned to position 1' of inositol. We conclude that the additional phosphate group in GPIP is located at position 1' of the inositol moiety, and that the intermediate is glycero(X)-phospho(3')-*myo*-inositol 1'-phosphate.

GPI configuration in *Eukarya*, *Archaea* and *Bacteria*

To test the thermostabilizing properties of GPI on model enzymes, like malate dehydrogenase or lactate dehydrogenase, high amounts of the pure solute are needed. GPI can be obtained from its natural hyperthermophilic producers, *Archaeoglobus fulgidus* and *Aquifex* spp.. The purification of GPI from *Archaeoglobus fulgidus* extract is difficult, due to the presence of similar solutes such as diglycerol phosphate and DIP that also accumulate in this archaeon. On the other hand, we failed to obtain the required amounts of *Aquifex* spp. biomass. Alternatively, the chemical synthesis of GPI was difficult and led to low product yield (R. Ventura and C. Maycock, unpublished results). Thus, we invested in the isolation of GPI from the membrane

phospholipids of beef liver. The process was time consuming and the yield of pure GPI obtained was around 33 mg/kg liver. The purity of the compound was evaluated by ^{1}H-, ^{13}C- and ^{31}P-NMR and the results confirmed the presence of glycero-phospho-*myo*-inositol.

Using NMR spectroscopy it is possible to identify an unknown compound by the addition of an aliquot of the suspected compound. If the unknown and the added compounds are identical no extra resonances will be apparent, and the intensity of the resonances will increase. ^{31}P-NMR spiking experiments showed that the GPI extracted from the phosphoinositides of the beef liver (*Eukarya*) is similar to the GPI accumulated in *Aquifex pyrophilus* (*Bacteria*), but different from the GPI accumulated in *Archaeoglobus fulgidus* (*Archaea*) (Fig. 4.8a). Curiously, the recombinant IPCT/DIPPS from *Archaeoglobus fulgidus* and *Aquifex aeolicus* produce equivalent forms of GPI when commercial CDP-glycerol (racemic) and D,L-*myo*-inositol 1-phosphate are used as substrates (Fig. 4.8b). It is important to recall that the DIPP synthase is specific for L-*myo*-inositol 1-phosphate (Chapter3). Therefore, these results indicate that the stereochemistry of GPI produced with the recombinant enzymes is determined by the stereochemistry of CDP-glycerol.

It would be interesting to investigate the formation of GPI in cell extracts of the natural producers, *Archaeoglobus fulgidus* or *Aquifex pyrophilus*. For that matter, we performed reactions using cell free extracts of *Archaeoglobus fulgidus*. The extracts were incubated with CTP, L-*myo*-inositol 1-phosphate and *rac*- glycerol 1-phosphate or *sn*-glycerol 3-phosphate and the reaction products were analyzed by ^{31}P-NMR. Unfortunately, due to the high affinity of the enzymes diglycerol phosphate synthase and IPCT/DIPPS for those substrates, the major products were DGP and DIP. GPI was formed in small amounts and the poor signal to noise ratio of the GPI resonance did

not allow us to draw any conclusion about the stereoconfiguration of that solute (data not shown).

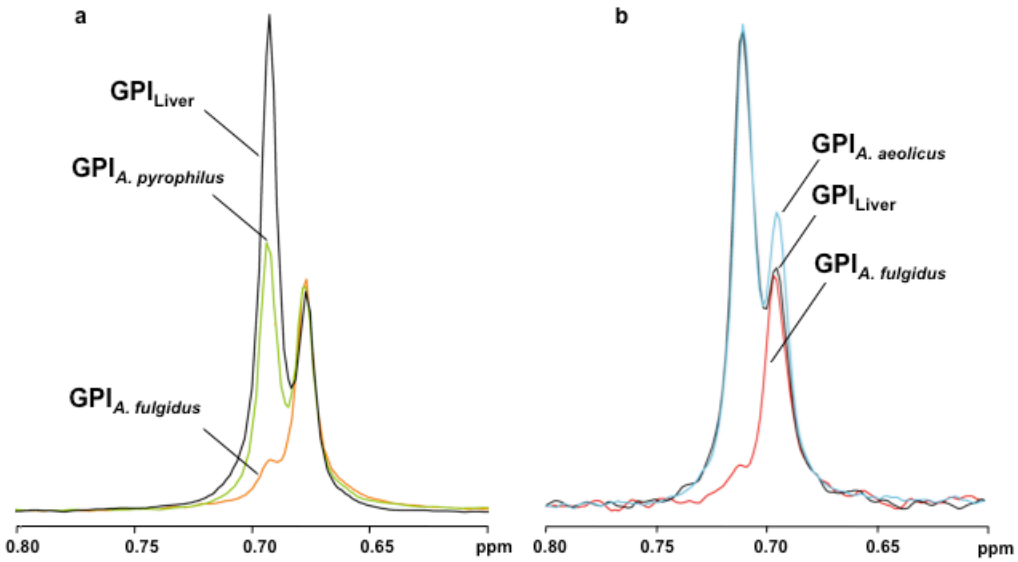


Figure 4.8. ^{31}P-NMR spectra of a sample of GPI partially purified from *Archaeoglobus fulgidus* extracts (panel **a**, orange spectrum) and of a sample of GPI produced with the recombinant DIPPS from *Archaeoglobus fulgidus* (panel **b**, red spectrum). An aliquot of an ethanolic extract of *Aquifex pyrophilus* containing GPI was added (panel a, green trace); subsequently, an aliquot of GPI purified from beef liver was also added (panel **a**, black trace). In panel b, the black trace refers to spiking of the sample with an aliquot of GPI purified from beef liver, and the blue trace was acquired after subsequent addition of GPI produced with the recombinant DIPPS from *Aquifex aeolicus* .

Discussion

The reaction scheme established for DIP synthesis points to a configuration of the inositol units as displayed in Structure II (Fig. 4.1), in disagreement with the stereochemical configuration earlier determined in DIP isolated from

Pyrococcus spp. (van Leeuwen 1994). This intriguing inconsistency was definitely resolved in this thesis work. Using labeled L-*myo*-[1-^{13}C]inositol 1-phosphate as a substrate either for the recombinant bifunctional IPCT/DIPPS and for the enzyme present in cell extracts of the natural producers, the ^{13}C-label was traced through the biosynthetic pathway. We showed that the two inositol moieties in DIP have the configurations represented by Structure II (Fig. 4.1) in all the organisms examined (*Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Aquifex aeolicus*, *Thermococcus kodakarensis*, and *Thermotoga maritima*). We propose that henceforth, the designation di-*myo*-inositol 1,3'-phosphate should be used to highlight the fact that the phosphate group is linked to stereochemically distinct carbon atoms of the two inositol groups.

We showed that GPI accumulating in the bacterium *Aquifex pyrophilus* is different from the GPI synthesized by the archaeon *Archaeoglobus fulgidus* (Fig. 4.8). GPI is synthesized by the bifunctional IPCT/DIPPS, due to the low specificity of CDP-inositol:inositol 1-phosphate transferase, which is able to use CDP-glycerol as well as CDP-inositol, and therefore produces GPI. This has been proven to occur in several hyperthermophilic archaea and bacteria, including organisms that do not accumulate GPI, such as *Thermotoga maritima*, *Pyrococcus furiosus*, and *Thermococcus kodakarensis* (Chapter 3). The use of carbon-13 labeled L-*myo*-inositol 1-phosphate allowed us to confirm that the inositol ring has the same configuration in GPI regardless the bacterial or archaeal origin of DIPP synthase. Therefore, we conclude that the difference in configuration relies on the glycerol moiety (Fig. 4.9). In fact, different enantiomers of glycerol phosphate are used in the glycerophosphate backbones of membrane phospholipids of *Bacteria* and *Archaea*. While *sn*-glycerol 1-phosphate is found in the glycerophospholipids of members of the domain *Archaea*, *sn*-glycerol 3-phosphate is found in glycerophospholipids of

Bacteria and *Eukarya* (Daiyasu et al. 2002). Both enantiomeric forms are derived from dihydroxyacetonephosphate (DHAP) and the chirality of glycerol phosphate is dictated by the stereoselectivity of the dehydrogenase of each organism. The enzyme *sn*-glycerol 1-phosphate dehydrogenase (G1PDH) produces *sn*-glycerol 1-phosphate, while the other enantiomer, *sn*-glycerol 3-phosphate, is produced by *sn*-glycerol 3-phosphate dehydrogenase (G3PDH). Until recently it was believed that G1PDH was restricted to *Archaea*, whereas G3PDH was confined to *Bacteria* and *Eukarya* (Daiyasu et al. 2002). This view seemed to correlate well with the occurrence of different stereoforms of GPI in *Archaea* and in *Bacteria*/*Eukarya*. However, it is known now that although G3PDH is widespread in *Bacteria*, it is also present in the genomes of the archaea *Archaeoglobus fulgidus* and *Methanothermobacter thermoautotrophicus* (Sakasegawa et al. 2004, Peretó et al. 2004). Likewise, G1PDH is present not only in archaeal species, but is also present in bacteria, like *Bacillus subtilis*, *Streptomyces* spp., *Streptococcus pneumoniae*, and *Thermotoga maritima* and (Peretó et al. 2004, Guldan et al. 2008). Therefore, a straightforward explanation for our findings with respect to the different configurations of GPI in *Archaea* and *Bacteria* is not available.

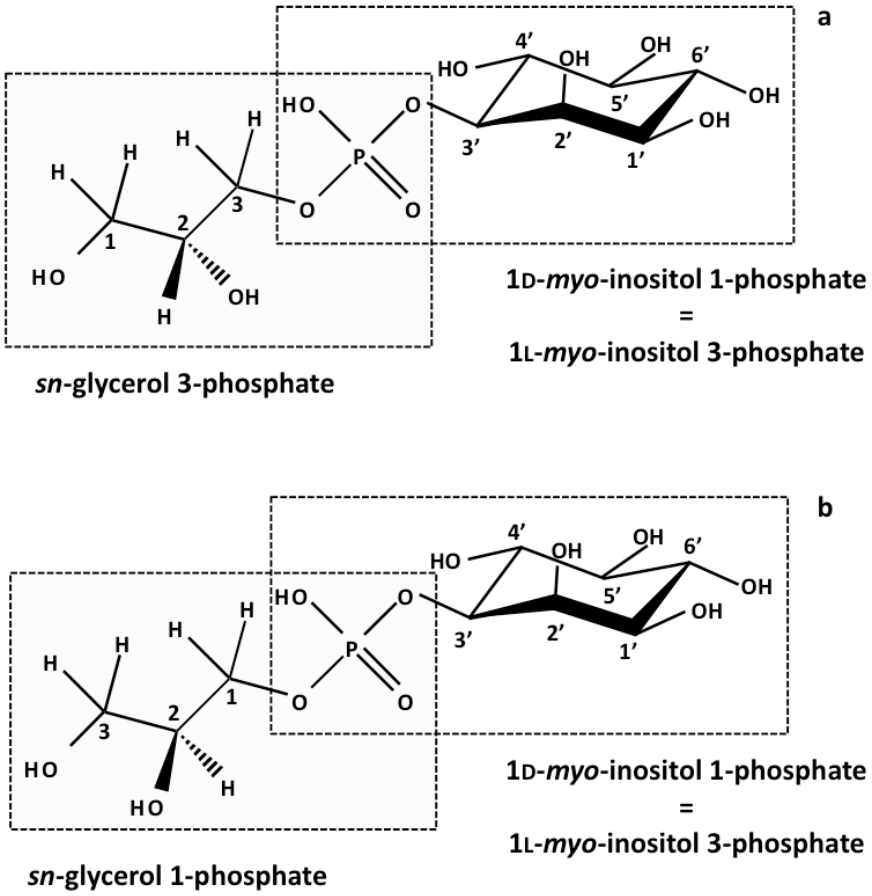


Figure 4.9. Structures of the two stereoisomers of GPI. We propose that structure (a) is present in the bacteria *Aquifex* spp. and in beef liver (*Eukarya*) and structure (b) is present in the archaeon *Archaeoglobus fulgidus*. The carbon atoms of glycerol are numbered stereospecifically, *i.e.*, the carbon atom that appears on the top of the Fischer projection that has the hydroxyl group at carbon-2 pointing to the left is designated C-1 (The nomenclature of lipids, 1978).

Curiously, the form of GPI that is produced by the recombinant IPCT/DIPPS from *Aquifex aeolicus* and *Archaeoglobus fulgidus* is equivalent to the natural form of GPI accumulating in *Archaeoglobus fulgidus*. This is an intriguing result given that the commercial CDP-glycerol used as substrate for

the reaction is a racemic mixture, resultant from the chemical reaction of coupling CMP and glycerol phosphate (Sigma-Aldrich communication). Hence, both forms of CDP-glycerol were available as substrates for the enzymes. However, all the recombinant enzymes examined, either from bacterial or archaeal origin, seem to prefer the same isomer of CDP-glycerol, *i.e.*, the compound that carries the *sn*-glycerol 1-phosphate backbone. On the other hand, when a cell extract of *Archaeoglobus fulgidus* was supplied with CDP-glycerol and L-*myo*-inositol 1-phosphate, similar amounts of the two isomers of GPI were formed (results not shown). This could suggest a deficient folding of the recombinant enzyme which led to a different stereospecificity. If this hypothesis holds true, then the GPI conformers actually accumulating in the native organisms might be determined by the availability of stereoforms of CDP-glycerol.

The glycerol moiety of GPI is provided by CDP-glycerol, a metabolite produced by the enzyme glycerol 3-phosphate cytidylyltransferase. To our knowledge this enzyme has been studied only in bacteria of the genera *Bacillus* and *Staphylococcus* and in the yeast *Saccharomyces cerevisiae*. According to the literature the enzyme uses CTP and *sn*-glycerol 3-phosphate as substrates, as expected since they belong to the domains *Bacteria* and *Eukarya*, where *sn*-glycerol 3-phosphate is the predominant isomer of glycerol phosphate (Park et al. 1993; 1997, Bandurina et al. 2003, Pattridge et al. 2003, Yim et al. 2001, Fong et al. 2006). Hence, it is essential to determine the configuration of CDP-glycerol produced by *Archaeoglobus fulgidus*.

Acknowledgments and work contributions

M. Gonçalves performed most of the cloning experiments of *ipct*/*dipps* genes. C. Fernandes and N. Empadinhas cloned the gene Rxyl_1212 in pET30a. N. Borges was involved in the phylogenetic analysis of IPCT/DIPPS. P. Lamosa was responsible for the rational behind the NMR experiments with ^{13}C labeled compounds and for the acquisition of the NMR spectra that led to the identification of the stereochemistry of DIP and MDIP. The NMR spectrometers at CERMAX are part of the National NMR Network and were acquired with funds from FCT and FEDER. This work was funded by Fundação para a Ciência e a Tecnologia, POCTI Portugal, and FEDER (Projects A004/2005 Action V.5.1., POCI/BIA-PRO/57263/2004, POCI/BIA-MIC/56511/2004, PTDC/BIA-MIC/71146/2006 and PTDC/BIO70806/2006. M. V. Rodrigues, N. Borges, P. Lamosa, C. Fernandes and N. Empadinhas acknowledge FCT for research fellowships. We thank Dr. David Turner for helpful discussions.

CHAPTER 5

A unique β-mannosyltransferase from *Thermotoga maritima* that is involved in the synthesis of mannosyl-di-*myo*-inositol phosphate

Part of this Chapter is published in:

Rodrigues MV, Borges N, Almeida CP, Lamosa P, Santos H (2009). A unique β-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-*myo*-inositol phosphate as the mannosyl acceptor. J Bacteriol **191**:6105-15.

Chapter 5 – Contents

Summary

In addition to di-*myo*-inositol 1,3'-phosphate (DIP), a compatible solute widespread in hyperthermophiles, the organic solute pool of *Thermotoga maritima* comprises 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (MDIP) and 2-(*O*–β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (MMDIP), two newly identified β-1,2-mannosides. In cells grown under heat stress, MDIP was the major solute accounting for 43% of the total pool; MMDIP and DIP accumulated to similar levels, each corresponding to 11.5% of the total pool. The synthesis of MDIP involved the transfer of the mannosyl group from GDP-mannose to DIP in a single-step reaction catalyzed by MDIP synthase. This enzyme used MDIP as an acceptor of a second mannose residue, yielding the di-mannosylated compound. Minor amounts of the tri-mannosylated form were also detected. Using a genomic approach, putative genes for MDIP synthase were identified in the genome of *Thermotoga maritima* and the assignment was confirmed by functional expression in *Escherichia coli*. Genes with significant sequence identity were found only in the genomes of *Thermotoga* spp., *Aquifex aeolicus*, *Ferroglobus placidus* and *Archaeoglobus profundus*. MDIP synthase of *Thermotoga maritima* had maximal activity at 95ºC and apparent K_m values of 16 mM and 0.7 mM for DIP and GDP-mannose, respectively. The stereochemistry of MDIP was established resorting to isotopic labeling and NMR: DIP selectively labeled with carbon-13 on C_1 of the L-inositol moiety was synthesized and used as substrate for MDIP synthase. The mannose residue is linked to carbon 2 of the inositol group that is in the L-configuration. This β-1,2-mannosyltransferase is unrelated with known glycosyltransferases and, within the domain *Bacteria*, it is restricted to members of the two deepest lineages, *i.e.*, the *Thermotogales*

and the *Aquificales*. To our knowledge this is the first β-1,2-mannosyltransferase characterized thus far.

Introduction

Thermotoga maritima was first isolated from hot, marine sediments in the Vulcano Island, Italy, being able to grow between 55ºC and 90ºC (Huber et al. 1986). This strictly anaerobic bacterium ferments a variety of simple and complex carbohydrates to acetate, hydrogen and CO_2 (Conners et al. 2006). In line with these metabolic traits, a substantial percentage of the genes annotated in the genome of this hyperthermophile are allocated to the metabolism of mono- and polysaccharides (Chhabra et al. 2003, Nelson et al. 1999). Therefore, *Thermotoga maritima* has been pointed out as a source of glycoside hydrolases with potential industrial relevance, namely in processes of biomass conversion into biofuels (Blumer-Schuette et al. 2008, VanFossen et al. 2008).

Like many other hyperthermophiles isolated from marine environments, *Thermotoga maritima* is slightly halophilic (optimum NaCl concentration of 2.7% (wt/vol)), and developed biochemical strategies to counterbalance the external osmotic pressure. The accumulation of low-molecular mass organic compounds in the cytoplasm is the most common osmoadaptation mechanism, which enables a rapid response to fluctuations in the salinity of the external medium. Interestingly, the organic solutes encountered in organisms adapted to thrive in hot environments are clearly different from those used by mesophiles, leading to the view that osmolytes of (hyper)thermophiles could play an additional role as protectors of macromolecules and other cellular components against heat damage. This

notion is further fuelled by the finding that the total pool of organic solutes of (hyper)thermophiles increases notably not only at supra-optimal salinity but also in response to heat stress conditions (Santos et al. 2007).

Despite the scarcity of genetic tools for the manipulation of marine hyperthermophiles, a number of novel organic solutes were identified and the corresponding biosynthetic pathways characterized at the genetic and biochemical levels (Jorge et al. 2007, Lamosa et al. 2006, Santos et al. 2007), providing the critical knowledge to engage in the elucidation of the molecular basis of the whole process, from sensing stress to the synthesis of specific osmolytes. In this context, we reported recently the characterization of the pathway for the synthesis of di-*myo*-inositol 1,3'-phosphate (Chapter 2). Additionally, the genes and enzymes involved in the relevant reaction steps were disclosed (Chapetr 3, Rodionov et al. 2007). The synthesis proceeds via a phosphorylated form of DIP, and the respective synthase is a membrane associated enzyme that catalyzes the condensation of CDP-inositol with inositol 1-phosphate.

The solute pool in members of the order *Thermotogales* was investigated a few years ago (Martins et al. 1996). *Thermotoga neapolitana* responded to heat stress with a strong accumulation of DIP and DIP derivatives. One of the solutes was assigned to a mannosylated form of DIP, at that time designated di-mannosyl-di-*myo*-inositol 1,3'-phosphate; moreover, the presence of a second DIP-derivative was proposed but its structure remained elusive. Therefore, we set out to fully characterize the solute pool of *Thermotoga* spp. and to identify the genes and the enzyme(s) involved in the synthesis of the DIP derivatives. Members of the genus *Thermotoga* accumulated DIP and two mannosylated forms of this compound, herein fully characterized using NMR and mass spectrometry. Moreover, the pathway for the synthesis of these novel solutes was identified, leading to the

discovery of a unique β–1,2-mannosyltransferase that catalyzes the transfer of the mannosyl group from GDP-mannose to DIP.

Materials and methods

Materials

D,L-glycerol 3-phosphate, CTP, GDP, GDP-mannose, ADP-glucose, TDP-glucose, UDP-glucose, GDP-glucose, phosphoenolpyruvate, glycerol, *myo*-inositol, guanidinium chloride and pyruvate kinase from rabbit muscle were purchased from Sigma-Aldrich (St. Louis, MO). NADH and rabbit muscle lactate dehydrogenase were purchased from Roche Applied Science (Penzberg, Germany). D-*myo*-inositol 1-phosphate was obtained by chemical synthesis (Chapter 2). L-*myo*-inositol 1-phosphate was produced from glucose 6-phosphate and NADH using recombinant *myo*-inositol phosphate synthase of *Archaeoglobus fulgidus* (Chapter 3). DIPP was obtained as described earlier (Chapter 4). Di-*myo*-inositol 1,3'-phosphate (DIP) was supplied by bitop AG (Witten, Germany), and glycerophospho-*myo*-inositol (GPI) was isolated from *Archaeoglobus fulgidus* (Lamosa et al. 2006).

Strains and culture conditions

The type strain of *Thermotoga maritima* (strain 3109^T) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany). *Aquifex aeolicus* strain VF5 (around 5 g of wet weight) was kindly provided by M. Tomm and H. Huber, University of Regensburg, Germany. *Thermotoga maritima* was cultivated in a medium containing (per liter) 27 g of

NaCl, 0.5 g of yeast extract, 3 g of PIPES, 2.5 g of starch, 2 mg of $(NH_4)_2Ni(SO_4)_2 \bullet 7H_2O$, 1.75 g of $MgSO_4 \bullet 7H_2O$, 0.5 g of $MgCl_2 \bullet 6H_2O$, 0.16 g of KCl, 25 mg of NaBr, 7.5 mg of H_3BO_3, 3.8 mg of $SrCl_2 \bullet 6H_2O$, 0.025 mg of KI, 0.1 g of $CaCl_2 \bullet 2H_2O$, 0.35 g of K_2HPO_4, 1 g of NH_4Cl, 3 g of cysteine-HCl, and 10 ml of trace minerals solution (Balch et al. 1979). The final pH was adjusted to 6.5. The medium was degassed with N_2, and sterilized by autoclaving. Prior to inoculation, the medium was supplemented with biotin (20 μg/l), thiosulfate (20 mM) and Na_2S (0.05% wt/vol). The organism was cultured in 2-liter static vessels, under anaerobiosis (N_2 gas phase); cell growth was assessed by measuring optical density at 600 nm.

To examine the effect of osmotic and heat stresses on the level of intracellular organic solutes *Thermotoga maritima* was grown at optimal conditions (2.7% NaCl (wt/vol), 80ºC), under osmotic stress (4.5% (wt/vol) NaCl, 80ºC) and under heat stress (2.7% (wt/vol) NaCl, 88ºC); combined stress conditions (4.5% (wt/vol) NaCl, 88ºC) were also examined but growth was not observed.

Extraction, identification and quantification of intracellular solutes

Cells were harvested during late-exponential phase of growth by centrifugation (7,000 × *g*, 10 min, 4ºC), and washed twice with a NaCl solution identical in concentration to that of the growth medium. Cell pellets were suspended in water and disrupted by sonication. Aliquots were removed for determination of total protein content. The remaining cell extract was treated twice with boiling 80% ethanol as described previously (Santos et al. 2006). Freeze-dried extracts were dissolved in ^2H_2O and analyzed by NMR (Santos et al. 2006).

Purification of mannosyl-di-*myo*-inositol 1,3'-phosphate (MDIP)

Cells of *Thermotoga maritima* grown at 88ºC and 2.7% (wt/vol) NaCl were harvested during late exponential growth phase (optical density at 600 nm of 0.3) by centrifugation (7,000 × *g*, 10 min, 4ºC), and washed twice with a NaCl solution identical in concentration to that of the growth medium. The ethanolic cell extract, obtained as described above, was loaded onto a QAE-Sephadex A-25 column previously equilibrated with 5 mM sodium bicarbonate (pH 9.8), and elution was performed with a linear gradient of the same buffer (5 mM to 1 M). Fractions were analyzed by ^{31}P-NMR to monitor the presence of MDIP. Samples containing this compound were desalted using a column of activated Dowex 50W-X8. The pH of the eluted fractions was adjusted to 6 with 1 M KOH prior to lyophilization. The freeze-dried sample was dissolved in 1-propanol:ammonia (1:1.5, vol/vol) and loaded onto a Silica gel S column; elution was carried out with the same solvent mixture. The fractions enriched in MDIP were applied to a second Silica gel S column. The fractions containing pure MDIP were pooled, lyophilized and dissolved in $^{2}H_2O$ for structure determination by NMR. The yield of the purification was only 8% due to the difficulty in removing the contamination with MMDIP.

Preparation of crude cell extracts for detection of enzyme activity

Thermotoga maritima cells grown at 88ºC and 2.7% (wt/vol) NaCl were harvested by centrifugation and washed twice with Tris-HCl buffer (20 mM, pH 7.6) containing 2.5% (wt/vol) NaCl, under anaerobic conditions. The cell pellet was suspended in the same buffer and cells were disrupted in a French press. Cell debris was removed by centrifugation (30,000 × *g*, 45 min, 4ºC). Low molecular mass compounds were removed in a PD-10 column (GE Healthcare

Bio-Science AB, Uppsala, Sweden) equilibrated with Tris-HCl (20 mM, pH 7.6) containing 10 mM $MgCl_2$. Protein content was estimated by the Bradford method (Bradford 1976).

Cloning and expression of putative *mds* genes

Chromosomal DNA from *Thermotoga maritima* and *Aquifex aeolicus* was isolated according to the method of Ramakrishnan and Adams (Ramakrishnan et al. 1995). The genes of *Thermotoga maritima* (TM_0359) and *Aquifex aeolicus* (aq_1141), herein designated *mds* gene, were amplified by PCR using *Pfu* DNA polymerase (Fermentas, Burlington, Canada) and cloned in pET23a and pTRC99a, respectively, following standard protocols (Sambrook et al. 1989). The strategies used to clone the genes led to the addition of a methionine residue at the N-terminus of the protein from *Thermotoga maritima*, and 13 extra residues (Met-6His-Gly-4Asp-Lys) at the N-terminus of the *Aquifex aeolicus* protein. Enterokinase was used to remove the extra peptide, but the yield was very low (around 10%). *Escherichia coli* BL21(DE3) cells, bearing the constructs, were grown at 37ºC in LB medium with ampicillin (100 μg/ml) to an optical density of 0.5, and treated with 1 mM IPTG for 4 h. Cells were harvested, suspended in Tris-HCl (20 mM, pH 7.6) containing 10 mM $MgCl_2$ and disrupted in a French press; cell debris was removed by centrifugation (30,000 × *g*, 4ºC, 1 h).

Purification of recombinant *Thermotoga maritima* MDIP synthase

The cell extract (approximately 40 mg of protein/ml) resulting from the previous step was heated for 20 min at 70ºC to precipitate thermo-labile

proteins; after centrifugation the supernatant solution was applied onto a Resource Q column (GE Healthcare Bio-Science AB, Uppsala, Sweden) equilibrated with Tris-HCl buffer (20 mM, pH 7.6). Elution was carried out with a linear gradient (0-1 M NaCl). SDS-PAGE of fractions eluted between 0.1 and 0.2 M NaCl revealed a strong band with the expected molecular mass of 42 kDa. Moreover, the presence of this band correlated with MDIP synthase activity, as evaluated by ^{31}P-NMR to monitor the formation of MDIP using DIP and GDP-mannose as substrates. Fractions containing MDIP synthase were pooled, concentrated using centrifugal filter devices, and dialyzed against Tris-HCl buffer (20 mM, pH 7.6). The solution was applied onto a Mono Q column (GE Healthcare Bio-Science AB, Uppsala, Sweden), equilibrated with Tris-HCl buffer (20 mM, pH 7.6). Fractions containing MDIP synthase eluted between 0.2 and 0.3 M NaCl. This last chromatographic step was repeated and a fraction with pure MDIP synthase was obtained as judged by SDS-PAGE and silver staining.

Purification of recombinant *Aquifex aeolicus* MDIP synthase

The cell extracts were treated as above; after a first chromatographic step (Resource Q) the active fractions were applied onto an HisTrap™ HP column (GE Healthcare Bio-Science AB, Uppsala, Sweden). The protein eluted with 500 mM imidazole. The active sample was applied to a Phenyl-Sepharose column, and elution carried out with a linear gradient of ammonium sulfate (1.7 M – 0 M). The activities eluted at 0.5 M of $(NH_4)_2SO_4$.

Enzyme assays

MDIP synthase activity was detected in cell extracts of *Thermotoga maritima* and also in cell extracts of *E. coli* harboring *mds* genes either from *Thermotoga maritima* or *Aquifex aeolicus*. Reaction mixtures (0.4 ml), containing Tris-HCl buffer (20 mM, pH 7.6), 10 mM $MgCl_2$, 4 mM of the substrates, and cell extract (around 2 mg of total protein), were incubated for 1 h at 80ºC. The formation of the product, MDIP, was detected by the appearance of a resonance at -0.7 ppm in the ^{31}P-NMR spectra. Assays for the characterization of MDIP synthase were performed in 0.15 ml of solutions containing 50 mM Bis/Tris/Propane-HCl buffer, pH 7.6, 10 mM $MgCl_2$, 16 mM DIP and 2.2 mM GDP-mannose; the reaction mixtures were incubated at the specified temperatures in 2 ml glass tubes pressurized with 2 atm of argon and pre-heated for 2 min. The reaction was initiated by the addition of the enzyme (3 µg), and stopped at different times by immersion in liquid nitrogen. The MDIP synthase activity was measured from the amount of GDP released during incubation with GDP-mannose and DIP. GDP was quantified by monitoring the oxidation of NADH at 340 nm (Gu et al. 2005). This assay was performed at 30ºC in 1-ml reaction mixtures containing Tris-HCl (20 mM, pH 7.6), 10 mM $MgCl_2$, 0.4 mM phosphoenolpyruvate, 0.175 mM NADH, 2.7 U of pyruvate kinase, 4 U of lactate dehydrogenase, and 70 µl of the MDIP synthase assay mixture.

