-Omics study of *Sulfolobus solfataricus* species under different carbon sources

Wen Qiu

A thesis submitted for the degree of Doctor of Philosophy

Department of Chemical and Biological Engineering

The University of Sheffield

March 2015

Dedication

I dedicate this thesis to my parents and family members, especially my beloved parents. Also to my lovely friends, teachers and anyone who helps me in every aspect of my life.

Declaration

This is a declaration to state that this thesis is an account of the author's work which was conducted at the University of Sheffield, UK. This work has not been submitted for any other degree of qualification.

Acknowledgement

I would like to thank for all the assistances throughout my PhD research. It would not be possible to complete the whole work without their helps. A special acknowledgement to my supervisor, Professor Phillip C Wright, for his encouragement, inspiration, guidance and supports during my research.

I am really grateful to Dr. Khoa Trong Pham for his kindly helps and guidance, during my PhD study. I express my deep gratitude to Dr. Caroline Evans for her corrections especially for the short time frame to read whole thesis. I would like to express thanks to Dr. Josselin Noirel for checking one page essay to improve writing skills in the beginning of studies. I would like to express many thanks to Dr. Narciso Couto for advice and help with MS instrumentation.

I would like to pay my gratitude to previous post-doc Dr. Saw-Yen Ow and Dr. Malinda Salim for their enthusiastic guidance and support in the beginning of my PhD study. I also acknowledge my friend Xueyan Zhao for writing Matlab program for data analysis in Chapter 5. I would like to express my thanks to Andrew Landels for help me script for data analysis (the loop script is fantastic) in Chapter 5 and Rahul Kapoore for his help in writing Macros GC-MS data analysis.

A big thank you to all my friends in chemical and biological engineering department and my lovely flatmates and friends.

Finally, many sincere thanks to the China-UK scholarships for Excellence programme and China Scholarship Council (CSC) for providing financial support. And also thank ChELSI and EPSRC programs for their supports.

Abstract

The acidic thermophilic archaeon *Sulfolobus solfataricus* has been widely used as a model organism in archaeal research and it can utilise different compounds (tryptone or glucose) as carbon sources. However, the understanding of this archaeon in responding to these carbon sources (at quantitative proteome level) and its regulation (at phospho proteome level) has not been well studied yet. Therefore, I would like to investigate the effects of different carbon sources (tryptone and glucose) to different *S. solfataricus* strains (P2, PBL2025 and PBL2073). Furthermore, a P-peptide (phospho peptide) enrichment protocol was optimised. Quantitative phosphoproteomic analysis was performed for *S. solfataricus* P2 using optimised P-peptide enrichment and iTRAQ strategy.

In order to achieve these goals, firstly, the growth effect on different *S. solfatarcius* strains from different carbon sources was monitored to determine the sampling times. Sample preparation and technical optimisation for quantitative proteomic and phosphoproteomic studies were then performed.

Secondly, both proteomic (iTRAQ) and metabolomic (GC-MS) tools were applied to determine the global proteomic and metabolomic changes of three different *S. solfatarcius* strains in responding to two carbon sources (glucose and tryptone). In this thesis, an investigation of the global proteomic changes between *S. solfatarcius* strains PBL2025 and P2 was carried out. As a result, a total of 158 proteins (27% of quantified proteins) showed their abundance changes, when both strains grown on standard glucose media. Of these proteins, more than half of them (61%) involving in carbon fixation, butanoate and-so-forth showed their up regulation, whereas the others (39%) belonging to carbohydrate or amino acid metabolic pathways showed their down regulation. Moreover, 208 and 159 (27% and 34% of quantified proteins) proteins were affected under trypton *vs* glucose for PBL2025 and P2 respectively. The down regulation of detected amino acids pathways were significant in both strains in responding to carbon source change from glucose to tryptone, which was inferred to follow a feedback-inhibited pattern. Metabolomic data show high abundances of detected amino acids, which support for the proteomic observations.

Thirdly, MOAC (metal oxide affinity chromatography)-TiO_2, IMAC (immobilized metal ion affinity chromatography)-Fe and combination of these two (SIMAC) were tested for P-peptide enrichment of tryptic digested protein of *S. solfataricus* grown on standard media. Based on optimised P-peptides strategies, a combination of complementary strategies: TiO_2 magnetic beads and the SIMAC was applied to investigate the global phosphoproteomic changes in responding to different carbon sources (for *S. solfataricus* P2). Most of the quantified P-proteins involving in carbohydrate metabolism were unaffected, while P-proteins functioning in amino acid metabolism were up or down regulated.

This thesis reports the quantitative proteomic and metabolomic changes in three different *S. solfataricus* strains in responding to different carbon sources. An optimised P-peptide enrichment strategy was also established, which may be applied for other achaeal studies. Furthermore, quantitative phosphoproteomic study was performed for *S. solfataricus* P2, and this was recognised as the first global quantitative phosphoproteomic study in this archaeon.

List of abbreviations

ABC transporter:	ATP-binding cassette transporters
ABC:	Ammonium bicarbonate
ACN:	Acetonitrile
BSA:	Bovine serum albumin
CCM:	Central carbohydrate metabolism
CID:	Collision-induced dissociation
ERLIC:	Electrostatic repulsion hydrophilic interaction chromatography
ESI:	Electrospray ionization
ETD:	Electron transfer dissociation
GC-MS:	Gas chromatography–mass spectrometry
HILIC:	Hydrophilic-interaction liquid chromatography
HPLC:	High performance liquid chromatography
IMAC:	Immobilized metal ion affinity chromatography
IMAC-Fe:	Fe PHOS-Select iron affinity gel
iTRAQ:	Isobaric tags for relative and absolute quantitation
LoD:	Limit of detection
MOAC:	Metal oxide affinity chromatography
MS:	Mass spectrometry
OD:	Optical density
P-peptide:	Phospho peptide
P-protein:	Phospho protein
P-site	Phospho site
SCX:	Strong cation exchange
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIMAC:	Sequential elution