Characterization of *Thermotoga maritima* MDIP synthase

The temperature profile for the activity was determined between 70ºC and 105ºC. The pH of the buffer as measured at 25ºC was 7.6. Using the conversion factor $\Delta pK_a/\Delta T$ ºC = -0.015 for Bis/Tris/Propane we estimated that

the working pH value varied between 6.9 and 6.4 in the temperature range studied. Kinetic parameters (V_{max} and K_m) were determined at 95ºC and calculated pH of 5.7 in reaction mixtures (0.15 ml final volume) containing 5 mM GDP-mannose and 0-300 mM DIP or 0-3.5 mM GDP-mannose and 98 mM DIP. The reaction was started by the addition of 3 µg of MDIP synthase. Experiments were done in duplicate. K_m was determined by fitting the data to Michaelis-Menten equations using the software Origin 5.0 Professional (Microcal Software Inc., MA).

Preparation of di-*myo*-[1-^{13}C]inositol 1,3'-phosphate and determination of MDIP stereochemistry

myo-[1-^{13}C]inositol 1-phosphate and CDP-*myo*-[1-^{13}C]inositol were produced enzymatically. *myo*-[1-^{13}C]inositol 1-phosphate was produced from D-[6-^{13}C]glucose by coupling hexokinase from *Thermoproteus tenax* and *myo*-inositol 1-phosphate synthase from *Archaeoglobus fulgidus* as described in Chapter 4. Subsequently, CDP-*myo*-[1-^{13}C]inositol was synthesized from CTP and *myo*-[1-^{13}C]inositol 1-phosphate by using a cell extract of *E. coli* BL21(DE3) harboring the gene encoding IPCT/DIPPS of *Rubrobacter xylanophilus* (Chapters 3 and 4). Finally, the synthesis of di-*myo*-[1-^{13}C]inositol 1,3'-phosphate was performed in a reaction mixture containing 4 mM CDP-*myo*-[1-^{13}C]inositol, 4 mM of D,L-*myo*-inositol 1-phosphate, 10 mM $MgCl_2$, 4 mg total protein/ml of a cell extract of *Archaeoglobus fulgidus* in MOPS buffer (50 mM, pH 8.1). After 30 min of incubation at 80ºC, the reaction product, di-*myo*-[1-^{13}C]inositol 1,3'-phosphate, was purified by anion exchange chromatography as previously described (Chapter 4). The final product was 93% enriched in carbon 13 at position 1; the presence of minor

amounts of isotopomers without labeling at this position is due to the fact that the inositol moieties of inositol 1-phosphate and CDP-inositol can exchange to a low extent in the reaction catalyzed by DIPP synthase (Chapter 2). Labeled MDIP was produced from di-*myo*-[1-^{13}C]inositol-1,3'-phosphate and GDP-mannose supplied to cell extracts of *Thermotoga maritima* or to a partially purified preparation of MDIP synthase. The stereochemistry of MDIP was determined by using ^{1}H-, ^{31}P-, and ^{13}C-NMR.

NMR spectroscopy

For structure determination, spectra were acquired on a Bruker *AVANCE*^{III} spectrometer (Bruker, Rheinstetten, Germany) operating at 800.33 MHz for protons. ^{1}H, ^{13}C and ^{31}P chemical shifts are relative to 4,4-dimethyl-4-silapentane sodium sulfonate (DSS), external methanol designated at 49.3 ppm and external 85% H_3PO_4, respectively. The homonuclear proton correlation spectra (COSY), as well as heteronuclear correlation spectra (^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, and ^{1}H-^{31}P HSQC) were acquired at 25^{o}C using standard Bruker pulse programs. The homonuclear correlation spectra were acquired with pre-saturation of the water signal. In the HSQC sequences, $J_{C,H}$ and $J_{P,H}$ of 145 Hz and 7 Hz were used to calculate the delays for evolution of scalar couplings; in the ^{1}H-^{13}C HMBC spectra, 8 Hz was used for $J_{C,H}$. For quantification of the organic solutes accumulating in *Thermotoga maritima*, ^{1}H-NMR spectra were acquired on a Bruker *AVANCE*^{II}500 spectrometer equipped with a broadband inverse probe head. Spectra were acquired with pre-saturation of the water signal using a repetition delay of 60 s. Formate was used as an internal concentration standard.

Differential scanning calorimetry

Measurements were performed on a MicroCal VP-DSC MicroCalorimeter as previously described (Faria et al. 2008). Samples of MDIP synthase from *Aquifex aeolicus* (0.35 mg/ml) were prepared by extensive dialysis against 10 mM phosphate buffer, pH 6.2. Guanidinium chloride (1.5 M final concentration) was added before the assay. Scans were run at a heating rate of 1°C/min from 25 to 120°C with an overpressure of approximately 30 psi. Baselines were acquired with the same experimental conditions as for the assays both in the absence and in the presence of guanidinium chloride. The melting temperature was determined using MicroCal software.

Results

Identification of intracellular organic solutes in *Thermotoga maritima*

^{1}H-, ^{31}P- and ^{13}C-NMR spectra of ethanol extracts of *Thermotoga maritima* showed the presence of DIP, several DIP-derivatives, and glutamate (α- and β-forms) (Fig. 5.1a). The ^{1}H-^{31}P correlation spectra of extracts of cells grown at 80ºC showed three signals previously assigned to DIP, to a symmetric mannosylated derivative of DIP, and to an unknown compound, tentatively assigned to a different DIP stereoisomer (Fig. 3 in Martins et al. (1996)). The enhanced resolution provided by a high field spectrometer allowed us to conclude that the three sets of resonances corresponded to two compounds only, DIP and an asymmetric mannosylated derivative of DIP, herein firmly identified as 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (MDIP) on the

basis of an array of spectra (^{1}H COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC and ^{1}H-^{31}P HSQC) of the purified MDIP (Fig. 5.1b).

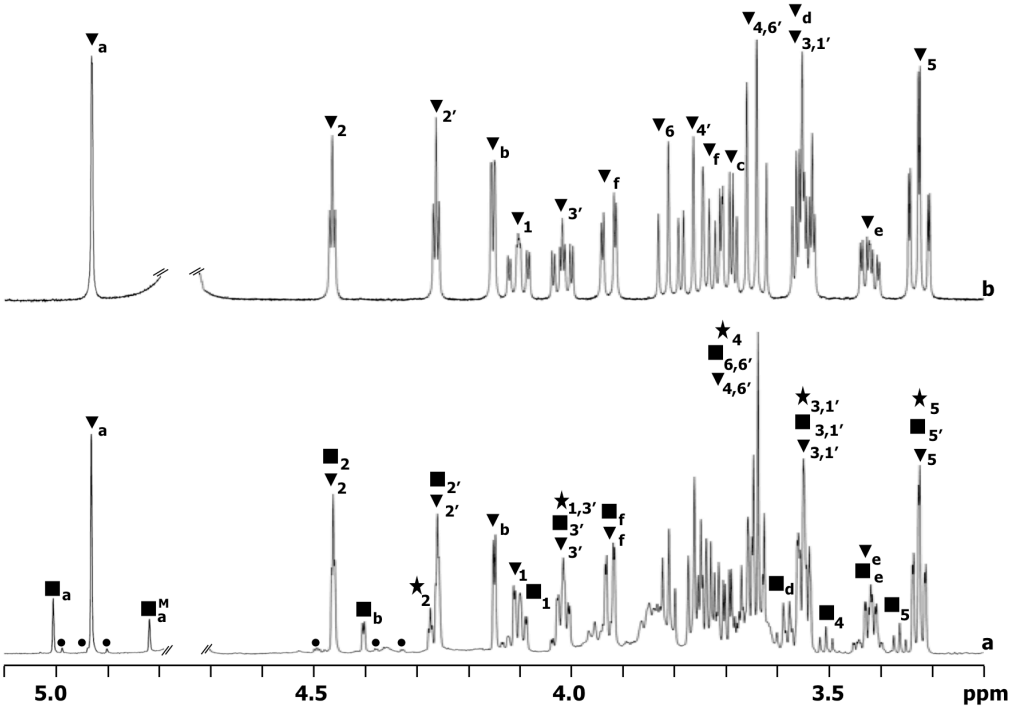


Figure 5.1. ^{1}H-NMR spectra of an ethanol extract of *Thermotoga maritima* grown at 88ºC in medium containing 2.7% (wt/vol) NaCl (a), and pure 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate (b). Resonances due to protons H_1, H_2, H_3, H_4, H_5 and H_6 of the *myo*-inositol moieties are identified by numbers (1, 2, 3, 4, 5 and 6, respectively) and labeled as follows: di-*myo*-inositol 1,3′-phosphate, DIP (★), mannosyl-di-*myo*-inositol 1,3′-phosphate, MDIP (▼), di-mannosyl-di-*myo*-inositol 1,3′-phosphate, MMDIP (■) and tri-mannosyl-di-*myo*-inositol 1,3′-phosphate (•). Resonances due to protons of the mannose moieties labeled with a, b, c, d, e, and f designate H_1, H_2, H_3, H_4, H_5 and $H_{6a\ or\ b}$, respectively. The label M designates the resonances due to the mannose moiety of MMDIP not directly linked to inositol. The nomenclature used for inositol is according to Chapter 4.

Moreover, the mass spectrum of a partially purified extract of *Thermotoga maritima* revealed two major signals, one with *m/z* of 420.9 that matched perfectly the molecular mass of DIP, and the other with *m/z* of 583.1 that was consistent with MDIP. The NMR parameters of MDIP are presented in Table 5.1. The value of 162 Hz for the $^{1}J_{C,H}$ coupling constant at position 1 of the mannose residue clearly indicates a β-configuration of this sugar moiety (Bock et al. 1974). The ^{1}H-^{13}C HMBC spectrum placed the glycosidic linkage of the mannosyl residue at position 2 of one inositol group, while the ^{1}H-^{31}P HSQC showed that the phosphate group was linked to positions 1 and 3′ of the inositol moieties. The set of resonances herein assigned to MDIP are also present in spectra of cell extracts of *Aquifex pyrophilus* (Lamosa et al. 2006) and *Aquifex aeolicus* (data not shown); therefore, we conclude that MDIP is synthesized also by these two members of the genus *Aquifex.*

In addition to resonances assigned to DIP and MDIP, the proton NMR spectra of extracts of *Thermotoga maritima* grown at 88^{o}C showed an extra set of resonances which comprised two signals at 5.00 and 4.81 ppm, the region typical of the anomeric protons of hexoses (Fig. 5.1a). Coupling constants $^{3}J_{H1,H2} \leq 2$ Hz and $^{1}J_{C1,H1}$ of 161 and 162 Hz were determined, and denote mannose residues in the β-configuration.

A partially purified preparation of the unknown compound was analyzed by ^{1}H-, ^{13}C-, and ^{31}P-NMR. The set of NMR data were consistent with the structure depicted in Figure 5.2; the relevant parameters are presented in Table 5.1. The compound was designated 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate (MMDIP) (Fig. 5.2). This identification was corroborated by the observation in the mass spectrum of a signal at *m/z* of 745.1, corresponding to the expected molecular mass of MMDIP.

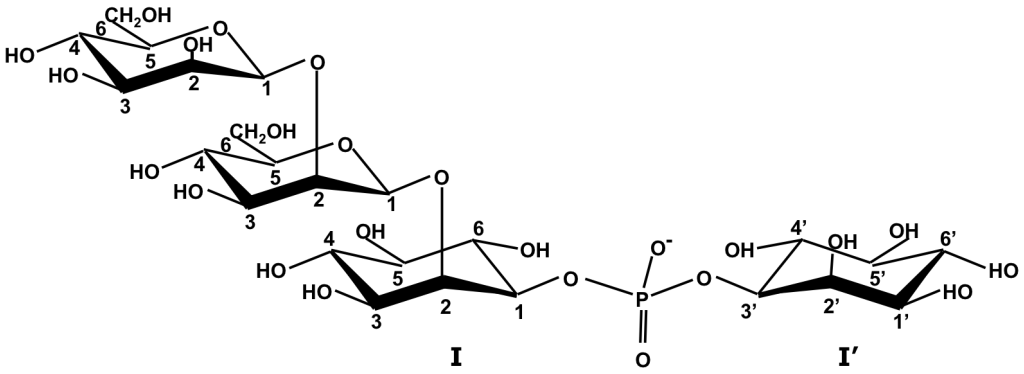


Figure 5.2. Molecular representation of 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (MMDIP), a newly identified organic solute accumulating in cells of *Thermotoga maritima* grown under heat stress conditions.

The NMR spectra of the ethanolic extract (Fig. 5.1a) revealed yet another set of very weak resonances which comprised three signals in the region of the anomeric protons of hexoses, at 4.98 ppm, 4.92 ppm and 4.89 ppm. These three signals belong to a vestigial tri-mannosylated derivative of DIP, but the unambiguous assignment of all the signals was precluded by overlapping with the much stronger signals of MMDIP. Nevertheless, it is clear from the ^{1}H-^{13}C HMBC spectra that the three mannosyl residues are linked sequentially through β–(1,2) glycosidic bonds. Therefore, we propose that this minor compound is 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate, which is in accordance with a peak at *m/z* of 907.0 detected in the mass spectrum.

Table 5.1. NMR parameters of 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate (MDIP) and 2-(*O*–β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate (MMDIP). The ^{31}P chemical shifts were derived directly from the one-dimensional ^{31}P-NMR spectra. ^{1}H and ^{13}C chemical shifts were obtained from the ^{1}H-^{13}C HSQC spectra; the $^{1}J_{C,H}$ coupling constants were measured in ^{1}H-^{13}C HSQC spectra acquired without carbon decoupling.

Moiety and position	MDIP[a]			MMDIP[a]		
	^{13}C-NMR[b]		^{1}H-NMR[c]	^{13}C-NMR[b]		^{1}H-NMR[c]
	δ (ppm)	$^{1}J_{C,H}$ (Hz)	δ (ppm)	δ (ppm)	$^{1}J_{C,H}$ (Hz)	δ (ppm)
Inositol (I)						
C-1	76.3	143	4.10	75.9	143	4.12
C-2	79.8	153	4.46	80.5	153	4.46
C-3	70.5	145	3.54	70.0	143	3.57
C-4	73.0	148	3.66	73.1	145	3.51
C-5	74.2	144	3.32	73.7	143	3.36
C-6	71.9	148	3.81	72.2	146	3.67
Inositol (I′)						
C-1′	70.8	144	3.56	76.7	144	4.02
C-2′	71.5	151	4.26	71.8	151	4.26
C-3′	76.8	143	4.03	70.8	143	3.55
C-4′	71.7	146	3.76	72.3	145	3.64
C-5′	74.1	144	3.33	74.1	143	3.32
C-6′	72.6	146	3.64	71.6	147	3.76
Mannose						
C-1	100.9	162	4.93	101.1	162	5.01
C-2	70.8	150	4.15	78.7	150	4.41
C-3	73.1	141	3.69	72.2	141	3.71
C-4	67.1	145	3.55	67.2	145	3.59
C-5	76.4	142	3.41	76.4	142	3.44
C-6	61.3	150	3.91; 3.73	60.9	nd	3.92; 3.75
Mannose						
C-1				101.1	161	4.81
C-2				70.7	148	4.11
C-3				72.9	141	3.67
C-4				66.8	145	3.58
C-5				76.5	142	3.41
C-6				61.1	nd	3.93; 3.74

[a] Measurements were carried out at 35°C and pH 6.5. The phosphorus resonances of MDIP and MMDIP appear at -0.7 ppm relative to external 85% H_3PO_4. [b] Methanol was used as reference designated at 49.3 ppm. [c] 4,4-dimethyl-4-silapentane sodium sulfonate (DSS) was used as reference. nd, not determined.

Effect of salt and growth temperature on the accumulation of organic solutes in *Thermotoga maritima*

Under optimal growth conditions (2.7% (wt/vol) NaCl, 80ºC), the total amount of organic solutes in *Thermotoga maritima* was around 0.2 µmol·mg of protein^{-1} and increased up to 0.7 µmol·mg of protein^{-1} in cells grown under heat or osmotic stress (Fig. 5.3 and Table 5.2). The proportion of organic solutes was dependent on the type of stress imposed; for example, the level of glutamate (α- and β- forms) increased notably in response to osmotic stress, whereas MDIP and MMDIP clearly increased under heat stress. Curiously, the ratio β-glutamate/α-glutamate was 1.8 in cells grown under osmotic stress, but decreased to 0.1 in cells grown at supra-optimal temperature. When *Thermotoga maritima* was cultivated in medium with 4.5% NaCl (wt/vol) (osmotic stress conditions), β- and α-glutamate dominated the solute pool, reaching 77% (molar percentage) of the total organic compounds. In contrast, DIP and DIP-derivatives represented 65% of the solute pool of *Thermotoga maritima* grown at 88ºC (up-shift of 8ºC relative to optimal), while β- and α-glutamate amounted to the remaining 35%. MDIP was the predominant compound (43% of the total), corresponding to 10-fold increase in concentration in respect to the optimal temperature conditions.

The newly identified DIP-derivative, MMDIP, was only detected in cells subjected to heat stress. Interestingly, the level of the parent compound, DIP, was fairly constant in the range of growth conditions examined.

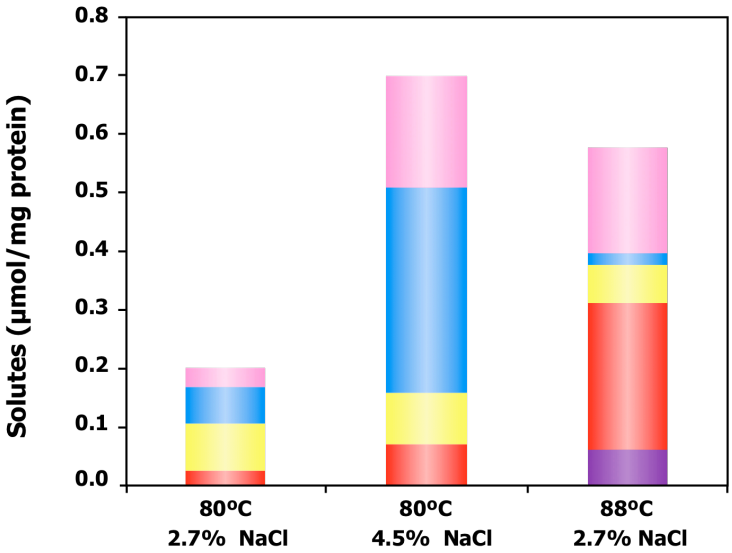


Figure. 5.3. Effect of the growth temperature and the NaCl concentration on the accumulation of α–glutamate (■ pink), β-glutamate (■ blue), di-*myo*-inositol 1,3'-phosphate (■ yellow), mannosyl-di-*myo*-inositol 1,3'-phosphate (■ red), di-mannosyl-di-*myo*-inositol 1,3'-phosphate (■ purple) in *Thermotoga maritima*. Optimal growth conditions are 80ºC and 2.7% (wt/vol) NaCl. Cultures were harvested during the late exponential phase of growth. The data are averages of two independent growths at each condition.

Table 5.2. Intracellular content of organic solutes in *Thermotoga maritima* grown at different temperatures and NaCl concentrations. Quantification was performed by ^{1}H-NMR.

Growth temp (ºC)	% NaCl (wt/vol)	Intracellular concentration (µmol·mg protein^{-1})[a]				
		α-glu	β-glu	DIP	MDIP	MMDIP
80	2.7	0.03± 0.00 (7.1)	0.06± 0.01 (13.8)	0.08± 0.01 (18.0)	0.02± 0.00 (5.5)	-
88	2.7	0.18± 0.03 (39.8)	0.02± 0.01 (4.3)	0.07± 0.01 (14.8)	0.25± 0.00 (55.4)	0.06± 0.00 (13.9)
80	4.5	0.19± 0.01 (42.5)	0.35± 0.01 (77.6)	0.09± 0.00 (19.6)	0.07± 0.01 (15.6)	-

[a] The intracellular concentration (mM) based on a cell volume of 4.5 µl·mg protein^{-1} determined for *Pyrococcus furiosus* (Martins and Santos 1995) is given in parenthesis. α-glu, α-glutamate; β-glu, β-glutamate; DIP, di-*myo*-inositol 1,3'-phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate.

Identification of the gene(s) for the synthesis of MDIP

The synthesis of MDIP was investigated by ^{31}P-NMR spectroscopy in cell extracts of *Thermotoga maritima* using GDP-mannose and DIP as putative sugar donor and acceptor, respectively. In addition to the DIP resonance at -0.57 ppm, a new resonance appeared at -0.7 ppm that was due to the reaction product, MDIP (Fig. 5.4a). This identification was firmly established by spiking with pure MDIP. To identify the gene encoding this activity, we searched Pfam (http://pfam.sanger.ac.uk/) for glycosyltransferases of *Thermotoga maritma*. Fourteen genes were found whose predicted products belong to the family 1 of glycosyltransferases. Eight of them are annotated, leaving six genes (TM_0392, TM_0359, TM_0619, TM_0628, TM_0744 and TM_1230) as putative candidates to encode MDIP synthase. A Blast search in public databases revealed that the gene TM_0359 had homologs (around 35% sequence identity) in the genomes of *Aquifex aeolicus* (aq_1141), a hyperthermophilic bacterium that accumulates MDIP (data not shown) and of *Archaeoglobus profundus* (Arcpr_1596) and *Ferroglobus placidus* (Ferp_1658 and Ferp_1639), two organisms in which the presence of MDIP has not yet been reported. Nearly identical genes (99% identity) were present in the genomes of *Thermotoga neapolitana*, *Thermotoga petrophila*, *Thermotoga naphthophila* and *Thermotoga* sp. RQ2. No hits were found in members of genera other than *Thermotoga*, *Aquifex*, *Ferroglobus* and *Archaeoglobus*. The genes TM_0359 and aq_1141 were selected for further analysis. The genes were separately cloned and cell extracts of *E. coli* BL21(DE3) harboring pET23a:TM_0359 or pET19b:aq_1141 were examined for the presence of MDIP synthase activity. The formation of MDIP was proven by ^{31}P-NMR analysis (Fig. 5.4c). We verified that MDIP was not formed in cell extracts of *E. coli* bearing empty plasmids (pET23a or pET19b) (Fig. 5.4b). These results

showed definitely that TM_0359 and aq_1141 encode enzymes capable of synthesizing MDIP.

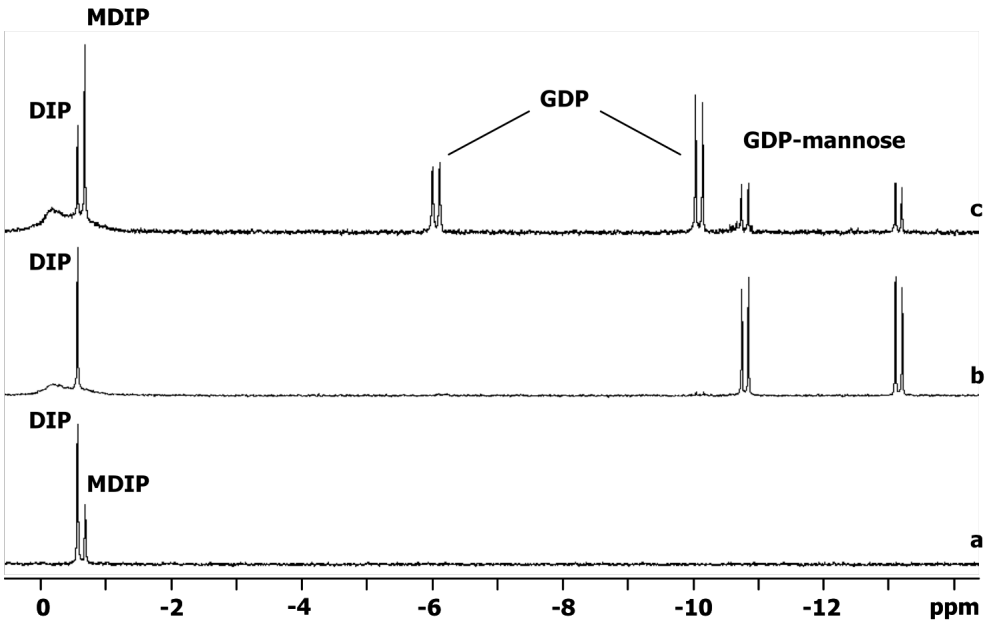


Figure. 5.4. ^{31}P-NMR spectra of the final products resulting from incubation at 80ºC of reaction mixtures containing 4 mM GDP-mannose, 4 mM DIP, 10 mM $MgCl_2$ in 20 mM Tris-HCl buffer, pH 7.6, with: (a) Cell extract of *Thermotoga maritima* grown under heat stress conditions (b) *E. coli* cell extract harboring pET23a, after heat treatment at 70ºC for 10 min, and (c) *E. coli* cell extract harboring the construction pET23a:*mds*, after heat treatment at 70ºC for 10 min. Spectra were run at a sample temperature of 27ºC. DIP, di-*myo*-inositol-phophate; MDIP, 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate. The two resonances due to GDP show different intensities due to incomplete relaxation of these signals.

Catalytic properties of *Thermotoga maritima* MDIP synthase

An array of experiments was carried out to evaluate the specificity of MDIP synthase. ADP-glucose, GDP-glucose, UDP-glucose and TDP-glucose were examined as putative substrates together with DIP, but no reaction product

was detected by ^{31}P-NMR. GDP-mannose was the only glycosyl donor used by the enzyme. Mannose activated nucleotides other than GDP-mannose were not examined due to commercial unavailability. DIP, MDIP, DIPP, GPI, DGP, glycerol, DL-glycerol 3-phosphate, *myo*-inositol, L-*myo*-inositol 1-phosphate and D-*myo*-inositol 1-phosphate were tested as putative mannosyl acceptors. DIP was used by the enzyme resulting in the formation of MDIP and low amounts of MMDIP; it was also verified that the enzyme was able to use MDIP and GDP-mannose to yield MMDIP. Therefore, MDIP synthase catalyzed two reactions: (i) the transfer of the mannose group from GDP-mannose to DIP, producing MDIP; and (ii) the transfer of the mannose group from GDP-mannose to MDIP, yielding MMDIP. This means that the product of the first reaction is also a substrate of the enzyme. We set out to characterize the kinetic parameters of the enzyme using GDP-mannose and DIP as substrates.

The initial reaction rates were obtained from the quantification of GDP produced using a spectrophotometric method, which has the advantage of a greater sensitivity compared to ^{1}H-NMR. MDIP synthase exhibited Michaelis-Menten kinetics; the apparent K_m values for DIP and GDP-mannose are presented in Table 5.3.

Table 5.3. Kinetic properties of recombinant MDIP synthase of *Thermotoga maritima* and effect of Mg^{2+} and Na^{+} ions.

Parameters	Substrate	MDIP synthase[c]
K_m (mM)[a]	DIP	16 ± 4.0
	GDP-mannose	0.7 ± 0.1
V_{max} (µmol/min·mg protein^{-1})		> 60
Mg^{2+}		
0		31[b]
10 mM		100[b]
NaCl		
0		100[b]
400 mM		51[b]

[a] K_m values for MDIP synthase were determined in mixtures containing 5 mM GDP-mannose and 0-300 mM DIP or 98 mM DIP and 0-5 mM GDP-mannose. [b] Expressed as percentages of the maximum activity. [c] All assays were carried out at 95ºC.

It should be noted that the spectrophotometric assay assesses the sum of MDIP and MMDIP produced in the two reaction steps; however, if the contribution of the MMDIP-forming reaction is small, the values of the kinetic parameters determined are good approximations to describe the first catalytic reaction. To assess the contribution of the MMDIP-forming reaction, the protocol described to determine initial rates was followed (see Materials and Methods) with 100 mM DIP and 5 mM GDP-mannose, substrate concentrations intended to be saturating for MDIP synthase; the reaction was allowed to proceed for 30 seconds at 95ºC as in the spectrophotometric assay, and the reaction products MDIP and MMDIP were quantified by NMR. The concentration of MDIP after 30 sec was 0.11 mM; MMDIP was not detected (below the sensitivity limit, *i.e.*, 0.03 mM). This shows that the formation of MMDIP is negligible under our assay conditions since DIP is likely

to compete with MDIP for the same binding site and the DIP concentration is at least 1000-fold higher than that of MDIP. Therefore, it is reasonable to assume that the second reaction step does not affect appreciably the values determined for K_m and V_{max} (Table 5.3). The kinetic properties of the enzyme for the second reaction (MDIP and GDP-mannose as substrates) were not investigated due to difficulties in obtaining gram-amounts of pure MDIP required for the assays. However, an estimation of the K_m for MDIP was made on the basis of a kinetic model (see Discussion). The presence of Mg^{2+} was required for maximal activity of MDIP synthase; in the absence of this cation the activity was only 30% of the maximum value. The activity was 50% lower in the presence of 400 mM NaCl. The temperature for maximal activity was around 95ºC (Fig. 5.5).

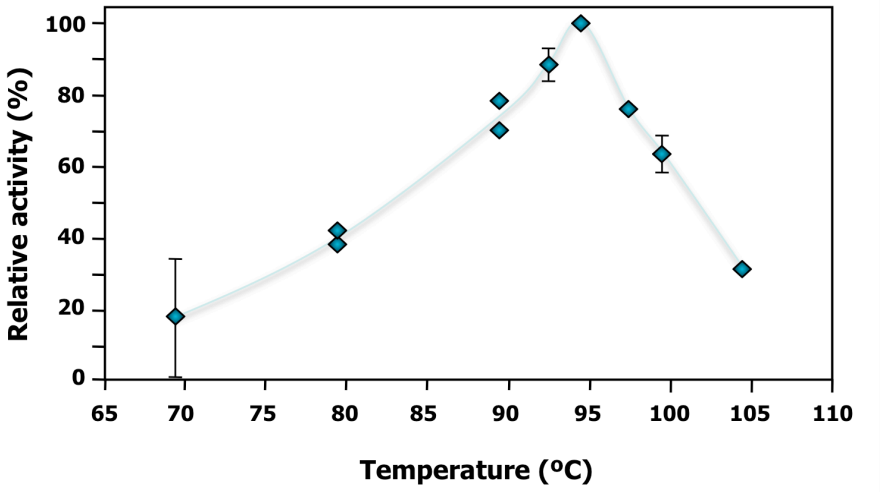


Figure 5.5. Temperature dependence of the activity of recombinant *Thermotoga maritima* MDIP synthase. The data for 98ºC and 105ºC are from single experiments. The experimental conditions are described in "Materials and Methods".

Thermodynamic stability of MDIP synthase

The stability of the recombinant MDIP synthase of *Aquifex aeolicus* was studied by differential scanning calorimetry. The enzyme from *Aquifex aeolicus* was used for these studies due to the high production yield. The protein was highly stable showing a transition temperature of 100ºC in the presence of 1.5 M guanidinium chloride. In the absence of denaturants no transition was observed in the temperature range examined (up to 120ºC).

Determination of the stereochemistry of MDIP

The array of NMR data presented in Table 5.1 led to the conclusion that the mannosyl group in MDIP is linked at position 2 of one of the inositol moieties, and the identification of this inositol moiety was achieved using specific isotopic labeling. For this purpose, DIP specifically labeled with carbon-13 in position 1 of the inositol moiety I was synthesized and used as substrate for MDIP synthase (see Fig. 5.6 and Materials and Methods).

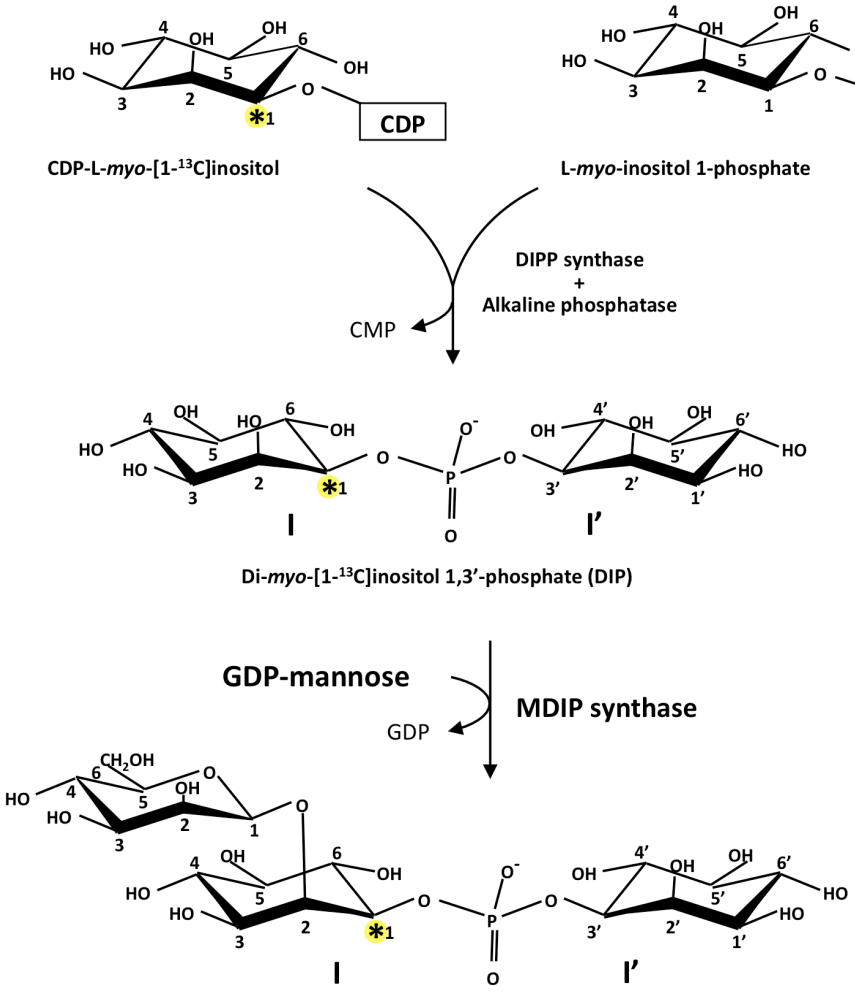


Figure 5.6. Reaction scheme used for the synthesis of mannosyl-di-*myo*-[1-^{13}C]inositol 1,3'-phosphate. CDP-L-*myo*-[1-^{13}C]inositol was the labeled precursor for the synthesis of di-*myo*-[1-^{13}C]inositol 1,3'-phosphate (DIP). The latter compound and GDP-mannose were incubated with a cell extract of *Thermotoga maritima* or with recombinant MDIP synthase to produce mannosyl-di-*myo*-[1-^{13}C]inositol 1,3'-phosphate (MDIP). I and I' designate the inositol moieties that are enantiomers (nomenclature according to Chapter 4). Asterisks designate the position of labeled carbon atoms.

The position of the mannosyl residue in MDIP was deduced from the analysis of ^{1}H-^{31}P correlation spectra (Fig. 5.7.). It is worth pointing out that the NMR resonance of a proton directly bound to a carbon-13 atom (-CH bond) will be split in two signals separated by a large coupling constant (typically greater than 120 Hz). It is also pertinent to note that proton 1 (of inositol I) and proton 3´ (of inositol I´) in DIP resonate at the same frequency since DIP is a symmetric molecule. However, when a mannosyl group is added to one of the inositol moieties of DIP, the symmetry is broken and those protons resonate at 4.10 ppm and 4.03 ppm, the former resonance being assigned to the inositol moiety carrying the mannose group (Table 5.1). The ^{1}H-^{31}P HSQC spectrum (Fig. 5.7) reveals a correlation between the phosphorus signal of MDIP (at -0.7 ppm) and the proton signal centered at 4.10 ppm which is split in two components separated by 143 Hz (≈0.18 ppm); the spectrum also shows a correlation with another proton at a chemical shift of 4.03 ppm whose resonance exhibits no splitting and coincidentally overlaps the component at lower frequency. These data show that the proton whose signal is split due to direct bonding with carbon-13 resonates at 4.10 ppm, hence belonging to the inositol moiety that carries the mannosyl residue (Table 5.1). Therefore, the mannosyl residue is linked to inositol I. The stereochemical representation of the MDIP molecule is shown in Figure 5.7.

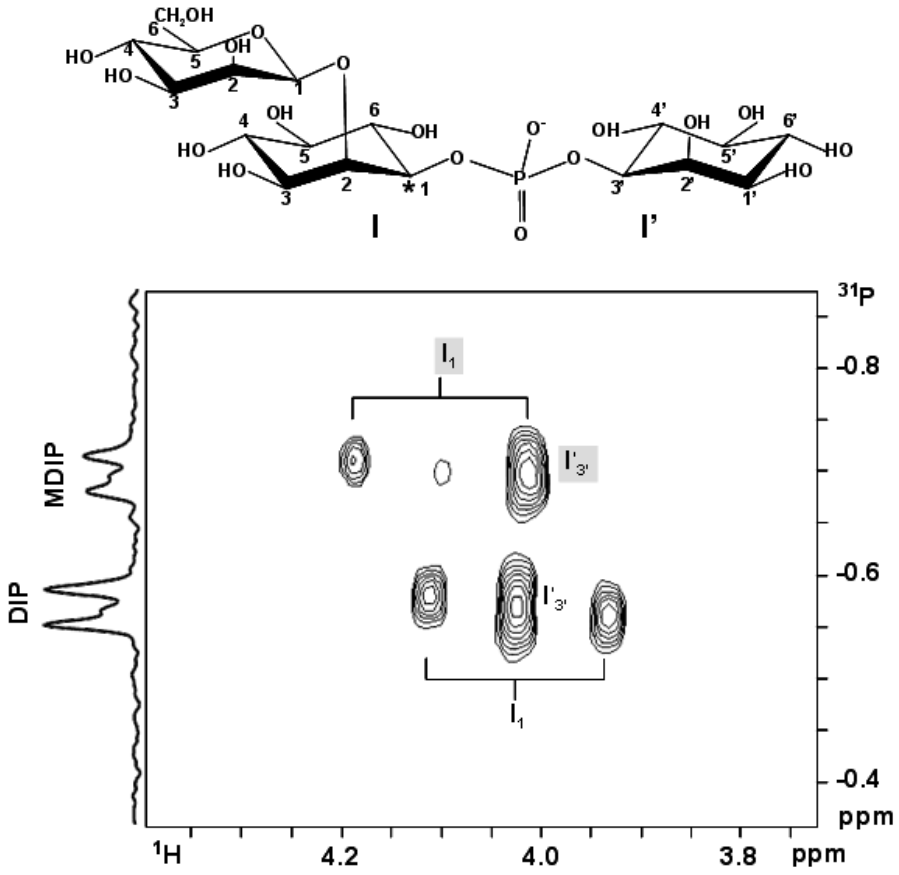


Figure 5.7. Section of a ^{1}H- ^{31}P HSQC correlation spectrum of mannosyl-di-*myo*-[1-^{13}C]inositol 1,3′-phosphate (MDIP) produced from the incubation of GDP-mannose and di-*myo*-[1-^{13}C]inositol 1,3′-phosphate (DIP) with a cell extract of *Thermotoga maritima*. The proton resonance at 4.10 ppm is due to proton 1 of the inositol moiety of MDIP that is directly linked to mannose; the splitting of this signal, due to scalar coupling with carbon-13, shows that the mannosyl group is definitely linked to the inositol moiety that carries the label *i.e.*, moiety I (see text for details). Shaded I_1 and $I'_{3'}$ designate proton 1 and proton 3′ of MDIP, respectively; non shaded I_1 and $I'_{3'}$ designate protons 1 and 3′ of the residual reagent, DIP, respectively.

Discussion

The solute pool of *Thermotoga maritima* is composed exclusively of negatively charged compounds, illustrating the strong preference of hyperthermophilic organisms for charged solutes (Santos et al. 2007). This bacterium accumulates α- and β-glutamate, DIP, and two novel β-1,2-mannosides, herein identified as mannosyl-di-*myo*-inositol 1,3'-phosphate and di-mannosyl-di-*myo*-inositol 1,3'-phosphate. Thus far, we found MDIP and MMDIP only in species of the genera *Thermotoga* (*Thermotoga neapolitana* and *Thermotoga maritima*) and *Aquifex* (*Aquifex aeolicus* and *Aquifex pyrophilus*). The rareness of these solutes and their close relationship with DIP, the most widespread solute in hyperthermophiles, motivated us to investigate their biosynthesis.

Herein, we report the characterization of a β-1,2-mannosyltransferase that uses DIP and GDP-mannose for the synthesis of MDIP; the same enzyme catalyzes a subsequent mannosylation reaction that results in the formation of MMDIP. A comparative genomic analysis led to the identification of the gene encoding these activities in the genome of *Thermotoga maritima*. Homologs with high sequence identity are present in other species of the genus *Thermotoga* (*Thermotoga neapolitana*, *Thermotoga naphthophila*, *Thermotoga petrophila,* and *Thermotoga* strain RQ2). Moreover, a gene with 36% identity was found in the genome of *Aquifex aeolicus*, and we confirmed that, despite the relatively low identity, the gene product exhibited MDIP synthase activity.

The enzymes implied in the synthesis of DIP, the substrate for MDIP synthase, also show a relatively low degree of sequence similarity (Chapter 3). For example, in *Thermotoga* spp. the relevant activities are encoded in two separate genes, while a fused gene is encountered in *Aquifex aeolicus* (Chapter 3); altogether, these findings suggest that the pathways for DIP and

MDIP synthesis diverged early during the evolution of these two bacterial genera.

Homologs of MDIP synthase are found only in the genera *Thermotoga*, *Aquifex, Archaeoglobus* and *Ferroglobus*, indicating that MDIP synthase is a singular, novel enzyme, thus far restricted to hyperthermophilic organisms. In line with this view, the thermophiles *Persephonella marina*, and *Hydrogenivirga* sp., members of the order *Aquificales*, lack a homologous enzyme, although they possess genes for the synthesis of DIP. The available results suggest a correlation between MDIP and hyperthermophily. It is curious, however, that MDIP synthase is absent within the most extreme hyperthermophiles, the archaeons *Pyrolobus fumarii* and *Pyrodictium occultum*, which accumulate DIP as a major solute (Fig. 5.8).

The MDIP synthase of *Thermotoga maritima* catalyzed not only the synthesis of MDIP, but also that of other DIP-derivatives containing two and three mannosyl groups linked via β-(1,2) glycosidic bonds, *i.e.*, MMDIP and minor amounts of the tri-mannosylated form. All the reactions involve an inverting mechanism in which the α-configuration of the substrate, GDP-mannose, was converted into the opposite anomeric configuration. Therefore, MDIP synthase is an inverting β-1,2-mannosyltransferase. The analysis of the amino acid sequence revealed no homology with any of the 91 glycosyltransferases families existing in the Carbohydrate Active Enzymes database (CAZy, www.cazy.org, (Cantarel et al. 2009)), (P. M. Coutinho, CAZy, personal communication), indicating that MDIP synthase is the first member of a new glycosyltransferase family. The number of known β-1,2-mannosyltransferases is extremely low and the CAZy database contains only one family of β-1,2-mannosyltransferases, comprising the enzymes implicated in the synthesis of glycans of glycoproteins in the cell wall of *Pichia pastoris* and *Candida albicans* (Mille et al. 2008). However, the amino acid sequences of

these β-1,2-mannosyltransferases and MDIP synthase are totally unrelated. Therefore, it appears that the MDIP synthase herein disclosed is a novel enzyme as no counterparts have been found thus far.

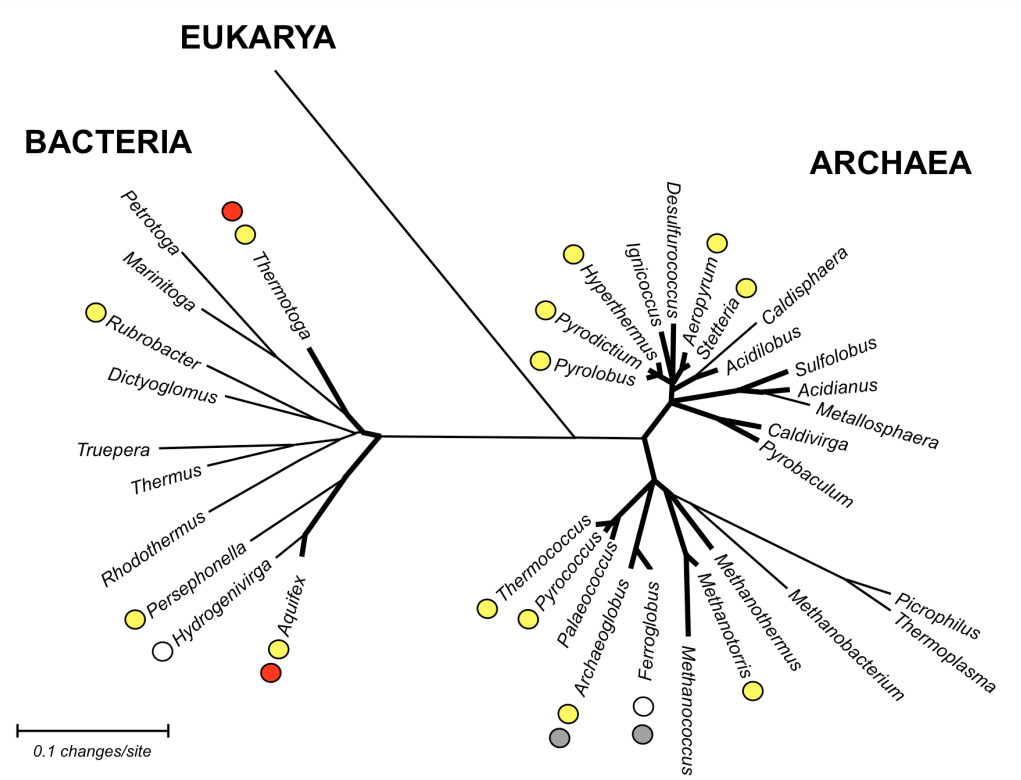


Figure 5.8. Unrooted phylogenetic tree based on 16S rRNA sequences of thermophilic (thin lines) and hyperthermophilic (thick lines) organisms, highlighting the distribution of DIP (yellow circles) and MDIP (red circles) (Empadinhas et al. 2006, Lamosa et al. 2006, Santos et al. 2007, and our unpublished results). The human 18S rRNA sequence was used as an outgroup. The white circle means that the genes for the synthesis of DIP are predicted but the accumulation of that solute has not been reported; the grey circle means that the MDIP synthase gene is predicted in *Archaeoglobus profundus* and *Ferroglobus placidus*, but the accumulation of MDIP has not been studied. The MEGA 4.0 software (Tamura et al. 2007) was used for sequence alignment and to generate the phylogenetic tree using the neighbor-joining method. Sequences of 16S and 18S rRNA are from the Ribosomal Database Project (http://rdp.cme.msu.edu/hierarchy/hb_intro.jsp).

The MDIP synthase of *Thermotoga maritima* has a high degree of specificity for GDP-mannose, as glycosyl donor, but showed some flexibility in regard to the mannosyl acceptor, using DIP, MDIP, and MMDIP as substrates. The relatively high K_m value (16 mM) for DIP correlates with the accumulation of this solute in the host organism; the intracellular concentration of DIP in *Thermotoga maritima* is 15-20 mM, calculated with the assumption that the intracellular volume of *Pyrococcus furiosus* is a reasonable approximation to that of *Thermotoga maritima* (Martins and Santos 1995). The experimental determination of the K_m value of MDIP synthase for MDIP was not possible due to substrate unavailability. The observation that MMDIP was detected only in cells grown under conditions leading to strong accumulation of MDIP (above 50 mM), suggested that the K_m value for MDIP should be considerably higher than that for DIP. Therefore, we deemed it interesting to estimate the K_m for MDIP using a kinetic model to describe the two reaction steps catalyzed by the enzyme (M represents GDP-mannose):

$$DIP + M \leftrightarrow MDIP$$
$$MDIP + M \leftrightarrow MMDIP$$

The model assumes Michaelis-Menten kinetics for both reactions and the establishment of a steady state condition in regard to MDIP, *i.e.*, the rate of MDIP formation was considered equal to the rate of MDIP consumption. Hence, the following relationship is derived:

$$\frac{V_{\max}^{DIP}}{V_{\max}^{MDIP}} = \frac{K_m^{DIP}}{K_m^{MDIP}} \times \frac{[MDIP]}{[DIP]} \qquad \text{(eq. 1)}$$

where, $V_{\max}^{DIP}$ and $V_{\max}^{DIP}$ designate the kinetic parameter V_{max} for DIP and MDIP, respectively.

With the reasonable assumption that the enzyme has equal V_{max} values for DIP and MDIP, the above relationship becomes:

$$\frac{K_m^{DIP}}{K_m^{MDIP}} = \frac{[DIP]}{[MDIP]} \qquad \text{(eq. 2)}$$

If the steady state condition is a good approximation in the cells, a K_m value of 60 mM for MDIP is estimated by introducing in equation 2 the DIP and MDIP concentrations determined in cells grown under heat stress (Table 5.2) and the K_m^{DIP} of 16 mM. The magnitude of the K_m^{MDIP} parameter explains the absence of MMDIP under optimal or osmotic stress conditions, when the intracellular concentration of MDIP is far below the estimated value.

It is interesting that the level of DIP in *Thermotoga maritima* did not change upon exposure to either heat stress or osmotic stress, in contrast with the observation that the DIP pool increases in response to heat stress in many of the hyperthermophiles examined, like *Pyrococcus furiosus*, *Archaeoglobus fulgidus*, *Aquifex pyrophilus* and *Methanotorris igneus* (Ciulla et al. 1994, Santos et al. 2007). The constancy of the DIP level is probably related with the use of this compound as an intermediate metabolite in the synthesis of MDIP and MMDIP, the solutes whose level increased notably under heat stress.

The presence of β-1,2-mannosides as components of the solute pool in hyperthermophilic organisms is surprising since the β-1,2-mannoside linkage is very rare in nature. This type of linkage has not been found in mammalian cells and, to our knowledge, has been reported only in a few organisms,

mostly pathogens: in glycoproteins of *Candida* spp., *Salmonella thompson*, and *Pichia pastoris* (Lamosa et al 2006, Mille et al. 2008, Suzuki et al. 1997), and in reserve oligosaccharides of *Leishmania* spp. (Ralton et al. 2003) and of the thermophilic bacterium *Truepera radiovitrix* (our unpublished results). The absence of β-1,2-mannosides in humans leads to the proposal that β–1,2-mannosyltransferases could be suitable targets for the development of drugs against infectious diseases.

The chemical composition of MDIP and MMDIP leads us to speculate about a connection with the polar heads in the glycolipids of *Thermotoga maritima*. Unfortunately, the information on the glycolipid composition of this bacterium is very limited and in particular phosphatidyl-*myo*-inositol mannosides have not been reported (Manca et al. 1992). This polar group has been detected in the cell wall of mycobacteria, but the stereochemistry is clearly distinct: the mannose moiety is bound via a α-1,2-linkage to the inositol phosphate group instead of a β-1,2-linkage found in MDIP and MMDIP; the inositol group found in mycobacteria has a D-configuration instead of the L-configuration found in the mannosylated forms of DIP (Kordulakova et al. 2002, Morita el al. 2004, Salman et al. 1999). Accordingly, the glycosyltransferases involved in the synthesis of the mycobacterial glycolipids are unrelated with the MDIP synthase of *Thermotoga maritima*.

In conclusion, this work led to the structural characterization of two β-1,2-mannosides accumulating in *Thermotoga maritima* in response to heat stress; these compounds are synthesized by the activity of a novel β-1,2-mannosyltransferase that uses DIP as a mannosyl acceptor, and appears to be the first member of a new family of glycosyltransferases. These unusual features fuel interest in the elucidation of the catalytic mechanism. Work is in progress to determine the crystallographic structure of this unique enzyme.

Acknowledgments and work contributions

C. P. Almeida cultivated *Thermotoga maritima* and performed the solute extractions. N. Borges analyzed the NMR data of the ethanolic extractions and searched the genome of *Thermotoga maritima* for the gene encoding MDIP synthase. C. Faria cloned and purified the recombinant MDIP synthase from *Aquifex aeolicus* used in the DSC experiments. P. Lamosa contributed to the identification of the novel solute acquiring and analyzing the one and two-dimensional NMR data. T. Q. Faria provided guidance on the DSC experiments and deduced the kinetic equations. Mass spectrometry data were obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, ITQB. The NMR spectrometers at CERMAX are part of the National NMR Network and were acquired with funds from FCT and FEDER. This work was funded by Fundação para a Ciência e a Tecnologia, POCTI Portugal, and FEDER Projects PTDC/BIA-MIC/71146/2006 and PTDC/BIO70806/2006. I thank Dr. Rita Ventura for the help with the nomenclature of MDIP and MMDIP. M. V. Rodrigues acknowledges FCT for the research fellowship.

CHAPTER 6

Final discussion

Chapter 6 – Contents

Inositol-containing compatible solutes in hyperthermophiles

Many hyperthermophiles are isolated from marine geothermal areas and it is assumed that the accumulation of compatible solutes allows them to cope with fluctuations in the osmolarity of the external medium. However, the accumulation of solutes also occurs in response to elevated temperature, supporting the view that compatible solutes of hyperthermophiles play a role in thermoadaptation.

Di-*myo*-inositol 1,3'-phosphate (DIP) is widely distributed among organisms thriving in hot environments (Fig. 6.1). In most hyperthermophiles, DIP accumulates primarily at supra-optimal growth temperatures, and has never been found in organisms thriving optimally below 60ºC.

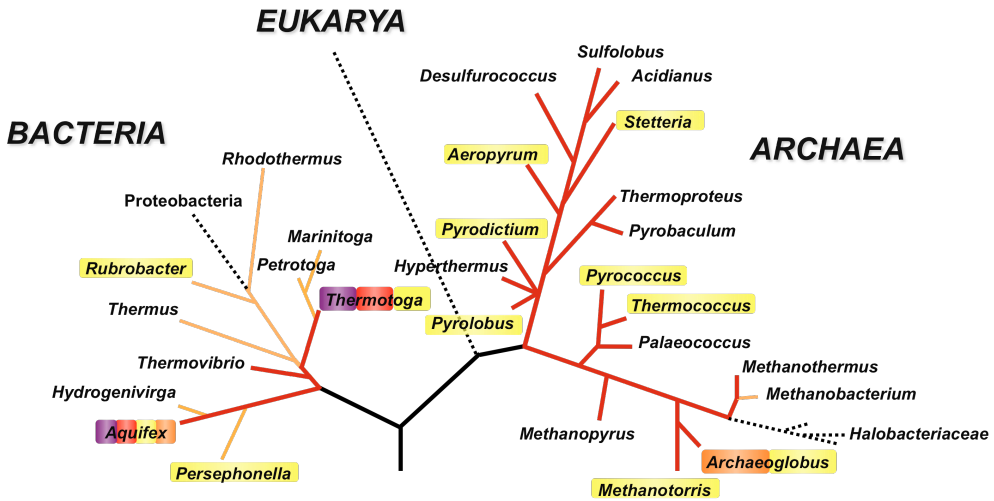


Figure 6.1. Distribution of inositol-containing solutes in the Tree of Life. Di-*myo*-inositol phosphate (yellow boxes), mannosyl-di-*myo*-inositol phosphate (red boxes), di-mannosyl-di-*myo*-inositol phosphate (purple boxes) and glycero-phospho-*myo*-inositol (orange boxes). Red and orange branches indicate hyperthermophiles and thermophiles, respectively.

The main goal of this thesis was to discover the biosynthetic pathways for DIP and other inositol-containing solutes. DIP was firstly identified in the archaeon *Pyrococcus woesei* in 1992 (Scholz et al. 1992). This solute is composed by two *myo*-inositol moieties linked through a phosphodiester bridge. van Leeuwen and coworkers (van Leeuwen et al. 1994) proposed that the phosphate group bridges the C_1 positions of the two *myo*-inositol residues, hence DIP should have an L,L-configuration. However, the biosynthetic pathway established for the synthesis of DIP in this thesis (Chapters 2 and 3) pointed towards a D,L-configuration of the inositol moieties in DIP, rather than the L,L configuration initially proposed. This intriguing inconsistency was definitely resolved by using labeled L-[1-^{13}C]*myo*-inositol 1-phosphate as a substrate for the bifunctional recombinant CTP:inositol 1-phosphate cytidylyltransferase/CDP-inositol:inositol 1-phosphate transferase (IPCT/DIPPS) from several hyperthermophilic archaea and bacteria (Chapter 4). By tracing the ^{13}C-label through the biosynthetic pathway we showed that the phosphate group linked C_1 of one of the inositol residues and C_3 of the other one (Fig. 6.2). Therefore, the designation di-*myo*-inositol 1,3'-phosphate is proposed here to highlight the fact that the phosphate group is linked to stereochemically distinct carbon atoms.

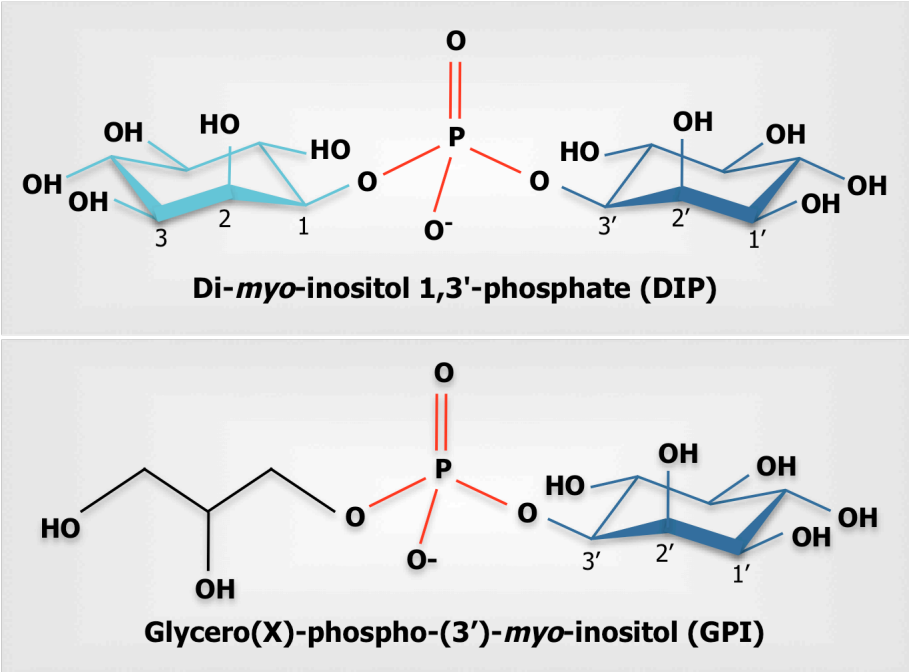


Figure 6.2. Structures of di-*myo*-inositol 1,3′-phosphate and glycero(X)-phospho-(3′)-*myo*-inositol. The L-numbering scheme is used for the nomenclature of the *myo*-inositol ring.

The same strategy allowed us to determine the configuration of the *myo*-inositol residue in glycero-phospho-*myo*-inositol (GPI). Using CDP-glycerol and L-[1-^{13}C]*myo*-inositol 1-phosphate as substrates for the same recombinant DIPPS, we concluded that the phosphate group was attached to C_3 of L-*myo*-inositol 1-phophate (Fig. 6.2).

By NMR analysis we showed that GPI accumulating in the bacterium *Aquifex pyrophilus* is different from GPI accumulating in the archaeon *Archaeoglobus fulgidus* but identical to GPI obtained from beef liver phospholipids (Chapter 4). The use of carbon-13 labeled L-*myo*-inositol 1-phosphate allowed us to confirm that the *myo*-inositol ring has the same configuration regardless of the bacterial or archaeal origin of the DIPPS used

to produce GPI (Chapter 4). Therefore, it is proposed that the difference in GPI´s from *Aquifex pyrophilus* and *Archaeoglobus fulgidus* relies on the configuration of the glycerol group, which is donated by CDP-glycerol.

The enantiomers of glycerol phosphate are generated from dihydroxyacetone phosphate (DHAP) and the chirality of the product is dictated by the stereospecificity of the dehydrogenase catalyzing the reaction (Fig. 6.3).

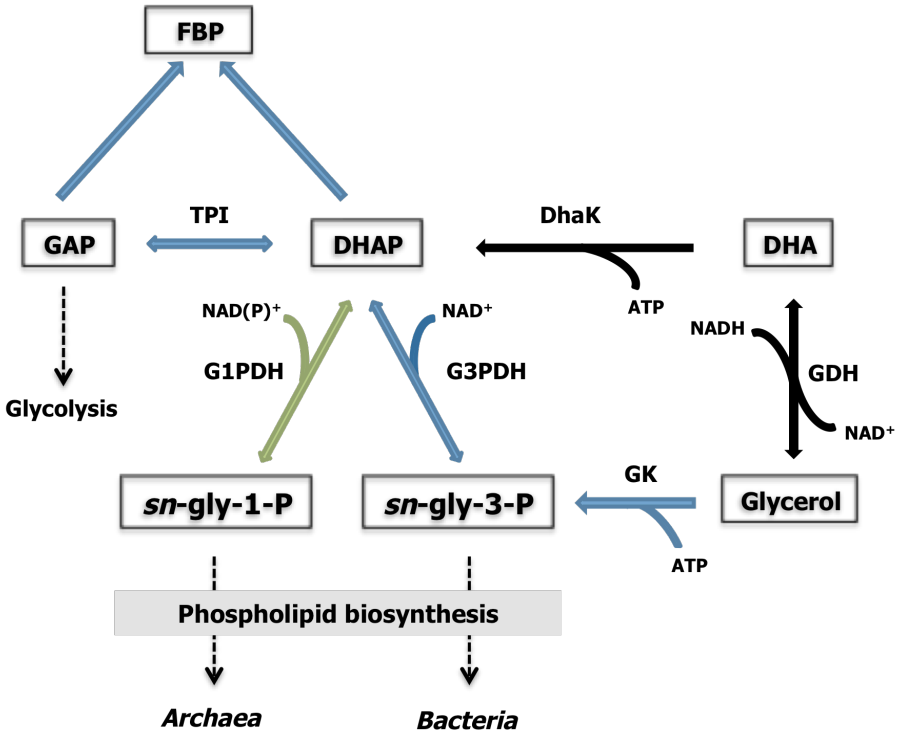


Figure 6.3. Biosynthesis of *sn*-glycerol 1-phosphate and *sn*-glycerol 3-phosphate. Colored arrows indicate that the genes encoding the respective enzymes were identified in members of the domain *Bacteria* (black), or of the domains *Archaea* and *Bacteria* (green), and the three domains of the Tree of Life (blue). FBP, fructose 1,6-bisphosphate; GAP, D-glyceraldehyde 3-phosphate; DHAP, dihydroxyacetone phosphate; DHA, dihydroxyacetone; *sn*-gly-1-P, *sn*-glycerol 1-phosphate; *sn*-glyecrol 3-phosphate; TPI, triosephosphate isomerase; DhaK, dihydroxyacetone kinase; GDH, glycerol dehydrogenase; GK, glycerol kinase; G3PDH, *sn*-glycerol 3-phosphate dehydrogenase; G1PDH, *sn*-glycerol 1-phosphate dehydrogenase. (Peretó et al. 2004, Koga and Morii 2007).

Currently, it is known that G3PDH is ubiquitous in *Bacteria*, but it is also annotated in the genomes of the archaea *Archaeoglobus fulgidus* and *Methanothermobacter thermoautotrophicus* (Peretó et al. 2004, Sakasegawa et al. 2004, Koga and Morii 2007). On the other hand, G1PDH is predominantly found in archaeal species, but homologous sequences have been predicted in the genome of the bacterium *Thermotoga maritima* and the activity confirmed in *Bacillus subtilis* (Peretó et al. 2004, Guldan et al. 2008). In summary, it seems that the occurrence of the specific enantiomers of glycerol phosphate is not restricted to a particular domain of Life. Hence, in order to complete the structural characterization of GPI it is necessary to determine the configuration of CDP-glycerol produced in the relevant organisms. We speculate that the CDP-glycerol pool in the archaeon *Archaeoglobus fulgidus* will be formed predominantly from *sn*-glycerol 1-phosphate since this is the sole enantiomer of glycerol phosphate present in the backbone of archaeal glycerophospholipids. However, we should take into account that the genome of *Archaeoglobus fulgidus* contains two genes coding for G3PDH, hence the presence of *sn*-glycerol 3-phosphate cannot be excluded. It is also interesting to note that GPI extracted from *Archaeoglobus fulgidus* gave rise to a minor ^{31}P-NMR resonance with a chemical shift coincident with that of GPI from *Aquifex* spp. or beef liver (Fig. 6.4). At this stage it is important to clarify the stereospecificity of the enzyme(s) that form(s) CDP-glycerol as well as that of DIPPS.

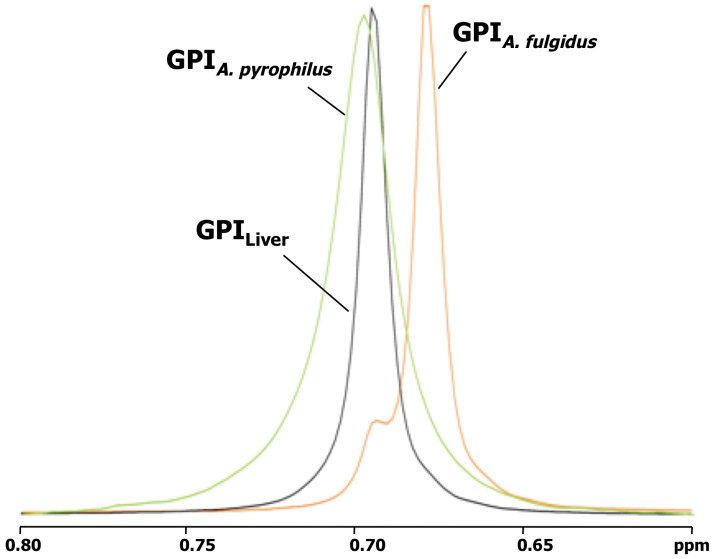


Figure 6.4. ^{31}P-NMR spectra of GPI partially purified from *Archaeoglobus fulgidus* extracts (orange spectrum), of GPI in an ethanolic extract of *Aquifex pyrophilus* (green spectrum), and of GPI derived from beef liver (black spectrum).

Biosynthetic pathways for DIP and GPI

Knowledge on the biosynthetic pathways of compatible solutes of hyperthermophiles is mandatory to understand the role played by these solutes in the mechanisms of thermoadaptation. When this work started, nothing was known about the synthesis of GPI, but the synthesis of DIP had been investigated in the late nineties by two independent groups. Chen and co-workers investigated the synthesis of this solute in *Methanocaldococcus igneus* (Chen et al. 1998) and proposed that *myo*-inositol 1-phosphate is synthesized from glucose 6-phosphate by *myo*-inositol 1-phosphate synthase and that part of this compound is dephosphorylated into *myo*-inositol while another part is presumably activated to CDP-inositol; subsequently both molecules are condensed to yield DIP. At the same time, Scholz and

colleagues (Scholz et al. 1998), proposed that the synthesis of DIP in *Pyrococcus woesei* involved the condensation of two molecules of *myo*-inositol 1-phosphate to yield DIP with the consumption of an NTP.

We investigated the synthesis of inositol-containing solutes in several hyperthermophiles (Chapters 2 and 3) and concluded that the synthesis of DIP involves solely L-*myo*-inositol 1-phosphate and CTP, *i.e.*, *myo*-inositol is not used as substrate, and no nucleoside triphosphate other than CTP can activate L-*myo*-inositol 1-phosphate.

We showed that the synthesis of DIP proceeds from glucose 6-phosphate via four steps: (1) glucose 6-phosphate was converted into L-*myo*-inositol 1-phosphate by L-*myo*-inositol 1-phosphate synthase; (2) L-*myo* inositol 1-phosphate was activated to CDP-inositol at the expense of CTP; (3) CDP-inositol was coupled with L-*myo*-inositol 1-phosphate to yield a phosphorylated intermediate, di-*myo*-inosityl 1,3'-phosphate 1'-phosphate (DIPP). For the first time, the synthesis of GPI was reported (Chapter 2).

GPI is synthesized similarly to DIP, except that the activated alcohol is CDP-glycerol instead of CDP-inositol. The synthesis proceeds as follows: CDP-glycerol is condensed with L-*myo*-inositol 1-phosphate yielding a phosphorylated form of GPI (GPIP), which is ultimately dephosphorylated into GPI. This was the first time that the involvement of phosphorylated intermediates in the synthesis of these compatible solutes was shown. These phosphorylated intermediates were purified and fully characterized by NMR (Chapter 4).

The pool of solutes in *Archaeoglobus fulgidus* includes a third polyol phosphodiester, diglycerol phosphate (DGP), which accumulates preferentially in response to osmotic stress (Gonçalves et al. 2003). The pathway for the synthesis of this rare compatible solute has some features in common with the synthesis of DIP and GPI (Fig. 6.5): i) in all cases, CTP is the only nucleoside

triphosphate used to activate the polyol, either glycerol or inositol, forming CDP-glycerol or CDP-inositol, respectively; ii) the polyol acceptor must be phosphorylated, *i.e.*, L-glycerol 3-phosphate or L-*myo*-inositol 1-phosphate; iii) the central step involves the condensation of CDP-polyol with the phosphorylated polyol to form a phosphorylated intermediate.; and iv) finally, the dephosphorylation of this intermediate leads to the final product, *i.e.*, diglycerol phosphate, glycero-phospho-*myo*-inositol or di-*myo*-inositol phosphate.

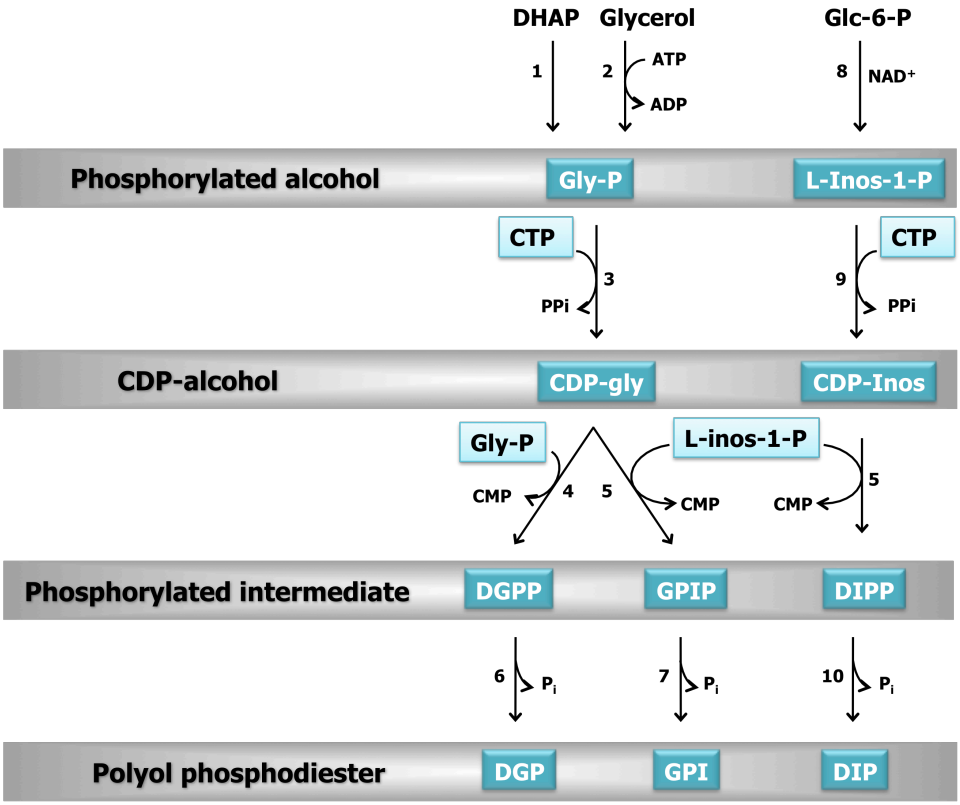


Figure 6.5. Similarities between the biosynthetic pathways for the three polyol phosphodiesters accumulating in *Archaeoglubus fulgidus*, DIP, GPI and DGP. Enzymes: (1) glycerol phosphate dehydrogenase; (2) glycerol kinase; (3) CTP:glycerol phosphate cytidylyltransferase, (4) DGPP synthase, (5) CDP-inositol:L-*myo*-inositol 1-phosphate transferase (DIPPS), (6) DGPP phosphatase, (7) GPIP phosphatase, (8)

L-*myo*-inositol 1-phosphate synthase, (9) CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase (IPCT), and (10) DIPP phosphatase. DGP, diglycerol phosphate; DGPP, X,X'-diglycerol phosphate X'-phosphate; DIP, di-*myo*-inositol 1,3'-phosphate, DIPP, di-*myo*-inositol 1,3'-phosphate 1'-phosphate; GPI, glycero(X)-phospho-(3')-*myo*-inositol; GPIP, (X-glycerol)-*myo*-inositol 3'-phosphate 1'-phosphate; DHAP, dihydroxyacetone phosphate; Glc-6-P, glucose 6-phosphate; CDP-gly, CDP-glycerol; CDP-Inos, CDP-L-*myo*-inositol; L-Inos-1-P, L-*myo*-inositol 1-phosphate; and Gly-P, glycerol phosphate (Borges et al. 2006, Chapter 2, and Peretó et al. 2004).

Interestingly, the scheme proposed for the synthesis of GPI and DIP shares some common aspects with the two-step pathway for the synthesis of sugar related compatible solutes, which involves the concerted action of a synthase and a phosphatase. Indeed, the formation of a phosphorylated intermediate is a well-known strategy for the synthesis of trehalose, sucrose, glucosylglycerol, or glucosylglycerate, solutes commonly found in mesophiles (Fig. 6.6) (Costa et al. 2006, Luley-Goedl and Nidetzky 2011). In the two-step pathway, a phosphorylated form of the product is formed by the action of a glycosyltransferase, which uses an NDP-sugar donor and a phosphorylated acceptor as substrates. Then a specific phosphatase converts the phosphorylated intermediate into the final product. The same strategy has been described for the synthesis of mannosylglycerate, a compatible solute also encountered in hyperthermophiles, and for the rare solute mannosylglucosylglycerate, that until now is restricted to the genus *Petrotoga* (Empadinhas et al. 2003, Fernandes et al. 2010). The concerted action of a synthase and a phosphatase, which ultimately results in the irreversible synthesis of the final product, may be the result of the evolutionary selection of an efficient strategy to enable the accumulation of high levels of compatible solutes in the cytoplasm.

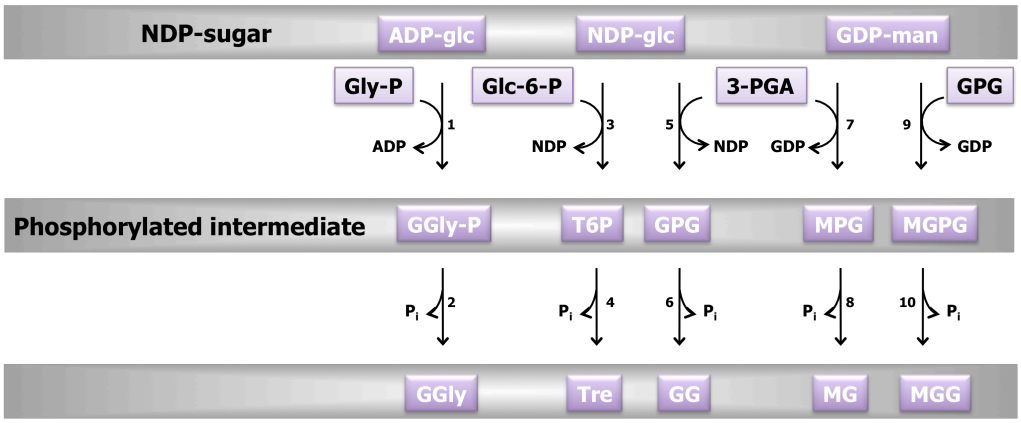


Figure 6.6. Concerted action of a synthase and a phosphatase in the biosynthesis of hexose derivatives. Enzymes: (1) glucosylglycerol phosphate synthase, (2) glucosylglycerol phosphate phosphatase, (3) trehalose phosphate synthase, (4) trehalose phosphate phosphatase, (5) glucosyl 3-phosphoglycerate synthase, (6) glucosyl 3-phosphoglycerate phosphatase, (7) mannosyl 3-phosphoglycerate synthase, (8) mannosyl 3-phosphoglycerate phosphatase, (9) mannosyl-glucosyl 3-phosphoglycerate synthase, (10) mannosyl-glucosyl 3-phosphoglycerate phosphatase. ADP-glc, ADP-glucose; NDP-glc, NDP-glucose; GDP-man, GDP-mannose; gly-P, glycerol 3-phosphate; glc-6-P, glucose 6-phosphate; 3-PGA, 3-phosphoglycerate; GPG, glucosyl 3-phosphoglycerate; T6P, trehalose 6-phosphate; MPG, mannosyl 3-phosphoglycerate; MGPG, mannosyl-glucosyl 3-phosphoglycerate; GGly-P, glucosylglycerol phosphate; Tre, trehalose; GG, glucosylglycerate; MG, mannosylglycerate; and MGG, mannosylglucosylglycerate (Costa et al. 2006, Elbein et al. 2003, Fernandes et al. 2010, Martins et al. 1999, Hagemann and Erdmann 1994).

However, there is an additional route for the synthesis of hexose derivatives involving a single-step reaction. In this pathway, a glycosyltransferase converts an NDP-sugar and a relevant non-phosphorylated acceptor into the final product (Fig. 6.7).

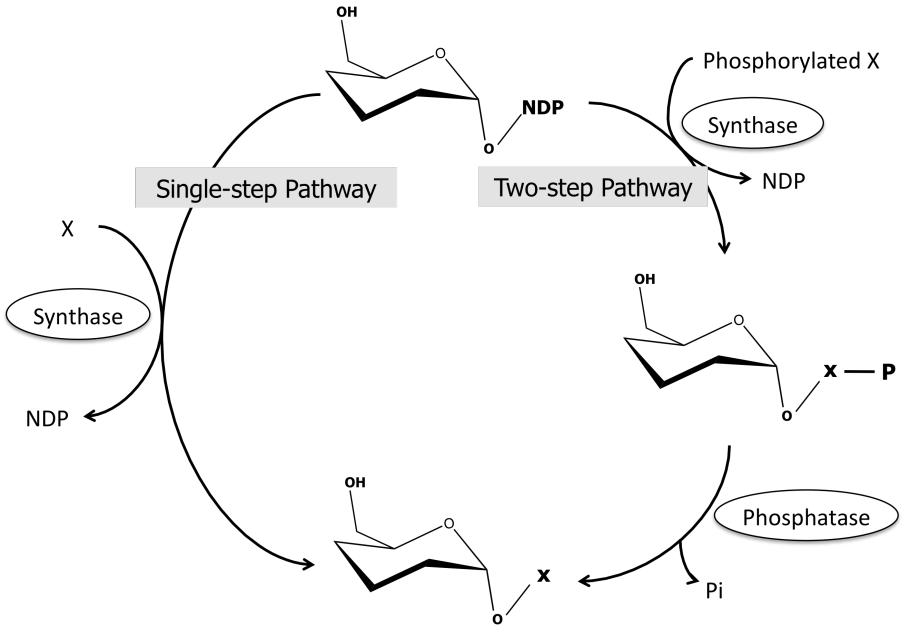


Figure 6.7. Multiple pathways for the synthesis of hexose derivatives. X, substrate.

The synthesis of sucrose, mannosylglycerate, glucosylglycerate, and mannosylglucosylglycerate can occur via either phosphorylating or non-phosphorylating routes (Curatti et al. 1998, 2000, Martins et al. 1999, Costa et al. 2006, Fernandes et al. 2007, Fernandes et al. 2010). The synthesis of trehalose is a particular case since five routes exist for the synthesis of this solute and only one involves a phosphorylated intermediate metabolite (Avonce et al. 2006). Thus far, only the synthase-phosphatase pathway has been described for the synthesis of glucosylglycerol (Hageman and Erdmann 1994). Interestingly, the enzymes involved in the synthesis of glucosylglycerol in *Pseudomonas* sp. and in *Stenotrophomonas rhizophila*, are encoded in a single gene. The bifunctional enzyme has a glycosyltransferase activity and a phosphatase activity in the C-terminal and N-terminal domains, respectively (Hagemann et al. 2008).

Recently, the elucidation of the pathway for the synthesis of archaetidyl-*myo*-inositol in *Methanothermobacter thermoautotrophicus* revealed that the synthesis of this phospholipid is different from that of phosphatidylinositol (eukaryal and bacterial) (Fig. 6.8) (Morii et al. 2009). While phosphatidylinositol is synthesized from CDP-diacylglycerol and free *myo*-inositol in a single step reaction, the synthesis of archaetidyl-*myo*-inositol proceeds from CDP-archaeol and L-*myo*-inositol 1-phosphate in a two-step reaction that resembles the synthesis of DIP (Fig. 6.8). Firstly, a phosphorylated intermediate is formed, archaetidyl-*myo*-inositol 1-phosphate, which is subsequently dephosphorylated into the final product, archaetidyl-*myo*-inositol (Morii et al. 2009). The synthesis of the same phosphorylated intermediate was detected in the membrane fraction of the archaeon *Pyrococcus furiosus.* The enzyme involved in this process is archaetidyl-*myo*-inositol phosphate synthase. The respective gene was identified in the genome of *Methanothermobacter thermoautotrophicus* and the activity confirmed by functional expression in *E. coli*. Homologous sequences are found in the genomes of *Sulfolobus solfataricus*, *Aeropyrum pernix, Haloferax volcanii*, and many other *Archaea*. Interestingly, a similar pathway was recently described for the synthesis of phosphatidylinositol in mycobacteria (Morii et al. 2010, Morita et al. 2011). Apparently, this group of bacteria can synthesize phosphatydylinositol by two alternative pathways, with or without the formation of a phosphorylated intermediate (Fig. 6.8). Thus far, only the phosphorylating route has been demonstrated for the synthesis of DIP (Chapters 2 and 3).

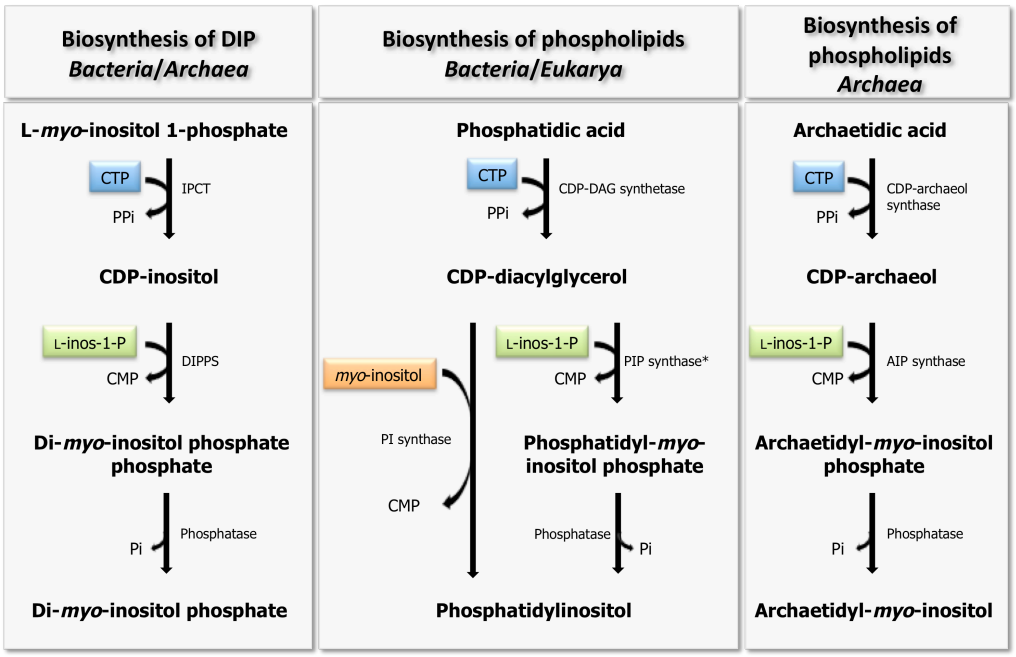


Figure 6.8. Similarities between the biosynthesis of phosphatidylinositol, archaetidyl-*myo*-inositol, and di-*myo*-inositol phosphate. CDP-DAG synthetase, CDP-diacylglycerol synthetase; PIP synthase, phosphatidylinositol phosphate synthase; PI synthase, phosphatidylinositol synthase; AIP synthase, archaetidyl-*myo*-inositol phosphate synthase; IPCT, CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase; DIPPS, CDP-inositol:L-*myo*-inositol 1-phosphate transferase. *Described in mycobacteria only. (Kent 1995, Chapter 3, Rodionov et al. 2007, Morii et al. 2000, 2009, Morita et al. 2011).

Although DIP was discovered nearly 20 years ago, the genes and enzymes directly involved in the synthesis of the solute were unknown. In this work a genomic approach was used for the identification of the genes involved in DIP biosynthesis (Chapter 3). Genes encoding IPCT (CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase) and DIPPS (CDP-inositol:L-*myo*-inositol 1-phosphate transferase) from *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Thermococcus kodakarensis*, *Aquifex aeolicus* and *Rubrobacter xylanophilus* were uncovered by functional expression in *E. coli*. In parallel,

the work by Rodionov and co-workers (Rodionov et al. 2007) confirmed the activity of the *ipct* and *dipps* genes from *Thermotoga maritima* and *Aeropyrum pernix*.

The DIPPS domain contains a conserved motif characteristic of the CDP-alcohol phosphatidyltransferase family (pfam01066). The members of this family have the ability to catalyze the displacement of CMP from a CDP-alcohol by the attack of a second alcohol, with formation of a phosphodiester bond. Most of these enzymes are involved in phospholipid metabolism, such as diacylglycerol cholinephosphotransferase (cholinephosphotransferase), CDP-diacylglycerol-serine-phosphatidyltransferase (phosphatidylserine synthase), CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase (phosphatidylglycerolphosphate synthase), or CDP-diacylglycerol-inositol 3-phosphatidyltransferase (phosphatidylinositol synthase). The latter enzyme catalyzes the reaction between *myo*-inositol and CDP-diacylglycerol, resulting in the formation of phosphatidyl-*myo*-inositol with the displacement of CMP. It is noteworthy the resemblance between the unit formed by the glycerol backbone and the phospholipid headgroup, with the compatible solute GPI (Fig. 6.9).

A similar observation can be made in respect with the couple diglycerol phosphate and phosphatidylglycerol. It has been postulated that the accumulation of GPI and other polyolphosphodiesters, could be related to the synthesis of these particular membrane components, but obviously this is not the case since it is presently known that distinct enzymes catalyze the synthesis of compatible solutes and phospholipids. However, it is conceivable that DIPP synthase has evolved from phosphatidylinositol synthase. It would be interesting to check whether the DIPP synthase is able to recognize CDP-diacylglycerol as substrate.

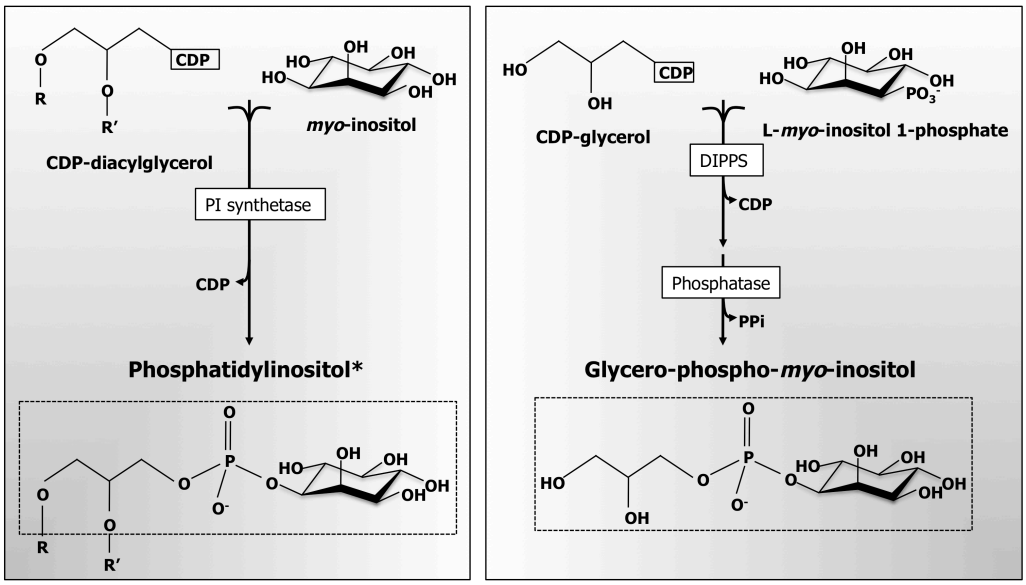


Figure 6.9. Comparison between the polar head group of phosphatidyl-*myo*-inositol and glycero-phospho-*myo*-inositol. DIPPS, CDP-inositol:L-*myo*-inositol 1-phosphate transferase; PI synthetase, phosphatidyinositol synthetase. R and R′, fatty acids. *In the domains *Eukarya*, *Bacteria* and *Archaea* the stereochemistry of the polar head group of archaetidyl- and/or phosphatidylinositol is the same regardless of the biosynthetic pathway (see Fig. 6.8).

Recently, the crystal structure of IPCT (N-terminal domain of the bifunctional enzyme IPCT/DIPPS) from *Archaeoglobus fulgidus* was determined (Brito et al. 2011). The authors showed that IPCT is barely related with other cytidyltransferases recognizing polyol-phosphates such as glycerol phosphate, ribitol phosphate or methylerythritol-phosphate. Instead, the enzyme shares some structural homology with pyrophosphorylases, namely with the N-terminal domain of the bifunctional N-acetylglucosamine 1-phosphate uridylyltransferases from *Streptococcus pneunominae*, *Escherichia coli* or *Vibrio cholerae*, and with glucose 1-phosphate thymidylyl-/uridylyl-transferases from *Corynebacterium glutamicum*, *Pseudomonas aeruginosa*, or *Escherichia coli*. Moreover, the high specificity of IPCT for the nucleotide CTP

that we have shown in Chapter 3 could be explained by Brito and co-workers as being due to the presence of a specific tight loop that allows the formation of an hydrogen bond between the NH_2 of the cytidine and the main chain oxygen of a proline residue (Brito et al. 2011). Unfortunately the crystal structure of the C-terminal domain, *i.e.*, DIPPS domain, is still missing, probably due to the difficulty inherent to expressing, purifying and crystallizing membrane proteins in a functional form. It would be interesting to resolve the structure of the whole protein in order to determine whether CDP-inositol is channeled from the N-terminal to the C-terminal domain of the protein, or instead if it is released and scavenged subsequently by the DIPPS domain. The channeling of metabolites is acknowledged as an efficient strategy utilized by hyperthermophiles to cope with elevated temperatures and make it possible the use of thermolabile substrates in metabolic reactions. Whether this holds true for CDP-inositol remains uncertain.

The wealth of genomic data currently available to the scientific community allows for the prediction of genes encoding IPCT and DIPPS by sequence homology. Genes with high similarity have been found in *Pyrococcus abysii*, *Hyperthermus butylicus*, several species of the genus *Thermococcus*, *Hydrogenivirga* or *Thermobaculum terrenum* (Chapter 3). In most cases the two activities are fused in a single gene product, resulting in a bifunctional enzyme, IPCT/DIPPS, with cytidilyltransferase activity in the N-terminal domain (IPCT), and CDP-alcohol phosphatidyltransferase activity in the C-terminal domain (DIPPS or DIPP synthase). However, in *Aeropyrum pernix*, *Hyperthermus butylicus* and *Thermotoga* spp. two separate genes are found.

Evolutionary studies of the IPCT/DIPPS are now possible using the several sequences available for the generation of phylogenetic trees. From the IPCT/DIPPS based unrooted tree two main clusters emerged (Fig. 6.10). One cluster is predominantly composed by the *Euryarchaeota* archaeal sequences,

with the inclusion of the bacterial sequences from *Hydrogenivirga* and *Aquifex aeolicus*. All these IPCT/DIPPS sequences are the result of fused *ipct/dipps* genes. The other main cluster is composed by species with two separate genes encoding IPCT and DIPPS, and includes the bacterial IPCT/DIPPS of *Thermotoga* spp. and the archaeal *Crenarchaeota*, *Aeropyrum pernix* and *Hyperthermus butylicus*. The location of the IPCT/DIPPS from *Thermobaculum terrenum* and *Rubrobacter xylanophilus* is intriguing; apparently it is positioned closer to the cluster comprising species where separate genes are found, but in these bacteria the two activities are encoded in a single gene. Also intriguing is the clustering between the bacteria *Persephonella marina* and the *Korarchaeota Candidatus Korarchaeum cryptophilum*, since they seem to be poorly related to any of the main clusters. In the 16S RNA based phylogenetic tree it is clear the dichotomy between the archaeal species and the bacterial species, however this separation is not apparent in the IPCT/DIPPS based tree; some archaeal sequences are located in branches predominantly populated by bacteria and vice versa (Fig. 6.10).

The occurrence of lateral gene transfer events among these organisms is highly likely, which could explain the mixed location of IPCT/DIPPS from *Archaea* and *Bacteria*. Lateral gene transfer events occurring among prokaryotes, namely (hyper)thermophiles, are supported by several pieces of evidence (Koonin et al. 2001, Robb and Newby 2007). For example, the genomes of *Aquifex aeolicus* and *Thermotoga maritima* are predicted to have 16% and 24% of genes that were acquired from hyperthermophilic archaea, while only 7% of the genes in *Bacillus subtilis* are similar to archaea (Aravind 1998, Nelson et al. 1999). A careful bioinformatic analysis would be needed to uncover the evolutionary history of the metabolic pathways leading to DIP synthesis.

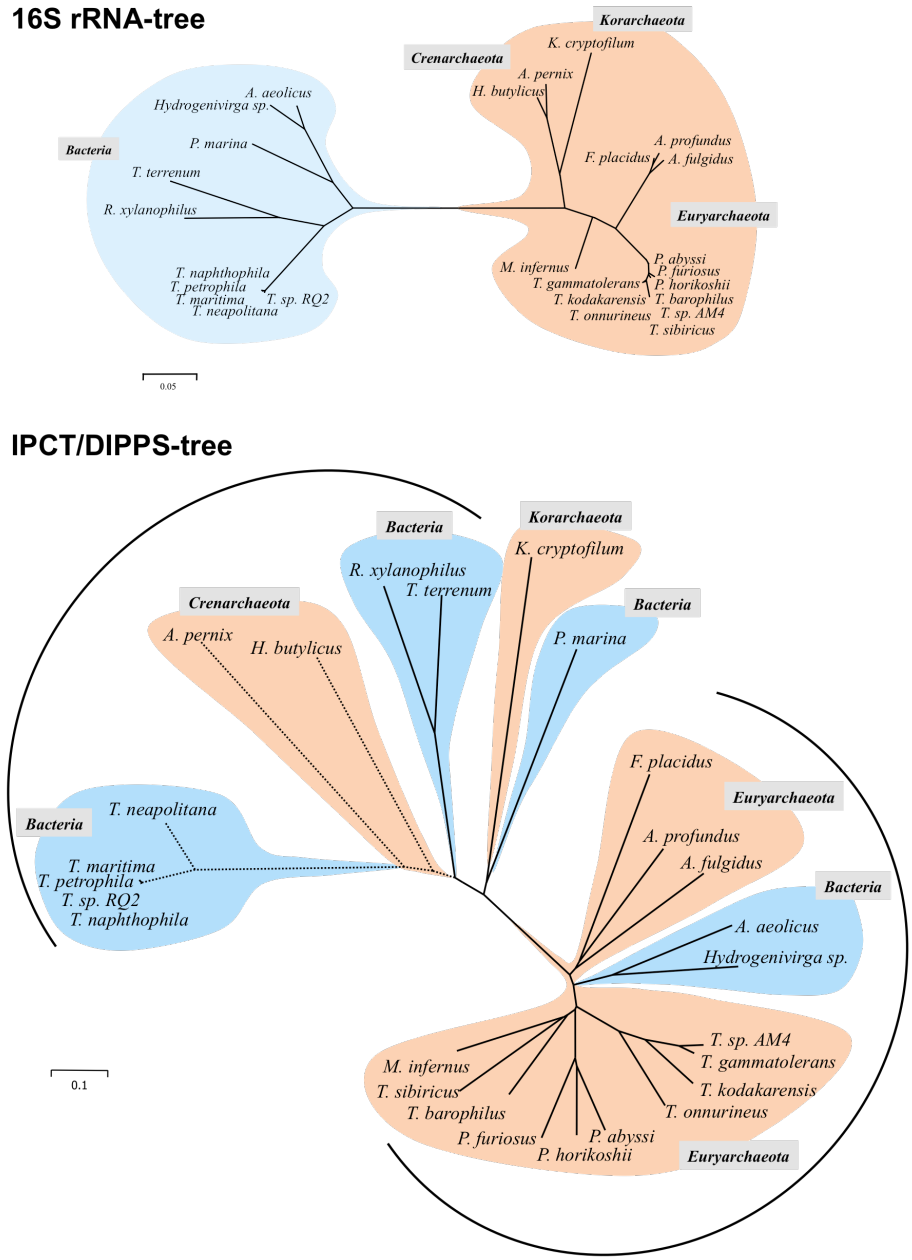


Figure 6.10. Unrooted phylogenetic tree based on 16S-rRNA sequences (upper panel), and amino acid sequences of IPCT/DIPPS (lower panel). The ClustalX program (Thompson et al. 1997) was used for sequence alignments. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary

distances used to infer the phylogenetic tree. Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007). *A. aeolicus*, *Aquifex aeolicus*, *Hydrogenivirga* sp., *Hydrogenivirga* sp. 128-5-R1-1, *P. marina*, *Persephonella marina*, *R. xylanophilus*, *Rubrobacter xylanophilus*, *T. terrenum*, *Thermobaculum terrenum*, *T. naphtophila*, *Thermotoga naphthophila*, *T. maritima*, *Thermotoga maritima*, *T. petrophila*, *Thermotoga petrophila*, *T. neapolitana*, *Thermotoga neapolitana*, *A. fulgidus*, *Archaeoglobus fulgidus*, *A. profundus*, *Archaeoglobus profundus*, *T. gammatolerans*, *Thermococcus gammatolerans*, *T. kodakarensis*, *Thermococcus kodakarensis*, *T. onnurineus*, *Thermococcus onnurineus*, *T. sibiricus*, *Thermococcus sibiricus*, *T. barophilus*, *Thermococcus barophilus*, *P. horikoshii*, *Pyrococcus horikoshii*, *P. abyssi*, *Pyrococcus abyssi*, *P. furiosus*, *Pyrococcus furiosus*, *F. placidus*, *Ferroglobus placidus*, *M. infernus*, *Methanocaldococcus infernus*, *H. butylicus*, *Hyperthermus butylicus*, *K. cryptofilum*, *Candidatus Korarchaeum cryptofilum*, *A. pernix*, *Aeropyrum pernix*. Dotted lines designate species where IPCT/DIPPS is the product of two separate genes. Blue, *Bacteria*; orange, *Archaea*.

Mannosylated derivatives of DIP

The solute pool of *Thermotoga neapolitana* was investigated and two new DIP derivatives were recognized (Martins et al. 1996). These solutes were tentatively identified as a symmetric di-mannosylated derivative of DIP and as a DIP stereoisomer. Therefore, we set out to fully elucidate the pool of solutes related to *myo*-inositol in *Thermotoga* spp.. A high-filed NMR spectrometer allowed us to conclude that the DIP related compatible solutes accumulating in *Thermotoga maritima* were in fact a mono-mannosylated and an asymmetric di-mannosylated compound, herein definitely identified as mannosyl-di-*myo*-inositol phosphate (MDIP) and di-mannosyl-di-*myo*-inositol phosphate (MMDIP), respectively (Fig. 6.11). The levels of MDIP and MMDIP increased in response to growth at supra-optimal temperatures.

The synthesis of MDIP involves the transfer of the mannosyl group from GDP-mannose to DIP in a single-step reaction catalyzed by a glycosyltransferase (Chapter 5). The gene for mannosyl-di-*myo*-inositol

phosphate synthase (*mds* gene) was identified in the genome of *Thermotoga maritima* and *Aquifex aeolicus* and the activity was confirmed by functional expression in *E. coli* (Chapter 5). The recombinant enzyme uses DIP and GDP-mannose for the synthesis of MDIP. This enzyme does not use the phosphorylated intermediate of DIP, in disagreement with the frequent biosynthetic scheme for polyolphosphodiesters and hexose derivatives. The same enzyme can use MDIP as an acceptor of a second mannose residue, yielding the di-mannosylated derivative of DIP, MMDIP (Chapter 5).

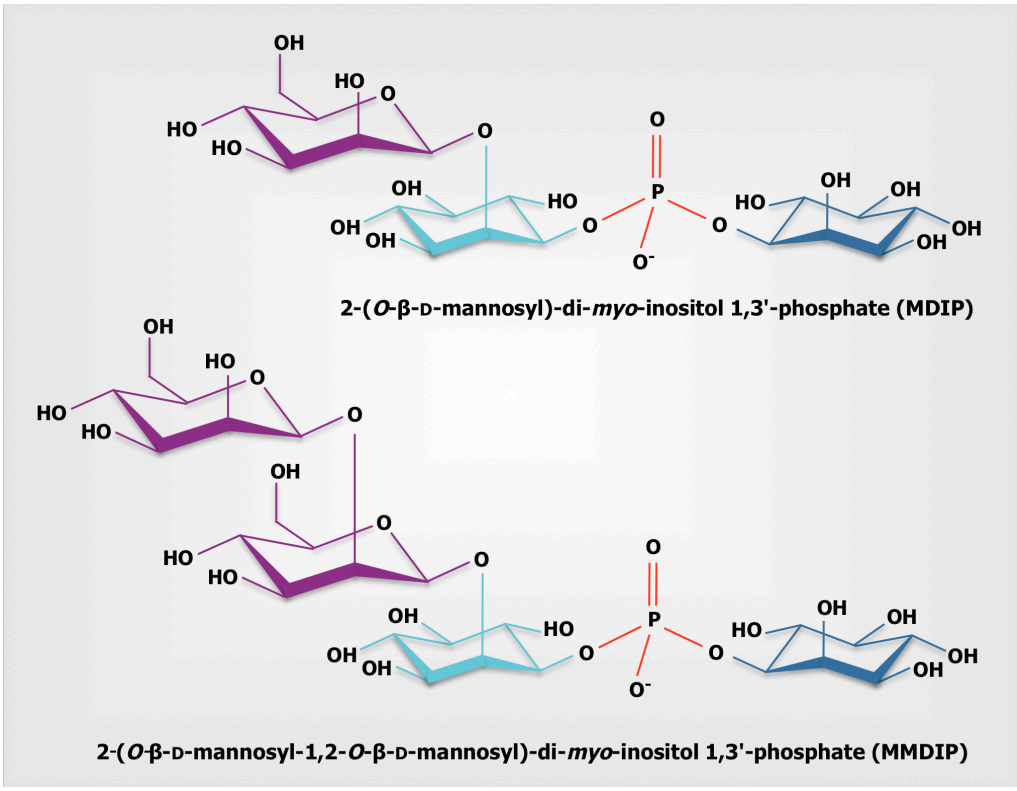


Figure 6.11. Structures of 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate and 2-(*O*-β-D-mannosyl-1,2-*O*-β-D-mannosyl)-di-*myo*-inositol 1,3′-phosphate.

MDIP synthase is a β-1,2-mannosyltransferase, unrelated with known glycosyltransferases. Apparently this enzyme is the first member of a new glycosyltransferase family and this justifies investing in its production for structural characterization purposes (work in progress in the lab). Within the domain *Bacteria*, it is restricted to members of the two deepest lineages, *i.e.*, the *Thermotogales* and the *Aquificales*. A homolog of MDIP synthase was identified in the genome of *Archaeoglobus profundus*. Apparently this enzyme is rare and restricted to hyperthermophiles, but the absence of the *mds* gene in the majority of hyperthermophiles accumulating DIP is an intriguing observation.

Mannosylated forms of phosphatidyl-*myo*-inositol (phosphatidylmannosides) can be found in the plasma membrane and/or the cell wall envelope of mycobacteria, and are essential components for growth and virulence of the pathogenic species. The major phosphatidyl-*myo*-inositol mannosides have two or six mannose residues and are usually tri-acylated as a result of the transfer of a fatty acid residue to the C_6 of one of the core mannose residues.

The mannosylated polar head group of phosphatidyl-*myo*-inositol mannosides resembles part of the molecule of mannosyl-di-*myo*-inositol phosphate, although with a different stereochemistry: in phosphatidyl-*myo*-inositol mannosides, the mannose residue is linked by an α-(1-2) linkage to the *myo*-inositol that is bound to phosphate in the 3 position (using the L-numbering scheme), while in MDIP the mannose residue is linked by an β-(1-2) linkage to the *myo*-inositol that is bound to phosphate in the 1 position (using the L-numbering scheme) (Fig 6.12). Accordingly, the glycosyltransferases involved in the synthesis of phosphatidylmannosides are unrelated to MDIP synthase of *Thermotoga maritima*. The first step in the mannosylation of phosphatidyl-*myo*-inositol is catalyzed by an α-mannosyltransferease (PimA) which mediates the transfer of a mannose

residue from GDP-mannose to the C_2 of the *myo*-inositol ring of phopshatidylinositol. Scarce information is available about the composition of the phospholipids of the membrane of *Thermotoga maritima*, and in particular the presence of phosphatidylinositol or phosphatidilinositol mannosides is unknown.

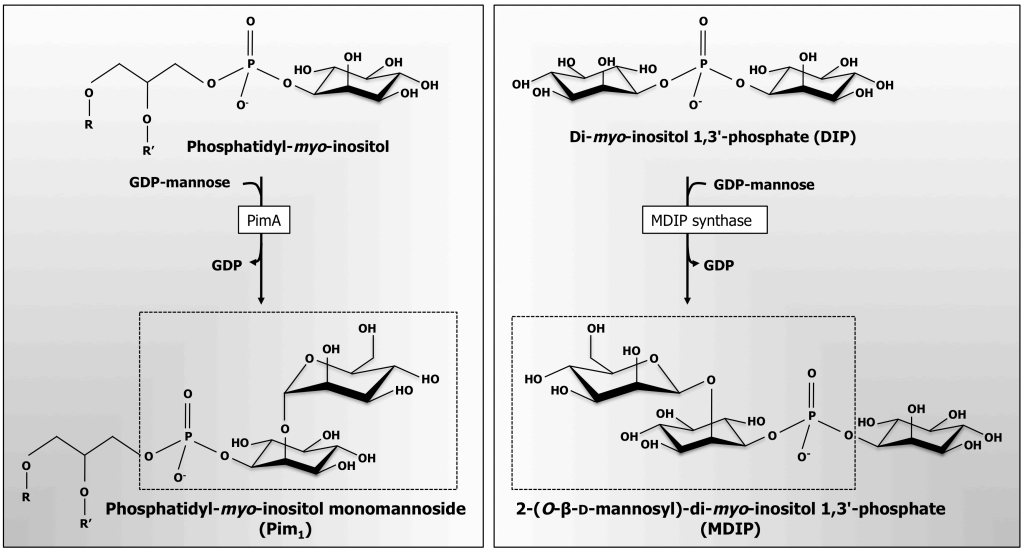


Figure 6.12. Similarities between phosphatidylinositol monomannoside (Pim_1) occurring in mycobacteria, and 2-(*O*-β-D-mannosyl)-di-*myo*-inositol 1,3'-phosphate (MDIP) accumulated in *Thermotoga* spp. and *Aquifex* spp.. PimA, α-D-mannose-α(1-2)-phosphatidyl-*myo*-inositol transferase; MDIP synthase, mannosil-di-*myo*-inositol 1,3'-phosphate synthase.

β-1-2 mannosidic linkages are uncommon in Nature and have only been reported in a few bacterial and yeast species. Most of these species are pathogenic and include *Leishmania* spp., *Candida* spp., and *Salmonella thompson*, but this type of linkage has also been identified in the non-pathogenic yeast *Pichia Pastoris* (Ralton et al. 2003, Vinogradov et al. 2000, Susuki et al. 1997, Shibata et al. 1985, Mille et al. 2008). β-1-2 mannans

occur in the cell wall of *Candida* spp. participating in the adhesion process and are recognized as antigenic factors (Shibata et al. 1992, Susuki et al. 1997, Mille et al. 2008). In *Leishmania* spp. these structures function as a major carbohydrate reserve (Ralton et al. 2003). To our knowledge β-1-2 linked oligosaccharides have not been found in mammalian cells and therefore the detection of such an antigen is useful for the diagnostics of fungal infections. Furthermore, it has been observed that the administration of synthetic β-1-2 oligomannosides reduces the colonization by *Candida* spp. of the intestinal track of mice, whereas α-1-2 oligomannosides did not (Dromer et al. 2002). This suggests that β-1-2 mannosides contribute very specifically to the virulence of these pathogens and thus they are interesting candidates to study as potential drug targets.

Transport, regulation and physiological role of inositol-containing solutes

The accumulation of compatible solutes for osmo- and/or thermoadaptation can be regulated at the level of transport systems and/or at the level of the *de novo* synthesis. Little is known about the regulation of the synthesis of compatible solutes in hyperthermophiles (Santos et al. 2007). The synthesis of mannosylglycerate is differentially regulated at the level of protein expression: the two-step pathway is activated under salt stress conditions, whereas the single-step pathway is enhanced under heat stress (Borges et al. 2004). However, nothing is known about the regulation of the synthesis of the inositol-containing compatible solutes.

We obtained preliminary results on the heat-shock response of *Thermotoga maritima* in respect to the accumulation of compatible solutes

and transcription profiles of genes encoding enzymes involved in the synthesis of DIP and MDIP (Appendix 1). Heat-shock led to an increase in the accumulation of α-glutamate, DIP, MDIP, and MMDIP. The transcriptional levels of the genes *mpg1*, *mds*, *suhB*, and *mips*, encoding, respectively, the enzymes mannose-1-phosphate guanylyltransferase, mannosyl-DIP synthase, *myo*-inositol-1-monophosphatase, *myo*-inositol 1-phosphate synthase, are essentially unchanged upon heat-shock. On the other hand, the genes *ipct* and *dipps*, encoding respectively CTP:*myo*-inositol 1-phosphate cytidylyltransferase, and CDP-inositol:*myo*-inositol 1-phosphate transferase, are down-regulated. Our approach was not the optimal one to study simultaneously compatible solute levels and mRNA levels and we think that a quantitative method should be used to determine the transcript levels. Most likely the synthesis of DIP and MDIP is regulated by the activity of the respective enzymes, or at a translational level. For this matter, a proteomics approach should be followed.

An alternative to the *de novo* synthesis of compatible solutes is their up-take from the environment, which in fact is energetically more favorable. In many cases, the presence of specific transport systems functions as a solution for leaky membranes and allow for the rescue of lost solutes from the exterior medium. Nothing is known about the transport of DIP and derivatives across the cellular membrane due to unavailability of radiolabeled solutes. The discovery of the biosynthetic genes and enzymes undertaken in the present thesis will enable the production of ^{14}C-labelled solutes, such as DIP, GPI and MDIP.

In fact, knowledge on solute uptake in the *Archaea* is restricted mainly to methanogens. *Methanohalophilus portucalensis*, *Methanosarcina thermophila* or *Methanosarcina mazei* are a few examples of organisms in which glycine betaine transporters have been identified (Proctor et al. 1997,

Lai et al. 2000, Roeβler et al. 2002). Genomic analysis revealed potential candidates for primary transporters of glycine betaine in the genomes of *Methanococcoides burtonii*, *Archaeoglobus fulgidus*, *Methanococcus maripalidus* and other *Methanosarcina* spp. (Müller et al. 2005). In *Methanosarcina* spp. the genes encoding potential glycine betaine transporters are regulated by osmolarity, which may be directly related to the accumulation of glycine betaine in hypersalinity conditions (Müller et al. 2005). Recently, a DIP-deficient mutant was constructed using the efficient gene disruption system available for *Thermococcus kodakarensis* (Borges et al. 2010). This mutant efficiently replaced DIP for aspartate under heat stress conditions. Surprisingly, the absence of DIP did not affect the growth rate of the mutant strain in comparison to the parental strain. Nevertheless, the results suggest that the adaptation of *Thermococcus kodakarensis* to heat-stress depends on the stabilizing/protecting effects provided by the accumulation of negatively charged organic solutes.

Final considerations

The results gathered in this thesis on the accumulation and biosynthesis of inositol-containing solutes definitely extended our knowledge on the physiology of hyperthermophiles. The discovery of the genes involved in the biosynthetic pathways represents an important step towards the ultimate goal of deciphering the *in vivo* role of these exquisite compatible solutes. Moreover, this information is crucial to the design of metabolic strategies envisaging the production of these compatible solutes in suitable industrial hosts, such as yeast, *Escherichia coli* or *Corynebacterium glutamicum*.

To obtain a global description of the heat stress response it is essential to understand the cascade of events starting with sensing stress, proceeding to the pathways of signal transduction and finally the activation of the biosynthetic genes (Fig. 6.13). This work paves the way to elucidate the regulatory mechanisms underlying solute accumulation in response to stress. A yet unsolved issue is the catabolism of DIP and derivatives. We assume that MDIP and MMDIP are acted upon by an unknown mannosidase that releases the mannose residue, which can then be further metabolized, but the fate of DIP is undetermined.

The scarcity of genetic tools is still a major bottleneck to the advance of knowledge on the physiology of hyperthermophiles. The archaea *Sulfolobus solfataricus* and *Thermococcus kodakarensis* are the only hyperthermophiles for which a limited number of genetic manipulations have become feasible in recent years. Therefore, a great research effort should be directed to develop genetic tools for manipulation of *Archaeoglobus fulgidus* and *Thermotoga maritima*, two organisms adapted to thrive at high temperature, which accumulate an interesting palette of compatible solutes.

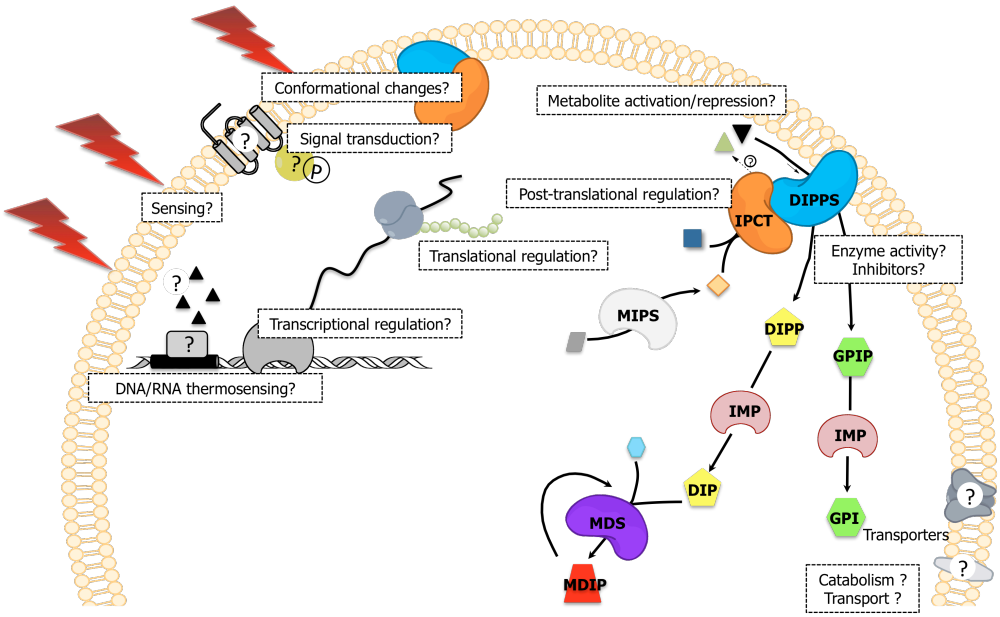


Figure 6.13. Schematic overview of the biosynthetic pathways for DIP, GPI, MDIP, and MMDIP disclosed in this thesis. The stress sensing mechanisms and signal transduction pathways, the regulation of the biosynthetic pathways, and the catabolism of these solutes remain elusive. MIPS, L-*myo*-inositol 1-phosphate synthase; IPCT, CTP:L-*myo*-inositol 1-phosphate cytidylyltransferase; DIPPS, CDP-inositol:L-*myo*-inositol 1-phosphate transferase; IMP, inositol monophosphatase; MDS, mannosyl-di-*myo*-inositol 1,3′-phosphate synthase; DIP, di-*myo*-inositol 1,3′-phosphate; DIPP, di-*myo*-inositol 1,3′-phosphate 1'-phosphate; GPI, glycero(X)-phospho(3′)-*myo*-inositol; GPIP, glycero(X)-phospho(3′)-*myo*-inositol(1′)-phosphate; and MDIP, mannosyl-di-*myo*-inositol 1,3′-phosphate. Symbols: grey ■, glucose 6-phosphate; orange ◆, L-*myo*-inositol 1-phosphate; dark blue ■, CTP; black ▼, CDP-glycerol; green ▲, CDP-L-*myo*-inositol; light blue ⬣, GDP-mannose.

References

Achenbach-Richter L, Stetter KO, Woese CR (1987) A possible biochemical missing link among archaebacteria. Nature **327**:348-349.

Agranoff BW (1978) Cyclitol confusion. Trends Biochem Sci **3**:N283–N285.

Agranoff BW (2009) Turtles All the Way: Reflections on *myo*-Inositol. J Biol Chem **284**:21121-21126.

Antranikian G, Vorgias CE, Bertoldo C (2005) Extreme environments as a resource for microorganisms and novel biocatalysts. Adv Biochem Eng Biotechnol **96**:219-262.

Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. Trends Genet **14**:442-444.

Arnott MA, Michael RA, Thompson CR, Hough DW, Danson MJ (2000) Thermostability and thermoactivity of citrate synthases from the thermophilic and hyperthermophilic archaea, *Thermoplasma acidophilum* and *Pyrococcus furiosus*. J Mol Biol **304**:657-668.

Atomi H, Matsumi R, Imanaka T (2004) Reverse gyrase is not a prerequisite for hyperthermophilic life. J Bacteriol **186**:4829-4833.

Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G (2006) Insights on the evolution of trehalose biosynthesis. BMC Evol Biol **6**:109.

Badurina DS, Zolli-Juran M, Brown ED (2003) CTP:glycerol 3-phosphate cytidylyltransferase (TarD) from *Staphylococcus aureus* catalyzes the cytidylyl transfer via an ordered Bi-Bi reaction mechanism with micromolar $K_{(m)}$ values. Biochim Biophys Acta **1646**:196-206.

Balch WE, Fox GE, Magrum LJ, Woese CR, Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. Microbiol Rev **43**:260-296.

Barns SM, Delwiche CF, Palmer JD, Pace NR (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A **93**:9188-9193.

Beeder J, Nilsen RK, Rosnes JT, Torsvik T, Lien T (1994) *Archaeoglobus fulgidus* isolated from hot North Sea oil field waters. Appl Environ Microbiol **60**:1227-1231.

Birrien JL, Zeng X, Jebbar M, Cambon-Bonavita MA, Querellou J, Oger P, Bienvenu N, Xiao X, Prieur D (2011) *Pyrococcus yayanosii* sp. nov.,the first obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol. [Epub ahead of print].

Blochl E, Rachel R, Burggraf S, Hafenbradl D, Jannasch HW, Stetter KO (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.

Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM (2008) Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr Opin Biotechnol **19**:210-217.

Bock K, Pedersen C (1974) A study of ^{13}CH coupling constants in hexopyranoses. J Chem Soc Perkin II **3**:293-297.

Boonyaratanakornkit BB, Simpson AJ, Whitehead TA, Fraser CM, El-Sayed NM, Clark DS (2005) Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock. Environ Microbiol **7**:789-797.

Boonyaratanakornkit BB, Miao LY, Clark DS (2007) Transcriptional responses of the deep-sea hyperthermophile *Methanocaldococcus jannaschii* under shifting extremes of temperature and pressure. Extremophiles **11**:495-503.

Borges N, Marugg JD, Empadinhas N, da Costa MS, Santos H (2004) Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of *Rhodothermus marinus*. J Biol Chem **279**:9892-9898.

Borges N, Goncalves LG, Rodrigues MV, Siopa F, Ventura R, Maycock C, Lamosa P, Santos H (2006) Biosynthetic pathways of inositol and glycerol phosphodiesters used by the hyperthermophile *Archaeoglobus fulgidus* in stress adaptation. J Bacteriol **188**:8128-8135.

Borges N, Matsumi R, Imanaka T, Atomi H, Santos H (2010) *Thermococcus kodakarensis* mutants deficient in di-*myo*-inositol phosphate use aspartate to cope with heat stress. J Bacteriol **192**:191-197.

Bouveng H, Lindberg B, Wickberg B (1955) Low-molecular carbohydrates in algae. Structure of the glyceric acid mannoside from red algae. Acta Chem Scand **9**: 807–809.

Bowers KJ, Wiegel J (2011) Temperature and pH optima of extremely halophilic archaea: a mini-review. Extremophiles **15**:119-128.

Bradford MM (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.

Bredberg K, Persson J, Christiansson M, Stenberg B, Holst O (2001) Anaerobic desulfurization of ground rubber with the thermophilic archaeon *Pyrococcus furiosus*: a new method for rubber recycling. Appl Microbiol Biotechnol **55**:43-48.

Breitmaier E, Voelter W (1989) Carbon-13 NMR spectroscopy: high-resolution methods and applications in organic chemistry and biochemistry. p. 401. 3rd Ed., VCH, Weinheim, Germany.

Bremer E, Kramer R (2000) Coping with osmotic challanges: osmoregulation through accumulation and release of compatible solutes in Bacteria. p. 79-96. *In* G. Storz and R. Hengge-Aronis (*ed.*), Bacterial stress responses. ASM Press, Washington, DC.

Brito JA, Borges N, Vonrhein C, Santos H, Archer M (2011) Crystal structure of *Archaeoglobus fulgidus* CTP:inositol-1-phosphate cytidylyltransferase, a key enzyme for di-*myo*-inositol-phosphate synthesis in (hyper)thermophiles. J Bacteriol **193**:2177-2185.

Brochier-Armanet C, Forterre P (2007) Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. Archaea **2**:83-93.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. Nat Rev Microbiol **6**:245-252.

Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F, Forterre P (2005) *Nanoarchaea*: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to *Thermococcales*? Genome Biol **6**:R42.

Brock TD (1967) Life at high temperatures. Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed. Science **158**:1012-1019.

Brock TD, Brock KM, Belly RT, Weiss RL (1972) *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch Mikrobiol **84**:54-68.

Brock TD (1985) Life at high temperatures. Science **230**:132-138.

Brockerhoff H, Hanahan DJ (1959) Studies on naturally occurring phosphoinositides. J Am Chem Soc **81**:2591-2592.

Brown AD, Simpson JR (1972) Water relations of sugar-tolerant yeasts: the role of intracellular polyols. J Gen Microbiol **72**:589-591.

Brown AD (1976) Microbial water stress. Bacteriol Rev 40:803-846.

Burgess EA, Wagner ID, Wiegel J (2007) Thermal environments and biodiversity. p. 13-29. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res **37**:D233-D238.

Cary CS, Skank TM, Stein JL (1998) Worms bask in extreme temperatures. Nature **391**:545- 546.

Cavicchioli R (2002) Extremophiles and the search for extraterrestrial life. Astrobiology **2**:281-292.

Chen L, Spiliotis ET, Roberts MF (1998) Biosynthesis of di-*myo*-inositol-1,1'-phosphate, a novel osmolyte in hyperthermophilic archaea. Journal of Bacteriology **180**:3785-3792.

Chen L, Zhou C, Yang H, Roberts MF (2000) Inositol-1-phosphate synthase from *Archaeoglobus fulgidus* is a class II aldolase. Biochemistry **39**:12415-12423.

Chhabra SR, Shockley KR, Conners SB, Scott KL, Wolfinger RD, Kelly RM (2003) Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J Biol Chem **278**:7540-7552.

Ciulla RA, Burggraf S, Stetter KO, Roberts MF (1994) Occurrence and role of di-*myo*-Inositol-1,1'-phosphate in *Methanococcus igneus*. Appl Environ Microbiol **60**:3660-3664.

Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM (2006) Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. FEMS Microbiol Rev **30**:872-905.

Costa J, Empadinhas N, Goncalves L, Lamosa P, Santos H, da Costa MS (2006) Characterization of the biosynthetic pathway of glucosylglycerate in the archaeon *Methanococcoides burtonii*. J Bacteriol **188**:1022-1030.

Cowan DA (2004) The upper temperature for life – where do we draw the line? Trends Microbiol **12**:58-60.

Curatti L, Folco E, Desplats P, Abratti G, Limones V, Herrera-Estrella L, Salerno G (1998) Sucrose-phosphate synthase from *Synechocystis* sp. strain PCC 6803: identification of the spsA gene and characterization of the enzyme expressed in *Escherichia coli*. J Bacteriol **180**:6776-6779.

Curatti L, Porchia AC, Herrera-Estrella L, Salerno GL (2000) A prokaryotic sucrose synthase gene (*susA*) isolated from a filamentous nitrogen-fixing cyanobacterium encodes a protein similar to those of plants. Planta **211**:729-735.

da Costa MS, Santos H, Galinski EA (1998) An overview of the role and diversity of compatible solutes in *Bacteria* and *Archaea*. Adv Biochem Eng Biotechnol **61**:117-153.

Daiyasu H, Hiroike T, Koga Y, Toh H (2002) Analysis of membrane stereochemistry with homology modeling of *sn*-glycerol-1-phosphate dehydrogenase. Protein Eng **15**:987-995.

Damste JS, Rijpstra WI, Hopmans EC, Schouten S, Balk M, Stams AJ (2007) Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order *Thermotogales*. Arch Microbiol **188**:629-641.

Daniel RM, Cowan DA (2000) Biomolecular stability and life at high temperatures. Cell Mol Life Sci **57**:250-264.

Das F, Gerstein M (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison. Funct Integr Genomics **1**:76-88.

de Macario EC, Macario AJL (1994) Heat-shock response in *Archaea*. Trends Biotechnol **12**:512-518.

Di Giulio M (2003) The universal ancestor and the ancestor of bacteria were hyperthermophiles. J Mol Evol **57**:721-730.

Di Giulio M (2011) The last universal common ancestor (LUCA) and the ancestors of archaea and bacteria were progenotes. J Mol Evol **72**:119-126.

Dorr C, Zaparty M, Tjaden B, Brinkmann H, Siebers B (2003) The hexokinase of the hyperthermophile *Thermoproteus tenax*: ATP-dependent hexokinases and ADP-dependent glucokinases, two alternatives for glucose phosphorylation in Archaea. J Biol Chem **278**:18744-18753.

Dreef CE, Douwes M, Elie CJ, Marel GA, Boom JH (1991) Application of the bifunctional phosphorylating agent bis[6-(trifluoromethyl)benzotriazol-1-yl] methylphosphonate towards the preparation of isosteric D-*myo*-inositol phospholipid and phosphate analogues. Synthesis **1991**:443-447.

Driessen AJM, Albers SV (2007) Membrane adaptations of (hyper)thermophiles to high temperatures. p. 104-116. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Dromer F, Chevalier R, Sendid B, Improvisi L, Jouault T, Robert R, Mallet JM, Poulain D (2002) Synthetic analogues of β-1,2 oligomannosides prevent intestinal colonization by the pathogenic yeast *Candida albicans*. Antimicrob Agents Chemother **46**:3869-3876.

Dutta A, Chaudhuri K (2010) Analysis of tRNA composition and folding in psychrophilic, mesophilic and thermophilic genomes: indications for thermal adaptation. FEMS Microbiol Lett **305**:100-108.

Egorova K, Antranikian G (2005) Industrial relevance of thermophilic *Archaea*. Curr Opin Microbiol **8**:649-655.

Elbein AD, Pan YT, Pastuszak I, Carroll D (2003) New insights on trehalose: a multifunctional molecule. Glycobiology **13**:17R-27R.

Elcock AH (1998) The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. J Mol Biol **284**:489-502.

Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of the *Archaea*. Proc Natl Acad Sci U S A **105**:8102-8107.

Empadinhas N, Marugg JD, Borges N, Santos H, da Costa MS (2001) Pathway for the synthesis of mannosylglycerate in the hyperthermophilic archaeon *Pyrococcus horikoshii*, Biochemical and genetic characterization of key enzymes. J Biol Chem **276**:43580-43588.

Empadinhas N, Albuquerque L, Henne A, Santos H, da Costa MS (2003) The bacterium *Thermus thermophilus*, like hyperthermophilic archaea, uses a two-step pathway for the synthesis of mannosylglycerate. Appl Environ Microbiol **69**:3272-3279.

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

Empadinhas N, da Costa MS (2008) Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. Int Microbiol **11**:151-161.

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

Faria TQ, 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.

Fernandes C, Empadinhas N, da Costa MS (2007) Single-step pathway for synthesis of glucosylglycerate in *Persephonella marina*. J Bacteriol **189**:4014-9.

Fernandes C, Mendes V, Costa J, Empadinhas N, Jorge C, Lamosa P, Santos H, da Costa MS (2010) Two alternative pathways for the synthesis of the rare compatible solute mannosylglucosylglycerate in *Petrotoga mobilis*. J Bacteriol **192**:1624-1633.

Fong DH, Yim VC, D'Elia MA, Brown ED, Berghuis AM (2006) Crystal structure of CTP:glycerol-3-phosphate cytidylyltransferase from *Staphylococcus aureus*: examination of structural basis for kinetic mechanism. Biochim Biophys Acta **1764**:63-69.

Forterre P (1996) A hot topic: the origin of hyperthermophiles. Cell **85**:789-792.

Galinski EA, Trüper HG (1994) Microbial behavior in salt-stressed ecosystems. FEMS Microbiol Rev **15**:95-108.

Gallazzini M, Burg MB (2009) What's new about osmotic regulation of glycerophosphocholine. Physiology (Bethesda) **24**:245-249.

Galtier N, Lobry JR (1997) Relationships between genomc G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J Mol Evol **44**:632-636.

Giaever HM, Styrvold OB, Kaasen I, Strom AR (1988) Biochemical and genetic characterization of osmoregulatory trehalose synthesis in *Escherichia coli*. J Bacteriol **170**:2841-2849.

Glansdorff N (1999) On the origin of operons and their possible role in evolution toward thermophily. J Mol Evol **49**:432-438.

Glansdorff N, Xu Y, Labedan B (2008) The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol Direct **3**:29.

Goncalves LG, Huber R, da Costa MS, Santos H (2003) A variant of the hyperthermophile *Archaeoglobus fulgidus* adapted to grow at high salinity. FEMS Microbiol Lett **218**:239-244.

Gonçalves LG (2008) Osmo- and thermoadaptation in hyperthermophilic *Archaea*: identification of compatible solutes, accumulation profiles, and biosynthetic routes in *Archaeoglobus* spp. Ph.D. Thesis, Instituto de Tecnologia Química e Biológica, Portugal.

Goncalves LG, Lamosa P, Huber R, Santos H (2008) Di-*myo*-inositol phosphate and novel UDP-sugars accumulate in the extreme hyperthermophile *Pyrolobus fumarii*. Extremophiles **12**:383-389.

Goude R, Renaud S, Bonnassie S, Bernard T, Blanco C (2004) Glutamine, glutamate, and α-glucosylglycerate are the major osmotic solutes accumulated by *Erwinia chrysanthemi* strain 3937. Appl Environ Microbiol **70**:6535-6541.

Grabowski B, Kelman Z (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. Annu Rev Microbiol **57**:487-516.

Gribaldo S, Brochier-Armanet C (2006) The origin and evolution of *Archaea*: a state of the art. Philos Trans R Soc Lond B Biol Sci **361**:1007-1022.

Grogan DW (1989) Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J Bacteriol **171**:6710-6719.

Groisillier A, Herve C, Jeudy A, Rebuffet E, Pluchon PF, Chevolot Y, Flament D, Geslin C, Morgado IM, Power D, Branno M, Moreau H, Michel G, Boyen C, Czjzek M (2010) MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. Microb Cell Fact **9**:45.

Grosjean H, Oshima T (2007) How nucleic acids cope with high temperature. p. 39-56. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Gu X, Chen M, Wang Q, Zhang M, Wang B, Wang H (2005) Expression and purification of a functionally active recombinant GDP-mannosyltransferase (PimA) from *Mycobacterium tuberculosis* H37Rv. Protein Expr Purif **42**:47-53.

Guisbert E, Yura T, Rhodius VA, Gross CA (2008) Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. Microbiol Mol Biol Rev **72**:545-554.

Guldan H, Sterner R, Babinger P (2008) Identification and characterization of a bacterial glycerol-1-phosphate dehydrogenase: Ni(2+)-dependent AraM from *Bacillus subtilis*. Biochemistry **47**:7376-7384.

Gutmann DAP, Mizohata E, Newstead S, Ferrandon S, Henderson PJF, Van Veen HW, Byrne B (2007) A high-throughput method for membrane protein solubility screening: the ultracentrifugation sedimentation assay. Protein Sci **16**:1422-1428.

Hagemann M, Effmert U, Kerstan T, Schoor A, Erdmann N (2001) Biochemical characterization of glucosylglycerol-phosphate synthase of *Synechocystis* sp. strain PCC 6803: comparison of crude, purified, and recombinant enzymes. Curr Microbiol **43**:278-283.

Hagemann M, Erdmann N (1994) Activation and pathway of glucosylglycerol synthesis in the cyanobacterium Synechocystis sp. PCC 6803. Microbiol **140**:1427-1431.

Hagemann M, Ribbeck-Busch K, Klahn S, Hasse D, Steinbruch R, Berg G (2008) The plant-associated bacterium *Stenotrophomonas rhizophila* expresses a new enzyme for the synthesis of the compatible solute glucosylglycerol. J Bacteriol **190**:5898-5906.

Hanahan DJ, Dittmar JC, Warashima E (1957) A column chromatographic separation of classes of phospholipids. J Biol Chem **228**:685-700.

Heine M, Chandra SB (2009) The linkage between reverse gyrase and hyperthermophiles: a review of their invariable association. J Microbiol **47**:229-234.

Hensel R, Konig H (1988) Thermoadaptation of methanogenic bacteria by intracellular ion concentration. FEMS Microbiol Lett **49**:75-79.

Holden JF, Adams MWW, Baross J A (1999) Heat-shock response in hyperthermophilic microorganisms. *In* C. R. Bell, M. Brylinsky, P. Johnson-Green (ed.) Microbial biosystems, New frontiers: proceedings of the 8th International Symposium on Microbial Ecology. Atlantic Canada Society for Microbial Ecology, Halifax, Canada.

Horlacher R, Xavier KB, Santos H, Diruggiero J, Kossmann M, Boos W (1998) Archaeal binding protein-dependent ABC transporter: molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. J Bacteriol **180**:680-689.

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO (2002) A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. Nature **417**:63-67.

Huber R, Langworthy TA, König H, Thomm M, Woese CR, Sleytr UB, Stetter KO (1986) *Thermotoga maritima* sp. nov. representes a new genus of unique extremely thermophilic eubacteria growing up to 90ºC. Arch Microbiol **144**:324-333.

Huber R, Eder W, Heldwein S, Wanner G, Huber H, Rachel R, Stetter KO (1998) *Thermocrinis ruber* gen. nov., sp. nov., A pink-filament-forming hyperthermophilic bacterium isolated from yellowstone national park. Appl Environ Microbiol **64**:3576-3583.

Huber H, Stetter KO (1998) Hyperthermophiles and their possible potential in biotechnology. J Biotechnol **64**:39-52.

Huber R, Huber H, Stetter KO (2000) Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. FEMS Microbiol Rev **24**:615-623.

Hurme R, Berndt KD, Normark SJ, Rhen M (1997) A proteinaceous gene regulatory thermometer in Salmonella. Cell **90**:55-64.

Hurme R, Rhen M (1998) Temperature sensing in bacterial gene regulation-what it all boils down to. Mol Microbiol **30**:1-6.

Ikezawa H (2002) Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol Pharm Bull **25**:409-417.

Irvine RF, Schell MJ (2001) Back in the water: the return of the inositol phosphates. Nat Rev Mol Cell Biol **2**:327-338.

Johnson DB (2007) Physiology and ecology of acidophilic microorganisms. p. 257-271. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Jorge CD, 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.

Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L (2010) Metabolomic and transcriptomic stress response of *Escherichia coli*. Mol Syst Biol **6**:364.

Kanai T, Takedomi S, Fujiwara S, Atomi H, Imanaka T (2010) Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. J Biochem **147**:361-370.

Kanal H, Kobayashi T, Aono R, Kudo T (1995) *Natronococcus amylolyticus* sp. nov., a haloalkaliphilic archaeon. Int J Syst Bacteriol **45**:762-766.

Karsten U, Barrow KD, Mostaert AS, King RJ, West JA (1994) ^{13}C- and ^{1}H-NMR studies on digeneaside in the red alga *Caloglossa leprieurii*: a re-evaluation of its osmotic significance. Plant Physiol Biochem **32**:669–676

Kashefi K, Lovley DR (2003) Extending the upper temperature limit for life. Science **301**:934.

Kent C, Carman GM, Spence MW, Dowhan W (1991) Regulation of eukaryotic phospholipid metabolism. FASEB J 5:2258-2266.

Kent C (1995) Eukaryotic phospholipid biosynthesis. Annu Rev Biochem **64**:315-343.

Kets EP, Galinski EA, de WM, de Bont JA, Heipieper HJ (1996) Mannitol, a novel bacterial compatible solute in *Pseudomonas putida* S12. J Bacteriol **178**:6665-6670.

Kiewietdejonge A, Pitts M, Cabuhat L, Sherman C, Kladwang W, Miramontes G, Floresvillar J, Chan J, Ramirez RM (2006) Hypersaline stress induces the turnover of phosphatidylcholine and results in the synthesis of the renal osmoprotectant glycerophosphocholine in *Saccharomyces cerevisiae*. FEMS Yeast Res **6**:205-217.

Klahn S, Hagemann M (2011) Compatible solute biosynthesis in cyanobacteria. Environ Microbiol **13**:551-562.

Klinkert B, Narberhaus F (2009) Microbial thermosensors. Cell Mol Life Sci **66**:2661-2676.

Koga Y, Morii H (2007) Biosynthesis of ether-type polar lipids in archaea and evolutionary considerations. Microbiol Mol Biol Rev **71**:97-120.

Kollman VH, Hanners JL, London RE, Adame EG, Walker TE (1979) Photosynthetic preparation and characterization of ^{13}C-labeled carbohydrates in *Agmenellum quadruplicatum*. Carbohydr Res **73**:193-202.

Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol **55**:709-742.

Kordulakova J, Gilleron M, Mikusova K, Puzo G, Brennan PJ, Gicquel B, Jackson M (2002) Definition of the first mannosylation step in phosphatidylinositol mannoside synthesis: PimA is essential for growth of mycobacteria. J Biol Chem **277**:31335-31344.

Kowalak JA, Dalluge JJ, McCloskey JA, Stetter KO (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. Biochemistry **33**:7869-7876.

Labes A, Schonheit P (2001) Sugar utilization in the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324: starch degradation to acetate and CO2 via a modified Embden-Meyerhof pathway and acetyl-CoA synthetase (ADP-forming) Arch Microbiol **176**:329-338.

Labes A, Schonheit P (2003) ADP-dependent glucokinase from the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324. Arch Microbiol **180**:69-75.

Ladenstein R, Antranikian G (1998) Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. Adv Biochem Eng Biotechnol **61**:37-85.

Lai MC, Sowers KR, Robertson DE, Roberts MF, Gunsalus RP (1991) Distribution of compatible solutes in the halophilic methanogenic archaebacteria. J Bacteriol **173**:5352-5358.

Lai MC, Hong TY, Gunsalus RP (2000) Glycine betaine transport in the obligate halophilic archaeon *Methanohalophilus portucalensis*. J Bacteriol **182**:5020-5024.

Lamosa P, Martins LO, da Costa MS, Santos H (1998) Effects of temperature, salinity, and medium composition on compatible solute accumulation by *Thermococcus* spp. Appl Environ Microbiol **64**:3591-3598.

Lamosa P, Burke A, Peist R, Huber R, Liu MY, 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, Goncalves LG, Rodrigues MV, Martins LO, Raven ND, Santos H (2006) Occurrence of 1-glyceryl-1-*myo*-inosityl phosphate in hyperthermophiles. Appl Environ Microbiol **72**:6169-6173.

Larsen PI, Sydnes LK, Landfald B, Strom AR (1987) Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. Arch Microbiol **147**:1-7.

Lee AM, Sevinsky JR, Bundy JL, Grunden AM, Stephenson JL, Jr. (2009) Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods. J Proteome Res **8**:3844-3851.

Lehmacher A, Vogt AB, Hensel R (1990) Biosynthesis of cyclic 2,3-diphosphoglycerate: Isolation and characterization of 2-phosphoglycerate kinase and cyclic 2,3-diphosphoglycerate synthetase from *Methanothermus fervidus*. FEBS Lett **272**:94-98.

Lemaux PG, Herendeen SL, Bloch PL, Neidhardt FC (1978) Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts. Cell **13**:427-434.

Liu W, Vierke G, Wenke AK, Thomm M, Ladenstein R (2007) Crystal structure of the archaeal heat shock regulator from *Pyrococcus furiosus*: a molecular chimera representing eukaryal and bacterial features. J Mol Biol **369**:474-488.

Lopez-Garcia P, Forterre P (2000) Thermal stress in hyperthermophiles. p. 369-382. *In* G. Storz and R. Hengge-Aronis (ed.), Bacterial stress responses. ASM Press, Washington, DC.

Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet **9**:583-593.

Luke KA, Higgins CL, Wittung-Stafshede P (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. FEBS J **274**:4023-4033.

Luley-Goedl C, Nidetzky B (2011) Glycosides as compatible solutes: biosynthesis and applications. Nat Prod Rep **28**:875-896.

Lykidis A, Jackson PD, Rock CO, Jackowski S (1997) The role of CDP-diacylglycerol synthetase and phosphatidylinositol synthase activity levels in the regulation of cellular phosphatidylinositol content. J Biol Chem **272**:33402-33409.

Macario AJ, Lange M, Ahring BK, Conway de ME (1999) Stress genes and proteins in the archaea. Microbiol Mol Biol Rev **63**:923-67, table.

Manca MC, Nicolaus B, Lanzotti V, Trincone A, Gambacorta A, Peter-Katalinic J, Egge H, Huber R, Stetter KO (1992) Glycolipids from *Thermotoga maritima,* a hyperthermophilic microorganism belonging to *Bacteria* domain. Biochim Biophys Acta **1124**:249-252.

Marguet E, Forterre P (1994) DNA stability at temperatures typical for hyperthermophiles. Nucleic Acids Res **22**:1681-1686.

Marguet E, Forterre P (2001) Stability and manipulation of DNA at extreme temperatures. Methods Enzymol **334**:205-215.

Marin K, Huckauf J, Fulda S, Hagemann M (2002) Salt-dependent expression of glucosylglycerol-phosphate synthase, involved in osmolyte synthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803. J Bacteriol **184**:2870-2877.

Martins LO, Santos H (1995) Accumulation of mannosylglycerate and di-*myo*-inositol-phosphate by *Pyrococcus furiosus* in response to salinity and temperature. Appl Environ Microbiol **61**:3299-3303.

Martins LO, Carreto LS, da Costa MS, Santos H (1996) New compatible solutes related to di-*myo*-inositol-phosphate in members of the order *Thermotogales*. J Bacteriol **178**:5644-5651.

Martins LO, Huber R, Huber H, Stetter KO, da Costa MS, Santos H (1997) Organic solutes in hyperthermophilic archaea. Appl Environ Microbiol **63**:896-902.

Martins LO, Empadinhas N, Marugg JD, Miguel C, Ferreira C, da Costa MS, Santos H (1999) Biosynthesis of mannosylglycerate in the thermophilic bacterium *Rhodothermus marinus*: Biochemical and genetic characterization of a mannosylglycerate synthase. J Biol Chem **274**:35407-35414.

Massant J (2007) How thermophiles cope with thermolabile metabolites. p. 57-74. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Massy DJ, Wyss P (1990) Chemical synthesis of GPIs and GPI-anchored glycopeptides. Helv Chim Acta 73:1037-1057.

Michell RH (2008) Inositol derivatives: evolution and functions. Nat Rev Mol Cell Biol **9**:151-161.

Michell RH (2011) Inositol and its derivatives: Their evolution and functions. Adv Enzyme Regul **51**:84-90.

Mille C, Bobrowicz P, Trinel PA, Li H, Maes E, Guerardel Y, Fradin C, Martinez-Esparza M, Davidson RC, Janbon G, Poulain D, Wildt S (2008) Identification of a new family of genes involved in β-1,2-mannosylation of glycans in *Pichia pastoris* and *Candida albicans*. J Biol Chem **283**:9724-9736.

Moffatt JG, Khorana HG (1961) Nucleoside polyphosphates. XI.[1] The synthesis and some reactions of nucleoside-5′ phosphoromorpholidates and related compounds. Improved methods for the preparation of nucleoside-5′ polyphosphates. J Am Chem Soc. **83**:649-658.

Mongodin EF, Hance IR, DeBoy RT, Gill SR, Daugherty S, Huber R, Fraser CM, Stetter K, Nelson KE (2005) Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. J Bacteriol **187**:4935-4944.

Morii H, Nishihara M, Koga Y (2000) CTP:2,3-di-*O*-geranylgeranyl-*sn*-glycero-1-phosphate cytidyltransferase in the methanogenic archaeon *Methanothermobacter thermoautotrophicus*. J Biol Chem **275**:36568-36574.

Morii H, Kiyonari S, Ishino Y, Koga Y (2009) A novel biosynthetic pathway of archaetidyl-myo-inositol via archaetidyl-*myo*-inositol phosphate from CDP-archaeol and D-glucose 6-phosphate in methanoarchaeon *Methanothermobacter thermautotrophicus* cells. J Biol Chem **284**:30766-30774.

Morii H, Ogawa M, Fukuda K, Taniguchi H, Koga Y (2010) A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria. J Biochem **148**:593-602.

Morita YS, Patterson JH, Billman-Jacobe H, McConville MJ (2004) Biosynthesis of mycobacterial phosphatidylinositol mannosides. Biochem J **378**:589-597.

Morita YS, Yamaryo-Botte Y, Miyanagi K, Callaghan JM, Patterson JH, Crellin PK, Coppel RL, Billman-Jacobe H, Kinoshita T, McConville MJ (2010) Stress-induced synthesis of phosphatidylinositol 3-phosphate in mycobacteria. J Biol Chem **285**:16643-16650.

Morita YS, Fukuda T, Sena CB, Yamaryo-Botte Y, McConville MJ, Kinoshita T (2011) Inositol lipid metabolism in Mycobacteria: biosynthesis and regulatory mechanisms. Biochim Biophys Acta. **1810**:630-641.

Movahedzadeh F, Smith DA, Norman RA, Dinadayala P, Murray-Rust J, Russell DG, Kendall SL, Rison SC, McAlister MS, Bancroft GJ, McDonald NQ, Daffe M, Av-Gay Y, Stoker NG (2004) The *Mycobacterium tuberculosis ino1* gene is essential for growth and virulence. Mol Microbiol **51**:1003-1014.

Murthy PPN (2006) Structure and nomenclature of inositol phosphates, phosphoinositides, and glycosylphosphatidylinositols. p. 1-20. *In* A. L. Majumder and B. B. Biswas (ed.), Biology of inositols and phosphoinositides. Springer, Netherlands.

Müller V, Spanheimer R, Santos H (2005) Stress response by solute accumulation in archaea. Curr Opin Microbiol **8**:729-736.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between *Archaea* and *Bacteria* from genome sequence of *Thermotoga maritima*. Nature **399**:323-329.

Neves C, da Costa MS, Santos H (2005) Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: osmoadaptation and thermoadaptation in the order *Thermococcales*. Appl Environ Microbiol **71**:8091-8098.

Niegowski D, Hedren M, Nordlund P, Eshaghi S (2006) A simple strategy towards membrane protein purification and crystallization. Int J Biol Macromol **39**:83-87.

Nikawa J, Kodaki T, Yamashita S (1987) Primary structure and disruption of the phosphatidylinositol synthase gene of *Saccharomyces cerevisiae*. J Biol Chem **262**:4876-4881.

Numbering of atoms in *myo*-inositol. Recommendations 1988. Nomenclature Committee of the International Union of Biochemistry (1989) Biochem J **258**:1-2.

Nunes OC, Manaia CM, da Costa MS, Santos H (1995) Compatible solutes in the thermophilic bacteria *Rhodothermus marinus* and "*Thermus thermophilus*". Appl Environ Microbiol **61**:2351-2357.

Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee GJ, Hattori M, Kanai A, Atomi H, Takai K, Takami H (2011) Insights into the evolution of *Archaea* and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res **39**:3204-3223.

Oren A (1999) Bioenergetic aspects of halophilism. Microbiol Mol Biol Rev 63:334-348.

Paper W, Jahn U, Hohn MJ, Kronner M, Nather DJ, Burghardt T, Rachel R, Stetter KO, Huber H (2007) *Ignicoccus hospitalis* sp. nov., the host of '*Nanoarchaeum equitans*'. Int J Syst Evol Microbiol **57**:803-808.

Park YS, Sweitzer TD, Dixon JE, Kent C (1993) Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from *Bacillus subtilis*. J Biol Chem **268**:16648-16654.

Parthasarathy R, Eisenberg F, Jr. (1986) The inositol phospholipids: a stereochemical view of biological activity. Biochem J **235**:313-322.

Pattridge KA, Weber CH, Friesen JA, Sanker S, Kent C, Ludwig ML (2003) Glycerol 3-phosphate cytidylyltransferase: Structural changes induced by binding of CDP-glycerol and the role of lysine residues in catalysis. J Biol Chem **278**:51863-51871.

Peretó J, Lopez-Garcia P, Moreira D (2004) Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci **29**:469-477.

Peter H, Weil B, Burkovski A, Kramer R, Morbach S (1998) *Corynebacterium glutamicum* is equipped with four secondary carriers for compatible solutes: identification, sequencing, and characterization of the proline/ectoine uptake system, ProP, and the ectoine/proline/glycine betaine carrier, EctP. J Bacteriol **180**:6005-6012.

Pflüger K, Baumann S, Gottschalk G, Lin W, Santos H, Müller V (2003) Lysine-2,3-aminomutase and β-lysine acetyltransferase genes of methanogenic archaea are salt induced and are essential for the biosynthesis of *N*-ε-acetyl-β-lysine and growth at high salinity. Appl Environ Microbiol **69**:6047-6055.

Pflüger K, Müller V (2004) Transport of compatible solutes in extremophiles. J Bioenerg Biomembr **36**:17-24.

Pikuta EV, Hoover RB, Tang J (2007) Microbial extremophiles at the limits of life. Crit Rev Microbiol **33**:183-209.

Porchia AC, Salerno GL (1996) Sucrose biosynthesis in a prokaryotic organism: Presence of two sucrose-phosphate synthases in *Anabaen*a with remarkable differences compared with the plant enzymes. Proc Natl Acad Sci U S A **93**:13600-13604.

Proctor LM, Lai R, Gunsalus RP (1997) The methanogenic archaeon *Methanosarcina thermophila* TM-1 possesses a high-affinity glycine betaine transporter involved in osmotic adaptation. Appl Environ Microbiol **63**:2252-2257.

Purves, WK, Orians GH, Heller HC, Sadava D (1998) LIFE: The Science of Biology. 5th Ed. Sinauer Associates, Inc., W. H. Freeman and Company. U.S.A.

Pysz MA, Ward DE, Shockley KR, Montero CI, Conners SB, Johnson MR, Kelly RM (2004) Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8**:209-217.

Ralton JE, Naderer T, Piraino HL, Bashtannyk TA, Callaghan JM, McConville MJ (2003) Evidence that intracellular β-1-2 mannan is a virulence factor in *Leishmania* parasites. J Biol Chem **278**:40757-40763.

Ramakrishnan V, Adams MWW (1995) Preparation of genomic DNA from sulfur-dependent hyperthermophilic *Archaea*. p. 95-96. *In* F.T. Robb, and A.R. Place (ed.), *Archaea*: a Laboratory Manual - Thermophiles. Cold Spring Harbor, NY:Cold Spring Harbor Laboratory Press.

Ramakrishnan V, Verhagen M, Adams M (1997) Characterization of di-*myo*-inositol-1,1-phosphate in the hyperthermophilic bacterium *Thermotoga maritima*. Appl Environ Microbiol **63**:347-350.

Reed RH, Richardson DL, Warr SRC, Stewart WDP (1984) Carbohydrate accumulation and osmotic stress in cyanobacteria. J Gen Microbiol **130**:1–4.

Ritossa F (1962) A new puffing pattern induced by tempearture shock and DNP in *Drosophila*. Experientia **18**:571-573.

Robb FT, Newby DT (2007) Functional genomics in thermophilic microorganisms. p. 30-38. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Roberts MF, Lai MC, Gunsalus RP (1992) Biosynthetic pathways of the osmolytes N-ε-acetyl-β-lysine, β-glutamine, and betaine in *Methanohalophilus* strain FDF1 suggested by nuclear magnetic resonance analyses. J Bacteriol **174**:6688-6693.

Roberts MF (2004) Osmoadaptation and osmoregulation in archaea: update 2004. Front Biosci **9**:1999-2019.

Roberts MF (2005) Organic compatible solutes of halotolerant and halophilic microorganisms. Saline Systems **1**:5.

Roberts MF (2006) Inositol in *Bacteria* and *Archaea*. Subcell Biochem **39**:103-133.

Robertson DE, Roberts MF, Belay N, Stetter KO, Boone DR (1990) Occurrence of β-glutamate, a novel osmolyte, in marine methanogenic bacteria. Appl Environ Microbiol **56**:1504-1508.

Robertson DE, Noll D, Roberts MF (1992) Free amino acid dynamics in marine methanogens: β-amino acids as compatible solutes. J Biol Chem **267**:14893-14901.

Rodionov DA, Kurnasov OV, Stec B, Wang Y, Roberts MF, Osterman AL (2007) Genomic identification and in vitro reconstitution of a complete biosynthetic pathway for the osmolyte di-*myo*-inositol-phosphate. Proc Natl Acad Sci U S A **104**:4279-4284.

Roeβler M, Müller V (2001) Osmoadaptation in bacteria and archaea: common principles and differences. Environ Microbiol **3**:743-754.

Roeβler M, Pfluger K, Flach H, Lienard T, Gottschalk G, Müller V (2002) Identification of a salt-induced primary transporter for glycine betaine in the methanogen *Methanosarcina mazei* Go1. Appl Environ Microbiol **68**:2133-2139.

Rohlin L, Trent JD, Salmon K, Kim U, Gunsalus RP, Liao JC (2005) Heat shock response of *Archaeoglobus fulgidus*. J Bacteriol **187**:6046-6057.

Roseman S, Distler JJ, Moffatt JG, Khorana HG (1961) Nucleoside polyphosphates. XI.1 An improved general method for the synthesis of nucleotide coenzymes. syntheses of uridine-5′, cytidine-5′ and guanosine-5′ diphosphate derivatives. J Am Chem Soc **83**:659-663.

Rothschild LJ, Mancinelli RL (2001) Life in extreme environments. Nature **409**:1092-1101.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4**:406-425.

Sakasegawa S, Hagemeier CH, Thauer RK, Essen LO, Shima S (2004) Structural and functional analysis of the gpsA gene product of *Archaeoglobus fulgidus*: a glycerol-3-phosphate dehydrogenase with an unusual $NADP^+$ preference. Protein Sci **13**:3161-3171.

Salman M, Lonsdale JT, Besra GS, Brennan PJ (1999) Phosphatidylinositol synthesis in mycobacteria. Biochim Biophys Acta **1436**:437-450.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY:Cold Spring Harbor Laboratory Press.

Santos H, da Costa MS (2001) Organic solutes from thermophiles and hyperthermophiles. Methods Enzymol **334**:302-315.

Santos H, da Costa MS (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. Meth Microbiol **35**:173-199.

Santos H, Lamosa P, Faria TQ, Borges N, Neves C (2007) The physiological role, biosynthesis, and mode of action of compatible solutes from (hyper)thermophiles. p. 86-104. *In* C. Gerday and N. Glandorff (ed.), Physiology and biochemistry of extremophiles. ASM Press, Washington, DC.

Santos H, Lamosa P, Borges N, Gonçalves LG, Pais T, Rodrigues MV (2010) Organic compatible solutes of prokaryotes that thrive in hot environments: the importance of

ionic compounds for thermostablization. p.497-520. *In* K. Horikoshi (ed.) Extremophiles handbook. Springer Japan.

Sato T, Fukui T, Atomi H, Imanaka T (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Bacteriol **185**:210-220.

Scholz S, Sonnenbichler J, Schafer W, Hensel R (1992) Di- *myo*-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei*. FEBS Lett **306**:239-242.

Scholz S, Wolff S, Hensel R (1998) The biosynthesis pathway of di-*myo*-inositol-1,1'-phosphate in *Pyrococcus woesei*. FEMS Microbiology Letters **168**:37-42.

Schumann W (2007) Thermosensors in eubacteria: role and evolution. J Biosci **32**:549-557.

Serebrov V, Clarke RJ, Gross HJ, Kisselev L (2001) Mg^{2+}-induced tRNA folding. Biochemistry **40**:6688-6698.

Servant P, Grandvalet C, Mazodier P (2000) The RheA repressor is the thermosensor of the HSP18 heat shock response in *Streptomyces albus*. Proc Natl Acad Sci U S A **97**:3538-3543.

Shibata N, Ichikawa T, Tojo M, Takahashi M, Ito N, Okubo Y, Suzuki S (1985) Immunochemical study on the mannans of *Candida albicans* NIH A-207, NIH B-792, and J-1012 strains prepared by fractional precipitation with cetyltrimethylammonium bromide. Arch Biochem Biophys **243**:338-348.

Shibata N, Arai M, Haga E, Kikuchi T, Najima M, Satoh T, Kobayashi H, Suzuki S (1992) Structural identification of an epitope of antigenic factor 5 in mannans of *Candida albicans* NIH B-792 (serotype B) and J-1012 (serotype A) as β-1,2-linked oligomannosyl residues. Infect Immun **60**:4100-4110.

Shockley KR, Ward DE, Chhabra SR, Conners SB, Montero CI, Kelly RM (2003) Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl Environ Microbiol **69**:2365-2371.

Schumann W (2007) Thermosensors in eubacteria: role and evolution. J Biosci **32**:549-557.

Silva Z, Borges N, Martins LO, Wait R, da Costa MS, 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**:163-172.

Smith LT, Smith GM, Madkour MA (1990) Osmoregulation in *Agrobacterium tumefaciens*: accumulation of a novel disaccharide is controlled by osmotic strength and glycine betaine. J Bacteriol **172**:6849-6855.

Sowers KR, Robertson DE, Noll D, Gunsalus RP, Roberts MF (1990) N-ε-acetyl-β-lysine: an osmolyte synthesized by methanogenic archaebacteria. Proc Natl Acad Sci U S A **87**:9083-9087.

Stetter KO, Thomm M, Winter J, Wildgruber G, Huber H, Zillig W, Janecovic D, Konig H, Palm P, Wundel S (1981) *Methanothermus fervidus*, sp. nov., a novel extremely thermophilic methanogen isolated from an Icelandic hot spring. Zbl Bakt Hyg I Abt Orig C2:166-178.

Stetter KO (1988) *Archaeoglobus fulgidus* gen. nov., sp. nov.: new taxon of extremely thermophilic *Archaebacteria*. Syst Appl Microbiol **10**: 172-173.

Stetter KO (1999) Extremophiles and their adaptation to hot environments. FEBS Lett **452**:22-25.

Stetter KO (2006) History of discovery of the first hyperthermophiles. Extremophiles **10**:357-362.

Stieglitz KA, Yang H, Roberts MF, Stec B (2005) Reaching for mechanistic consensus across life kingdoms: structure and insights into catalysis of the *myo*-inositol-1-phosphate synthase (mIPS) from *Archaeoglobus fulgidus*. Biochemistry **44**: 213-224.

Streeter JG (1985) Accumulation of α,α-trehalose by *Rhizobium* bacteria and bacteroids. J Bacteriol **164**:78-84.

Suzuki A, Shibata N, Suzuki M, Saitoh F, Oyamada H, Kobayashi H, Suzuki S, Okawa Y (1997) Characterization of β-1,2-mannosyltransferase in *Candida guilliermondii* and its utilization in the synthesis of novel oligosaccharides. J Biol Chem **272**:16822-16828.

Tachdjian S, Kelly RM (2006) Dynamic metabolic adjustments and genome plasticity are implicated in the heat shock response of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus*. J Bacteriol **188**:4553-4559.

Tachdjian S, Shockley KR, Conners SB, Kelly RM (2008) Functional genomics of stress response in extremophilic *Archaea*. p. 219-232. *In* P. Blum (ed.) *Archaea*: new models for prokaryotic biology. Caister Academic Press. Norfolk, UK.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24**:1596-1599.

Terui Y, Ohnuma M, Hiraga K, Kawashima E, Oshima T (2005) Stabilization of nucleic acids by unusual polyamines produced by an extreme thermophile, *Thermus thermophilus*. Biochem J **388**:427-433.

The nomenclature of lipids (1967) J Lipid Res **8**:523-528.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25**:4876-4882.

Trauger SA, Kalisak E, Kalisiak J, Morita H, Weinberg MV, Menon AL, Poole FL, Adams MW, Siuzdak G (2008) Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. J Proteome Res **7**:1027-1035.

Ulrih NP, Gmajner D, Raspor P (2009) Structural and physicochemical properties of polar lipids from thermophilic archaea. Appl Microbiol Biotechnol **84**:249-260.

Unsworth LD, van der OJ, Koutsopoulos S (2007) Hyperthermophilic enzymes: stability, activity and implementation strategies for high temperature applications. FEBS J **274**:4044-4056.

van den Burg B (2003) Extremophiles as a source for novel enzymes. Curr Opin Microbiol **6**:213-218.

VanFossen AL, Lewis DL, Nichols JD, Kelly RM (2008) Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. Ann N Y Acad Sci **1125**:322-337.

van Leeuwen SH, van der Marel GA, Hensel R, van Boom JH (1994) Synthesis of L,L-di-*myo*-inositol-1,1'-phosphate: a novel inositol phosphate from *Pyrococcus woesei*. Recl Trav Chim Pays-Bas **113**:335-336.

van de Vossenberg JL, Driessen AJ, Konings WN (1998) The essence of being extremophilic: the role of the unique archaeal membrane lipids. Extremophiles **2**:163-170.

Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev **65**:1-43.

Vierke G, Engelmann A, Hebbeln C, Thomm M (2003) A novel archaeal transcriptional regulator of heat shock response. J Biol Chem **278**:18-26.

Vinogradov E, Petersen BO, Duus JO (2000) Isolation and characterization of non-labeled and ^{13}C-labeled mannans from *Pichia pastoris* yeast. Carbohydr Res **325**:216-221.

von Blohn C, Kempf B, Kappes RM, Bremer E (1997) Osmostress response in *Bacillus subtilis*: characterization of a proline uptake system (OpuE) regulated by high

osmolarity and the alternative transcription factor sigma B. Mol Microbiol **25**:175-187.

Watanabe Y, Yamamoto T, Okazaki T (1997) Synthesis of 2,6-di-*O*-α-D-mannopyranosylphosphatidyl-D-*myo*-inositol: utilization of glycosylation and phosphorylation based on phosphite chemistry. Tetrahedron **53**:903-918.

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A **87**:4576-4579.

Woese CR (1993) The *Archaea*: their history and significance. *In* M. Kates, D.J. Kushner, A. T. Matheson (ed) The Biochemistry of the *Archaea* (*Archaebacteria*) Elsevier Science Publishers B. V. Netherlands.

Wohlfarth A, Severin J, Galinski E A (1993) Identification of Nδ-acetylornithine as a novel osmolyte in some Gram-positive halophilic eubacteria. Appl Microbiol Biotechno **39**:568-573.

Worthington P, Hoang V, Perez-Pomares F, Blum P (2003) Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. J Bacteriol **185**:482-488.

Xavier KB, Martins LO, Peist R, Kossmann M, Boos W, Santos H (1996) High-affinity maltose/trehalose transport system in the hyperthermophilic archaeon *Thermococcus litoralis*. J Bacteriol **178**:4773-4777.

Xu Y, Zhou P, Tian X (1999) Characterization of two novel haloalkaliphilic archaea *Natronorubrum bangense* gen. nov., sp. nov. and *Natronorubrum tibetense* gen. nov., sp. nov. Int J Syst Bacteriol 49 Pt **1**:261-266.

Yamamori T, Yura T (1980) Temperature-induced synthesis of specific proteins in *Escherichia coli*: evidence for transcriptional control. J Bacteriol **142**:843-851.

Yim VC, Zolli M, Badurina DS, Rossi L, Brown ED, Berghuis AM (2001) Crystallization and preliminary X-ray diffraction studies of glycerol 3-phosphate cytidylyltransferase from *Staphylococcus aureus*. Acta Crystallogr D Biol Crystallogr **57**:918-920.

Yu K-L, Fraser-Reid B (1988) A novel reagent for the synthesis of *myo*-inositol phosphates: *N,N*-diisopropyl dibenzyl phosphoramidite. Tetrahedron Lett **29**:979-982.

Yura T, Kanemori M, Morita m T (2000) The heat shock response: regulation and function. p. 3-18. *In* G. Storz and R. Hengge-Aronis (ed.), Bacterial stress responses. ASMPress, Washington, DC.

Zaparty M, Esser D, Gertig S, Haferkamp P, Kouril T, Manica A, Pham TK, Reimann J, Schreiber K, Sierocinski P, Teichmann D, van WM, von JM, Wieloch P, Albers SV, Driessen AJ, Klenk HP, Schleper C, Schomburg D, van der OJ, Wright PC, Siebers B (2010) "Hot standards" for the thermoacidophilic archaeon *Sulfolobus solfataricus*. Extremophiles **14**:119-142.

Zellner G, Stackebrandt E, Kneifel H, Messner P, Sleytr UB, de Macario EC, Zabel H-P, Stetter KO, Winter J (1989) Isolation and characterization of a thermophilic, sulfate reducing archaebacterium, *Archaeoglobus fulgidus* strain Z. System Appl Microbiol **11**:151-160.

APPENDIX 1

Transcriptional regulation of the genes involved in the synthesis of di-*myo*-inositol phosphate and mannosyl-di-*myo*-inositol phosphate in *Thermotoga maritima*: preliminary results

Appendix 1 – Contents

Summary

The heat-shock response of *Thermotoga maritima* was investigated with respect to the accumulation of compatible solutes and transcription profiles of genes encoding enzymes involved in the synthesis of di-*myo*-inositol 1,3'-phosphate (DIP), mannosyl-di-*myo*-inositol 1,3'-phosphate (MDIP), and di-mannosyl-di-*myo*-inositol 1,3'-phosphate (MMDIP). To this end, *Thermotoga maritima* was allowed to grow at optimal temperature (80ºC) until the mid-exponential phase. Then, the temperature was increased to 88ºC and maintained at this value for *ca*. 3 hours. The pool of compatible solutes under optimal growth conditions was composed of proline, aspartate, α- and β-glutamate, DIP and MDIP. At time 90 min of exposure to heat-shock the total amount of solutes increased 1.6-fold with respect to optimal growth conditions, especially due to the accumulation of α-glutamate (1.9-fold increase), and MDIP (2.7-fold increase). Additionally, low levels of a second mannosylated DIP-derivative, MMDIP, were detected. The transcriptional levels of the genes *mpg1*, *mds*, *suhB*, *mips*, *ipct*, and *dipps* encoding, respectively, mannose-1-phosphate guanylyltransferase, MDIP synthase, *myo*-inositol-1-monophosphatase, L-*myo*-inositol 1-phosphate synthase, CTP:*myo*-inositol 1-phosphate cytidylyltransferase, and CDP-inositol:*myo*-inositol 1-phosphate transferase were determined by RT-PCR. Surprisingly, the expression of *ipct* and *dipps* was down-regulated in response to heat-shock. The transcript levels of *mpg1*, *mds*, *suhB*, and *mips* remained approximately constant. It is intriguing that the expression of *mds* (implicated in the synthesis of MDIP and MMDIP) was not affected despite the clear increase in the pools of these solutes.

Introduction

The stress response refers to the group of events occurring in the cell upon exposure to harmful agents with the purpose to prevent cell death by counteracting the damage provoked by stressful conditions. (de Macario and Macario 1994; Holden et al. 1999). In particular, when a culture is exposed to supra-optimal growth temperatures, the global heat-shock response involves not only the decrease in the number of viable cells but also the induction of the synthesis of proteins that prevent protein aggregation, reassemble misfolded proteins or degrade permanently damaged proteins. In addition to the collective response of molecular chaperones, chaperonins and proteases, other events like the accumulation of compatible solutes or biofilm formation are regarded as mechanisms that lead to thermo-tolerance (Holden et al. 1999).

Some compatible solutes are highly restricted to (hyper)thermophilic organisms leading to the view that they play a role in thermoadaptation (Santos et al. 2007). A classical example is di-*myo*-inositol 1,3'-phosphate (DIP), which is the most widespread compatible solute among hyperthermophilic bacteria and archaea and has never been found in organisms with optimal growth temperature below 60ºC (Santos et al. 2010). Moreover, glycero-phospho-*myo*-inositol (GPI), mannosyl-di-*myo*-inositol 1,3'-phosphate (MDIP) and di-mannosyl-di-*myo*-inositol 1,3'-phosphate (MMDIP) are exclusive to hyperthermophilic organisms. GPI has been encountered in the archaeon *Archaeoglobus fulgidus* and species of the genus *Aquifex*, while MDIP and MMDIP are found in species of the genera *Aquifex* and *Thermotoga* (Lamosa et al. 2006, Chapter 5).

Thermotoga maritima is a model organism to study hyperthermophilic bacteria since it is easy to cultivate, can use many substrates for growth and

has great biotechnological potential (Conners et al. 2006). The pool of compatible solutes in *Thermotoga maritima* comprises α- and β-glutamate, DIP, MDIP and MMDIP; the level of the latter two solutes increase notably in cells grown at supra-optimal temperature, suggesting a role in thermoprotection. However, virtually nothing is known about the regulation of the biosynthesis of these solutes. An effective approach to fill this gap in knowledge must involve a systematic study of the heat shock response, comprising transcriptional, proteomic and metabolic (solute pools) analyses.

The transcriptional analysis of the heat-shock response of hyperthermophilic organisms has been studied in *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Methanocaldococcus jannaschii* and *Thermotoga maritima* (Boonyaratanakornkit et al. 2007, Shockley et al. 2003, Rohlin et al. 2005, Pysz et al. 2004). In particular, the transcriptional analysis of the heat-shock response of *Thermotoga maritima* revealed that, although it shares many elements with the heat-shock response of the mesophilic bacterium *Escherichia coli*, the regulatory strategies are significantly different. The common core elements are the up-regulation of heat-shock operons such as *hrcA-grpE-dnaJ*, *groES-groEL*, and *dnaK-sHSP* and the induction of the *rpoE*|*sigW* and *rpoD*|*sigA* homologs. In contrast to *Escherichia coli*, ATP-dependent proteases like *clpP*, *clpQ*, *clpY*, *lonA* and, *lonB* are down-regulated (Pysz et al. 2004).

The present study was motivated by our interest in understanding the chain of events leading to the final accumulation of solutes in response to heat shock. As a first step to this goal, we set out to obtain information on the heat shock response of *Thermotoga maritima* with respect to the solute pools and the transcript levels of the biosynthetic genes implicated in the synthesis of DIP and derivatives.

Materials and methods

Strains and culture conditions

The type strain of *Thermotoga maritima* (DSM 3109^T) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany). *Thermotoga maritima* was cultivated anaerobicaly at 80°C in a medium containing (per liter) 22 g of NaCl, 3 g of PIPES, 1.75 g of $MgSO_4 \bullet 7H_2O$, 0.5 g of $MgCl_2 \bullet 6H_2O$, 0.16 g of KCl, 25 mg of NaBr, 7.5 mg of H_3BO_3, 3.75 mg of $SrCl_2 \bullet 6H_2O$, 0.01 mg of KI, 0.25 g of $CaCl_2 \bullet 2H_2O$, 0.4 g of KH_2PO_4, 1 g of NH_4Cl, 0.4 g of cysteine-HCl, 5 g of tryptone, 3.16 g of thiosulfate, 4 mg of $NiCl_2 \bullet 6H_2O$, 2.5 mg of citric acid and 10 ml of modified trace minerals solution. Trace mineral solution was prepared according to Balch et al. (Balch et al. 1979) excluding $MgSO_4 \bullet 7H_2O$, NaCl and $CaCl_2 \bullet 2H_2O$. The final pH was adjusted to 6.5. For cultivation in bottles the medium was degassed with N_2 and then sterilized by autoclaving; for growth in fermentor, the medium was sterilized by autoclaving and degassed with N_2 prior to inoculation. In both cases, prior to inoculation the culture medium was supplemented with 0.05% (wt/vol) of Na_2S and 0.5% (wt/vol) of glucose. A 150 ml batch culture in mid exponential phase of growth was used to inoculate a 5 l fermentor containing 4 l of medium prepared as described above; the fermentor was equipped with an internal temperature controller and the growth proceeded with continuous gassing with pure N_2 and mild stirring at 40 rpm. Cell growth was monitored by measurement of optical density at 600 nm. For the heat-shock experiment, after 7 h of growth at 80°C (mid-exponential phase of growth), a 250 ml-sample was collected; then the temperature set point was raised, taking approximately 20 min to reach 88°C. Samples were taken at 15, 40, 60, 90, 150, and 200 min after reaching 88°C.

Then the temperature was shifted back to 80ºC, taking approximately 20 min to reach this temperature, and four more samples were collected (at 15, 40, 90 and 180 min). At each time point approximately 250 ml of sample was withdrawn for intracellular solute analysis and RNA extraction: approximately 30 ml were immediately centrifuged (1,600 × *g*, 10 min at 4°C) and stored at –80°C until processed for RNA extraction; the remaining sample was centrifuged (7,000 × *g*, 10 min at 4°C) and washed twice with an NaCl solution (2.2% (wt/vol)), and stored at -20°C until processed for solute extraction. *Thermotoga maritima* was also cultivated at the optimal growth temperature (80°C) and at 88°C. Samples were collected during mid exponential, late exponential and stationary phases of growth and treated as described above. Doubling time was calculated from the slope of semilog plots of optical density (exponential cell growth) over time.

Extraction, identification, and quantification of intracellular solutes

Cell pellets were suspended in water and disrupted by sonication. Aliquots were removed for determination of total protein content by the Bradford method (Bradford 1976). The remaining cell extract was treated twice with boiling 80% ethanol as described previously (Santos et al. 2006). Freeze-dried extracts were dissolved in $^{2}H_2O$ and analyzed by NMR (Santos et al. 2006).

Reverse transcription-PCR (RT-PCR) experiments

Total RNA was isolated from *Thermotoga maritima* cells using the RNeasy Mini Kit from Qiagen (QIAGEN). To avoid contamination with chromosomal DNA an additional incubation step with DNase I from the kit SV total RNA isolation

system (Promega) was performed. The samples were incubated with DNaseI for 30 min at 24ºC, according to the instructions on the kit (SV total RNA isolation system, Promega), and the subsequent steps of the protocol were performed as recommended. Total RNA (1.6 μg), deoxynucleoside triphosphates (final concentration of 0.48 mM) and random oligonucleotides (11.5 ng·μl^{-1}) (Invitrogen) were heated at 65°C for 5 min and chilled on ice. Dithiothreitol (final concentration of 4.8 mM), first-strand RT buffer, SuperScript™ III reverse transcriptase (0.96/20, vol/vol) (Invitrogen) and SUPERase-In (40 U) (Ambion) were added, and samples were incubated for 5 min at 25°C, 60 min at 50°C, and 15 min at 70°C for enzyme inactivation. An identical sample was treated in parallel, except for the addition of the reverse transcriptase. cDNA was subsequently used as a template for standard PCRs. To test for DNA contamination in the RNA preparations, the RNA samples without reverse transcriptase were used as negative controls for all conditions examined. Chromosomal DNA of *Thermotoga maritima* was used as a positive control for the PCR. The primer pairs (Table A1) were designed to amplify internal fragments of *myo*-inositol 1-phosphate synthase (*mips;* TM_1419), CTP:*myo*-inositol 1-phosphate cytidylyltransferase (*ipct;* TM_1418_a), CDP-inositol: *myo*-inositol 1-phosphate transferase (*dipps;* TM_1418_b), *myo*-inositol-1-monophosphatase (*suhB*; TM_1415), mannosyl-di-*myo*-inositol 1-phosphate synthase (*mds*; TM_0359), mannose 1-phosphate guanylyltransferase (*mpg1*; TM_1033), and DNA-binding protein HU (*hup*; TM_0266). The DNA-binding protein HU (*hup*) was used as a control. The cDNA samples were diluted 1:1 (single experiments), 1:3 and 1:5 (at least duplicate experiments) and used as template in the PCR analysis.

Table A1. Primers used in the RT-PCR experiments.

Primers	Sequence (5′ to 3′)	Chromosomal location
mips_F	CACTGCAACTCAGGTGTACGCTTACGC	447-474 bp downstream of *mips* SC
mips_R	TATGCGTACTCGATGTGAATGGC	853-872 bp downstream of *mips* SC
ipct_F	GCTCGTAAAATATCCGATGATCTCCC	98-124 bp downstream of *ipct* SC
ipct_R	CCTCAGTCCACGAAAAACTCTCCTTG	536-562 bp downstream of *ipct* SC
dipps_F	GATGGGTGGATTTCTTCTCTGATAAACAG	18-48 bp downstream of *dipps* SC
dipps_R	GGCAAAACCGGATAGCTTTCCGACG	455-480 bp downstream of *dipps* SC
suhB_F	CCCTGACGAAAACATCATGGCCG	98-121 bp downstream of *suhB* SC
suhB_R	CCTTCACGATTATCAGACCGGCTGC	534-559 bp downstream of *suhB* SC
mds_F	CCTGGACCTGCTCATAATAGAAACC	260-285 bp downstream of *mds* SC
mds_R	GCCACTGGGTGATCCACTCAGCCTCG	680-706 bp downstream of *mds* SC
hup_F	GACCAAGAAGGAACTCATCGAC	2-24 bp downstream of *hup* SC
hup_R	CTTTGAGGGCTTTTCCGGGTTTG	236-259 bp downstream of *hup* SC
mpg1_F	GAGCTTCCAGAGCTTCCAGACG	201-223 bp downstream of *mpg1* SC
mpg1_R	GGATCGACATCCTTCAGATTTTCG	635-659 bp downstream of *mpg1* SC

The genes *mips, ipct, dipps, suhB, mds mpg1*, and *hup* encode the enzymes *myo*-inositol 1-phosphate synthase, CTP:*myo*-inositol 1-phosphate cytidylyltransferase, CDP-inositol: *myo*-inositol 1-phosphate transferase, *myo*-inositol-1-monophosphatase, mannosyl-di-*myo*-inositol 1-phosphate synthase, mannose 1-phosphate guanylyltransferase, and DNA-binding protein HU, respectively. SC: Start codon.

NMR spectroscopy

For the quantification of organic solutes accumulating in *Thermotoga maritima*, ^{1}H-NMR spectra were acquired on a Bruker *AVANCE*^{II}500 spectrometer (Bruker, Rheinstetten, Germany) equipped with a 5-mm broadband inverse probe head. Spectra were acquired with pre-saturation of

the water signal using a repetition delay of 60 s. Formate was used as an internal concentration standard.

Results

Effect of supra-optimal growth temperature and growth phase on the accumulation of organic solutes in *Thermotoga maritima*

Thermotoga maritima grows optimally at 80ºC in a medium containing 2.7% (wt/vol) of NaCl (Huber et al. 1986). In this study *Thermotoga maritima* was cultivated in a medium containing 2.2% (wt/vol) of NaCl, as higher cell density was obtained in these conditions. At 80ºC the doubling time of *Thermotoga maritima* was 2.1 h and the maximal optical density was around 1.8. For cells grown at 88ºC (the upper limit for *Thermotoga maritima* growth is 90ºC), the doubling time was 4.7 h and the maximal optical density did not exceed 0.2.

^{1}H-NMR analysis revealed that the pool of organic compounds accumulating in *Thermotoga maritima* was composed of the amino acids proline, aspartate, α- and β-glutamate, and by the polyol phosphodiesters DIP, MDIP and MMDIP. Low amounts of MMMDIP, the tri-mannosylated derivative of DIP, were detected at mid exponential phase, when cells were grown at 88ºC. The pool of solutes reached a maximum during late exponential phase of growth, decreasing during the stationary phase.

The total amount of organic solutes accumulating at optimal temperature was around 0.3 µmol·mg protein^{-1} (Table A2 and Fig. A1), and at supra-optimal temperature (88ºC) the solute pool increased to 0.5 µmol·mg protein^{-1}. Amino acids dominated the solute pool, accounting to 64% of the

total organic solutes, when *Thermotoga maritima* was grown at 88ºC. The remaining 36% were composed of DIP (13%) and derivatives, MDIP (21%), and MMDIP (2%). On the other hand, when cells were cultivated at optimal temperature (80ºC), MMDIP was absent, DIP and MDIP represented 16% and 12% of the total pool, respectively, and amino acids (72%) were the major components.

Table A2. Intracellular content of organic solutes in *Thermotoga maritima* during growth at the optimal temperature (80ºC), and supra-optimal temperature (88ºC). Quantification of solutes was performed by ^{1}H-NMR.

Temp. (ºC)	Time of growth (h)	Intracellular concentration ($\times 10^2$ µmol·mg protein^{-1})[a] ± stdv ($\times 10^2$)							
		MMMDIP	**MMDIP**	**MDIP**	**DIP**	**β-glu**	**α-glu**	**Asp**	**Pro**
80	7	nd	nd	1.7 ± 0.2 (3.8)	2.9 ± 0.4 (6.4)	2.6 ± 1.2 (5.8)	7.2 ± 3.2 (16)	1.8 ± 0.4 (4.0)	7.0 ± 0.8 (15.6)
	12	nd	nd	3.4 ± 0.2 (7.6)	4.5 ± 0.2 (10.0)	5.1 ± 0.8 (11.3)	4.0 ± 3.8 (8.9)	1.3 ± 0.1 (2.9)	9.7 ± 1.0 (21.6)
	21	nd	nd	2.3* (5.1)	5.6* (12.4)	4.1* (9.1)	7.0* (15.6)	0.8* (1.8)	2.3* (5.1)
88	8	0.7 ± 0.6 (1.6)	1.6 ± 0.9 (3.6)	8.2 ± 0.5 (18.2)	4.8 ± 0.2 (10.7)	1.9 ± 0.2 (4.2)	5.1 ± 0.3 (11.3)	2.4 ± 0.6 (5.3)	11.6 ± 3.8 (25.8)
	12	nd	0.7 ± 0.4 (1.6)	10.2 ± 1.2 (22.7)	6.3 ± 0.1 (14.0)	2.6* (5.8)	13.5 ± 4.1 (30.0)	3.4 ± 1.5 (7.6)	11.2 ± 3.0 (24.9)
	24	nd	<0.5	4.6 ± 2.5 (10.2)	6.3 ± 1.0 (14.0)	<0.5	5.6 ± 1.7 (12.4)	1.4 ± 0.4 (3.1)	4.3 ± 1.9 (9.6)

[a] The intracellular concentration (mM) based on a cell volume of 4.5 µl·mg protein^{-1} as determined for *Pyrococcus furiosus* (Martins et al. 1995) is given in parenthesis. * Results obtained from a single experiment; all other results are averages of at least two independent determinations. DIP, di-*myo*-inositol 1,3'-phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate; MMMDIP, tri-mannosyl-di-*myo*-inositol 1,3'-phosphate; nd, not detected.

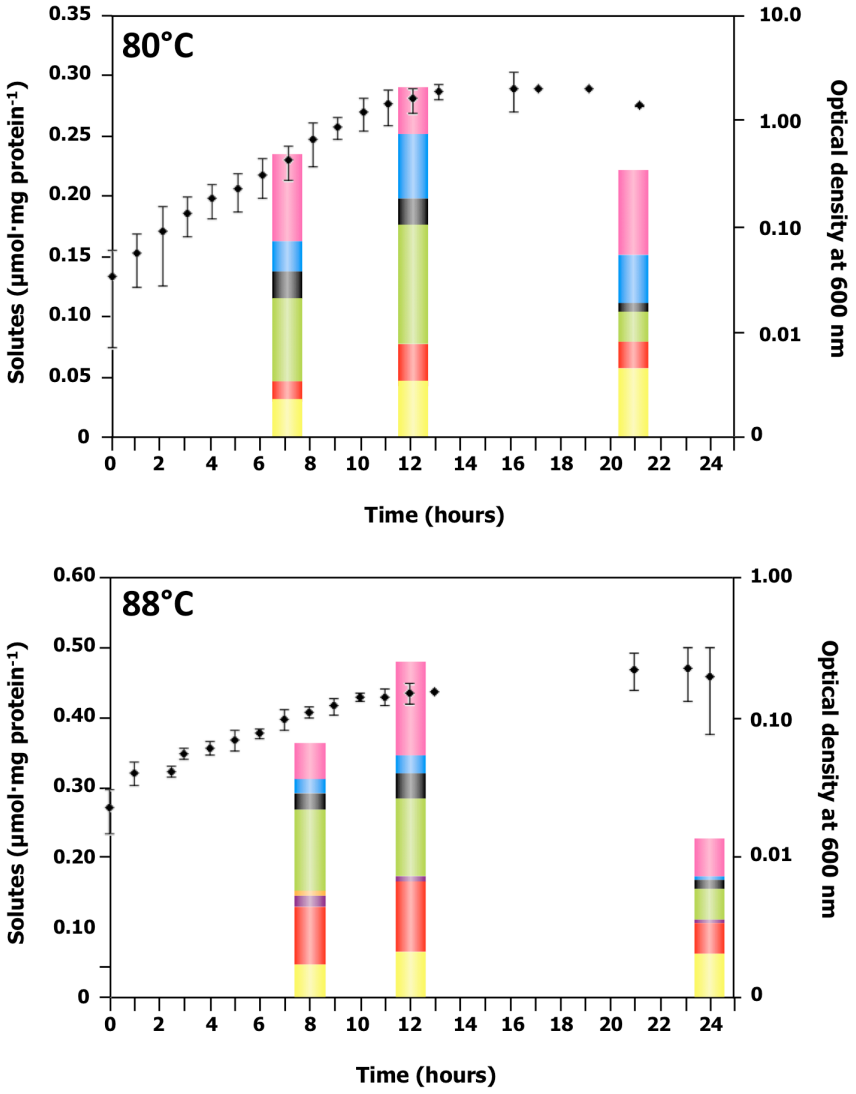


Figure A1. Growth (◆) and accumulation of compatible solutes (bars) in *Thermotoga maritima* cultivated in medium containing 2.2% (wt/vol) of NaCl at the following temperatures: upper panel, 80ºC (optimal growth temperature) and lower panel, 88ºC (supra-optimal temperature). The results are averages of at least two independent determinations. Proline (■, light green); aspartate (■, black); α-glutamate (■, pink); β-glutamate (■, blue); DIP (■, yellow); MDIP (■, red); MMDIP (■, purple); MMMDIP (■, orange). DIP, di-*myo*-inositol 1,3'-phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate; MMMDIP, tri-mannosyl-di-*myo*-inositol 1,3'-phosphate.

Effect of heat-shock on the accumulation of organic solutes in *Thermotoga maritima*

For determination of intracellular solutes the first sample was withdrawn at time 15 min after the target temperature was reached. It is important to note that due to the large culture volume (4 liter) used in the experiment, approximately 20 min were needed to reach the final temperature of 88ºC. Heat-shock was imposed for 200 min, after which the temperature was shifted back to 80ºC. Cell growth was definitely arrested at 88ºC (Fig. A2).

MMDIP was the only new solute accumulating in response to high temperature and was detected in the very first sampling (at 15 min). The α-glutamate content increased progressively until time 90 min. The level of MDIP increased 2.5-fold at time 200 min, while the level of DIP, aspartate, α– and β-glutamate remained approximately constant throughout the heat-shock challenge. Proline accumulates to a considerable level, but does not display a regular pattern.

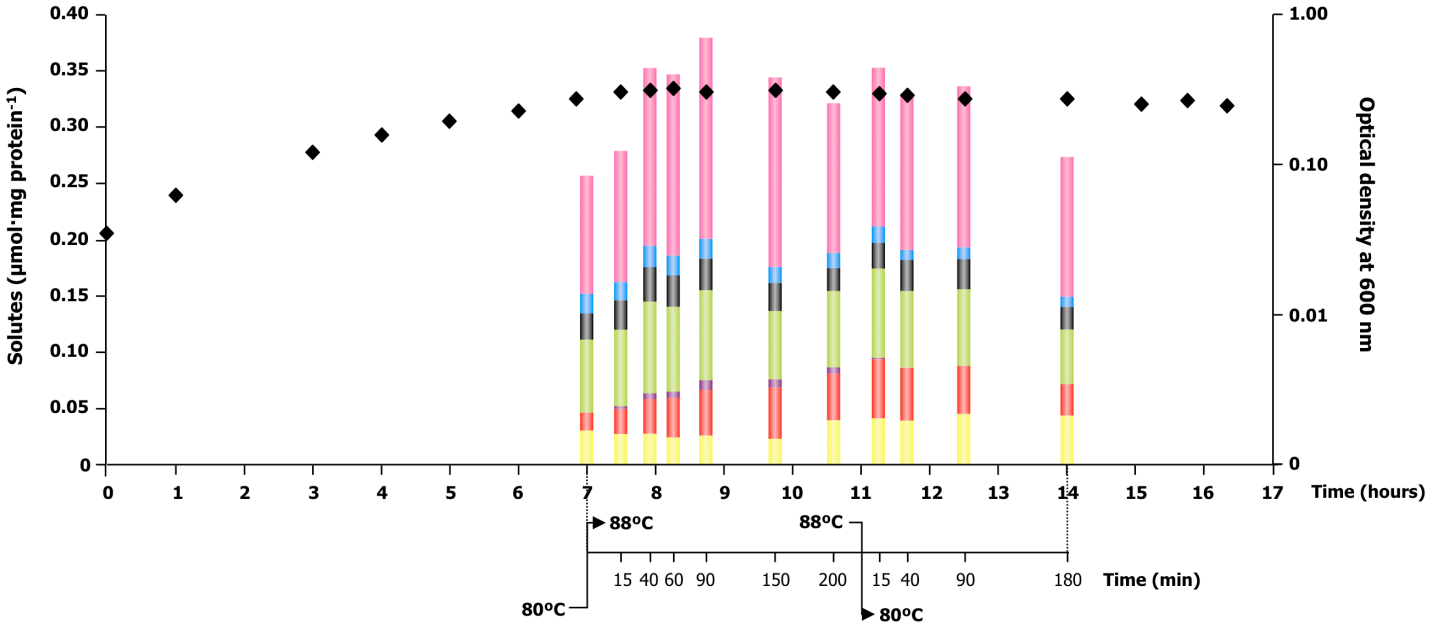


Figure A2. Heat-shock response of *Thermotoga maritima*: growth (◆) and accumulation of compatible solutes (bars). The growth media contained 2.2% (wt/vol) of NaCl. The results represent a single experiment. Proline (■, light green); aspartate (■, black); α-glutamate (■, pink); β-glutamate (■, blue); DIP (■, yellow); MDIP (■, red); and MMDIP (■, purple). DIP, di-*myo*-inositol 1,3'-phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate; MMMDIP, tri-mannosyl-di-*myo*-inositol 1,3'-phosphate.

The maximum pool of organic solutes was detected at 90 min after the temperature up-shift, reaching around 0.4 μmol·mg protein^{-1}, with α-glutamate as the predominant solute (0.18 μmol·mg protein^{-1}), followed by proline (0.08 μmol·mg protein^{-1}) and MDIP (0.04 μmol·mg protein^{-1}) (Table A3 and Fig. A2). At time 90 min at 88ºC, and until the end of the shock, there was a slight decrease in the amino acid content, while the levels of DIP, MDIP, and MMDIP remained fairly constant. At time 200 min, the heat shock was stopped by shifting the temperature back to the optimal growth temperature and the recovery of the cells in respect with the composition of the solute pool was examined. At time 15 min after the temperature down-shift, MMDIP practically disappeared, while MDIP increased by a similar amount; the levels of the other solutes remained unaltered. It is interesting to note that after 180 min at 80ºC the total amount of amino acids recovered the initial values (before the temperature increase) and that the total amount of DIP and derivatives increased 1.5-fold.

Table A3. Intracellular content of organic solutes in *Thermotoga maritima* throughout the heat-shock challenge. Cells were grown at optimal temperature (80ºC) until mid-exponential phase (7 h); then, the temperature was raised to 88ºC and samples were collected at several time points up to 200 min. After this time period, the temperature was shifted back to 80ºC (see Materials and methods for details). Quantification was performed by ^{1}H-NMR. Results refer to single experiments.

	Temp. (ºC)	Time of growth	Intracellular concentration (x10^2 μmol·mg protein^{-1})[a] ± stdv (x10^2)							
			MMMDIP	MMDIP	MDIP	DIP	β-glu	α-glu	Asp	Pro
Before heat-shock	**80**	7 h	nd	nd	1.6 (3.7)	2.7 (6.6)	1.9 (4.0)	8.8 (23.4)	2.0 (5.2)	6.5 (14.4)
Heat-shock	**88**	15 min	nd	<0.5	2.2 (4.9)	2.7 (5.9)	1.6 (3.5)	11.7 (25.9)	2.7 (6.1)	6.8 (15.0)
		40 min	nd	0.5 (1.2)	3.2 (7.1)	2.7 (6.0)	1.9 (4.3)	15.8 (35.0)	3.0 (6.8)	8.1 (18.0)
		60 min	nd	0.5 (1.2)	3.6 (8.0)	2.3 (5.3)	1.8 (4.0)	16.1 (35.7)	2.7 (6.1)	7.6 (16.9)
		90 min	nd	0.8 (1.8)	4.2 (9.3)	2.6 (5.7)	1.8 (4.0)	17.7 (39.4)	2.7 (6.0)	8.1 (17.9)
		150 min	nd	0.8 (1.7)	4.5 (10.0)	2.3 (5.2)	1.5 (3.3)	16.8 (37.2)	2.4 (5.4)	6.1 (13.7)
		200 min	nd	0.6 (1.2)	4.2 (9.2)	3.8 (8.4)	1.3 (2.8)	13.3 (29.7)	2.1 (4.8)	6.8 (15.1)
Recovery at optimal temperature	**80**	15 min	nd	<0.5	5.2 (11.6)	4.0 (8.8)	1.3 (3.0)	13.9 (30.8)	2.4 (5.4)	7.9 (17.5)
		40 min	nd	nd	4.7 (10.4)	3.9 (8.7)	0.9 (2.1)	13.8 (30.6)	2.7 (6.1)	6.8 (15.1)
		90 min	nd	nd	4.3 (9.6)	4.4 (9.7)	1.1 (2.4)	14.3 (31.7)	2.8 (6.2)	6.8 (15.1)
		180 min	nd	nd	2.8 (6.3)	4.3 (9.6)	0.9 (2.0)	12.4 (27.5)	2.0 (4.4)	5.0 (11.0)

[a] The intracellular concentration (mM) based on a cell volume of 4.5 μl·mg protein^{-1} as determined for *Pyrococcus furiosus* (Martins *et al*. 1995) is given in parenthesis. DIP, di-*myo*-inositol 1,3'-phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate; MMMDIP, tri-mannosyl-di-*myo*-inositol 1,3'-phosphate; nd, below the detection limit.

Effect of heat-shock on the transcription of genes involved in the synthesis of DIP, MDIP, MMDIP

RT-PCR was used to study the transcriptional heat-shock response of genes involved in the synthesis of DIP, MDIP and MMDIP in *Thermotoga maritima*. Figure A3 shows the transcript levels of the relevant genes as well as a schematic representation of the pathways for the synthesis of DIP and derivatives.

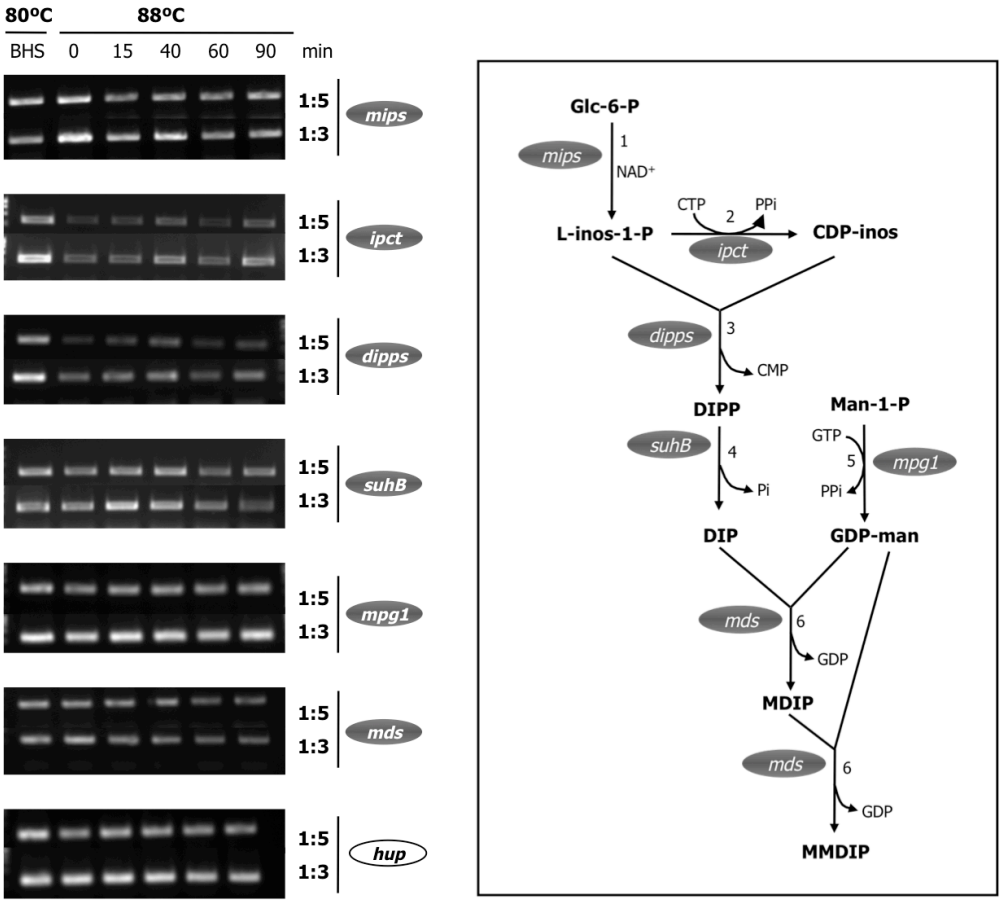


Figure A3. Transcriptional heat-shock response of the genes encoding enzymes implicated in DIP, MDIP and MMDIP synthesis in *Thermotoga maritima*. The cDNA samples were diluted 1:3 and 1:5 and used as template in the PCR analysis. PCR

experiments were performed using primers (Table A1) designed to amplify intragenic regions. The *hup* gene, encoding the DNA-binding protein HU, was used as a control. Control PCR experiments were performed (not shown) using as template chromosomal DNA of *Thermotoga maritima* (positive control) and RNA treated without reverse transcriptase (negative control). The inset shows a schematic representation of the biosynthesis of DIP, MDIP, and MMDIP. Enzymes (represented by numbers) and the respective genes (highlighted with grey boxes) are as follows: 1, *mips*, *myo*-inositol 1-phosphate synthase; 2, *ipct*, CTP:inositol 1-phosphate cytidylyltransferase; 3, *dipps*, CDP-inositol:inositol 1-phosphate transferase; 4, *suhB*, inositol 1-monophosphatase; 5, *mpg1*, mannose 1-phosphate guanylyltransferase; 6, *mds*, mannosyl-di-*myo*-inositol 1,3'-phosphate synthase. BHS, before heat-shock; Glc-6-P, glucose 6-phosphate; L-inos-1-P, L-*myo*-inositol 1-phosphate; CDP-inos, CDP-L-*myo*-inositol; Man-1-P, mannose 1-phosphate; DIP, di-*myo*-inositol 1,3'-phosphate; DIPP, di-*myo*-inositol 1,3'-phosphate phosphate; MDIP, mannosyl-di-*myo*-inositol 1,3'-phosphate; MMDIP, di-mannosyl-di-*myo*-inositol 1,3'-phosphate.

As stated above, 20 min was the time required to reach temperature equilibrium of the culture after an 8ºC increase. This may have led to loss of significant information as changes in transcript levels were detected at the end of this equilibration period (Table A4). The *mips* gene is up-regulated upon heat-shock, while the genes *ipct*, *dipps*, *suhB*, and *mpg1* are down-regulated. The transcript level of the *mds* gene was not affected by heat-shock.

Table A4. Changes in transcriptional levels of the target genes between the culture sample collected immediately before the temperature up-shift and the sample collected 20 min later, when the temperature of the culture reached equilibrium at 88ºC.

	mips	***ipct***	***dipps***	***suhB***	***mpg1***	***mds***
Response upon heat-shock[a]	↑	↓	↓	↓ (s)	↓ (s)	=

[a] Results from at least two independent PCR experiments, except for the *mips* gene (see materials and methods for details). ↑, up-regulated; ↓, down-regulated; ↓ (s), slightly down-regulated; =, maintained. Abbreviations as in Fig. A3.

The heat-shock conditions were maintained for 200 min; for the RT-PCR experiments samples were withdrawn at times 0, 15, 40, 60 and 90 min. In some cases the initial increase or decrease in the transcript level was not maintained in subsequent samples (Fig. A3). Transcription of *mips* appears to be slightly up-regulated by heat-shock (time zero), but subsequently the level of the transcript decreased nearly to the basal one (before heat-shock). The transcript profiles of *ipct* and *dipps* were very similar. The expression of these genes appears to be down-regulated upon heat-shock. On the other hand, the transcription of *suhB*, the gene encoding a putative DIPP phosphatase, exhibits minor differences along the heat-shock insult. Transcription of the *mds* gene was not affected.

Discussion

Thermotoga maritima was cultivated in a medium containing glucose as carbon source instead of starch as described in Chapter 5. This was done to allow the use of specifically labeled glucose to trace metabolic products derived from glucose by ^{13}C-NMR analysis.

The pool of organic solutes comprised α- and β-glutamate, DIP, MDIP, and MMDIP, similarly to what was reported earlier (Chapter 5). However, proline and aspartate were additional solutes. Also noteworthy is the higher level of amino acids accumulated. We believe that the differences in amino acid accumulation are mainly due to the different nitrogen sources used in the growth medium: tryptone was used as nitrogen source in this work, while yeast extract was used in the experiments described in Chapter 5.

There was a differential accumulation of solutes along growth. The pool of solutes reached a maximum during the late exponential phase of growth, being dominated by amino acids, specially proline and α-glutamate. In

the stationary phase, the total amount of solutes decreased, especially due to the decrease in the total amino acid pool. A decrease in the pool of solutes during the stationary phase of growth has been observed for other (hyper)thermophiles (Lamosa et al. 1998, Silva et al. 1999, Gonçalves et al. 2003). Although a definite explanation is not available, it has been proposed that the decrease in the amount of solutes is due to loss to the external medium or to the catabolism of these compounds to preserve cell viability under nutrient-depleted conditions. Nothing is known about the catabolism or transport of DIP and derivatives across the cellular membrane, thus this matter demands further investigation.

To our knowledge, this is the first work on the effect of heat-shock on the accumulation of compatible solutes in *Thermotoga maritima*. This bacterium responded to an 8ºC increase in the growth temperature with the accumulation of MMDIP. Upon heat-shock, there was a small decrease in the level of DIP, which apparently was counterbalanced by an increase in the pools of MDIP and MMDIP. During the final stages of the heat insult (time 200 min of growth at 88ºC) and the recovery period at 80ºC, it is notable that the total pool of DIP and MDIP was maintained. Although nothing is known about the catabolism of compatible solutes such as DIP or MDIP, we speculate that MDIP would be acted upon by a mannosidase, yielding DIP, hence the decrease in MDIP should lead to a concomitant DIP increase. As mentioned above, it would be interesting to understand the fate of such solutes, hence further investigation is needed on the catabolism of DIP and derivatives.

The main goal of this work was to perform an analysis of transcriptonal and intracellular solute changes in *Thermotoga maritima* as it undergoes a heat response from 80ºC to 88ºC. We focused on DIP and derivatives and combined the results from NMR spectroscopy, used for quantification of intracellular solutes, and RT-PCR analysis used to measure

the transcript levels of the genes involved in the synthesis of DIP and MDIP. However, the limitations of our experimental set up, especially the long time required for temperature equilibration, do not allow for definite conclusions.

Generally, stress responses are transient and after the initial changes, the expression of genes stabilizes in levels similar to those of unstressed cells even if the stress persists (Lopez-Maury et al. 2008, Jozefczuk et al. 2010). It is conceivable that due to our experimental approach/limitations, eventual initial changes in transcript levels were missed. Future efforts should be directed to increase the sensitivity of the solute detection method so that small culture volumes are manipulated. Mass spectrometry should be considered as an alternative analytical methodology.

Upon heat-shock, the transcript levels of *ipct*, *dipps*, and *mds* did not show a direct correlation with the levels of DIP, MDIP and MMDIP. In fact, while the amount of MDIP and MMDIP clearly increased upon heat-shock, the level of *mds* transcript remained unaltered throughout the experiment. These results combined with the fact that MDIP synthase has maximal activity at approx. 95ºC (15ºC above the optimal growth temperature of *Thermotoga maritima*) (Chapter 5), supports the view that the regulation of the synthesis of MDIP/MMDIP is under the control of enzyme activity.

A more intriguing result is the decrease in *ipct* and *dipps* transcripts in response to the heat challenge. Given that the level of MDIP increased with heat-shock, and DIP is the substrate for the synthesis of MDIP, it is clear that the activity of IPCT/DIPPS must go up. As for MDIP, it appears that the control of DIP synthesis occurs downstream the transcriptional step. Recent results on the biochemical characterization of IPCT from *Archaeoglobus fulgidus* show that the enzyme has optimal activity between 90º and 95ºC, approximately 10°C above the optimal growth temperature of the organism

(Brito et al. 2011). On the basis of the evidence available we postulate that the regulation of DIP synthesis occurs at the level of enzyme activity.

The genomic organization and flanking regions of the *ipct*/*dipps* genes in *Thermotoga martima* indicate an operon-like structure with the *mips* and *suhB* genes. In an operon, the genes are under the control of the same promoter, hence they are expected to have the same transcriptional regulation. However, different transcriptional responses were observed especially between the *mips* gene and the *ipct*, *dipps*, and *suhB* genes, suggesting the involvement of different promoters. Further experiments are needed to validate this hypothesis. Nevertheless, *ipct* and *dipps* are probably under the same type of transcriptional regulation given that in most organisms with homologous sequences, *ipct* and *dipps* are fused in a single gene product (Chapter 3, Rodionov et al. 2007).

The transcription of *ipct*/*dipps* genes has been earlier investigated in *Thermococcus kodakarensis* by Borges and colleagues (Borges et al. 2010), who showed that the genes are up-regulated when the organism was cultivated at supra-optimal temperature. Apparently, in this archaeon *ipct/dipps* are regulated by a putative transcriptional regulator, Phr, which represses the synthesis of these genes under optimal growth temperature (Kanai et al. 2010). A question then arises concerning the existence of different types of regulation depending on whether the cells were allowed to adapt to the supra-optimal temperatures (the case of *Thermococcus kodakarensis*) or were suddenly submitted to a heat shock (this work). Studies on the cyanobacterium *Synechocystis* sp. strain PCC 6803 revealed that cells acclimated to increasing salt concentrations show a positive correlation between the level of glucosylglycerol, the level of *ggpS* transcript, and the level of the respective enzyme (GgpS) (Marin et al. 2002). However, the same trend is not observed when *Synechocystis* cells were submitted to a sudden

osmotic shock. In this case, while the level of *ggpS* increased rapidly upon osmotic shock, and stabilizing subsequently at a level slightly higher than the original one, the levels of GgpS and glucosylglycerol increase smoothly during the osmotic challenge (Marin et al. 2002). In other words, while the transcription was abruptly disturbed, the translation and protein activity responded gradually to the imposed stress.

The heat-shock response of *Thermotoga maritima* has been investigated by Pysz and co-workers (2004) using cDNA microarrays to study the dynamic changes in gene expression. The cDNA microarray covered about 20% of the genome, but the set of genes examined in our work are not mentioned. The heat-shock response has been studied in two other DIP producers, *Archaeoglobus fulgidus* and *Pyrococcus furiosus* (Rohlin et al. 2005, Shockley et al. 2003). At that time the function of *ipct* and *dipps* were not known, but *mips* and *suhB* were assigned. The transcriptional heat-shock response of *Archaeoglobus fulgidus* revealed that the *mips* gene is not differentially regulated upon heat-shock (Rohlin et al. 2005). However, contradictory results have been reported for the closely related archaeon, *Pyrococcus furiosus*, in which the *mips* gene is strongly induced by thermal stress (Shockley et al. 2003). Our results suggested an initial increase in the transcript level, but then the amount of transcript returned to its "basal" level. These discrepant results indicate that *mips* is not a target for regulation of DIP synthesis. It is difficult to rationalize these results, given that the activity of MIPS and IMP are not exclusively committed to DIP synthesis. In fact, *mips* encodes an enzyme that operates in many organisms as the only supplier of the inositol ring, and it is the first enzyme in the process of biosynthesis of any inositol-containing molecule (Chen et al. 2000, Stieglitz et al. 2005).

Concluding remarks

The typical heat-shock machinery of hyperthermophiles, composed by the heat-shock proteins, is slowly being uncovered by studies on *Pyrococcus furiosus*, *Archaeoglobus fulgidus*, *Methanocaldococcus jannaschii*, *Sulfolobus solfataricus*, and *Thermotoga maritima* (Tachdjian and Kelly 2006, Shockley et al. 2003, Rohlin et al. 2005, Boonyaratanakornkit et al. 2007, Pysz et al. 2004). Other factors, such as biofilm formation or accumulation of compatible solutes are now recognized as important contributors to the thermotolerance of these organisms (Pysz et al. 2004).

Little is known about the regulation of the synthesis of compatible solutes in hyperthermophiles (Santos et al. 2007). In this work we combined transcriptional data on the relevant genes for the synthesis of DIP and MDIP, with the levels of these solutes upon heat shock. A direct correlation between transcript and solute levels was not observed. Curiously, the genes *ipct* and *dipps* were down-regulated upon heat-shock, while the levels of DIP derivatives were clearly increased. However, the transcriptional response may occur in a shorter time interval than that accessed in this work, so the hypothesis that he regulation of the synthesis of these solute may occur at the transcriptional level can not be ruled out definitely. Most likely, the regulation occurs at different levels, *i.e.*, transcriptional, translational, and at the level of metabolites. Several proteomics approaches are now available for the determination of protein levels, but measuring metabolite levels is not as straightforward as assessing protein or mRNA or levels. To our knowledge, the heat-shock response of *Thermotoga* species has not been investigated at the global proteomic level, and we consider this is an interesting line for future research.

Acknowledgments and work contributions

Cristiana Faria assisted in the growth and extraction of compatible solutes of *Thermotoga maritima*. I thank Dr. Paula Gaspar and Dr. Rute Castro for advice on the process of RNA extraction and Dr. Fátima Cairrão for the DNase I used to test the RNA extraction protocol. The NMR spectrometers at CERMAX are part of the National NMR Network and were acquired with funds from FCT and FEDER. This work was funded by Fundação para a Ciência e a Tecnologia, POCTI Portugal, and FEDER Projects PTDC/BIA-MIC/71146/2006. M. V. Rodrigues acknowledges FCT for the research fellowship.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt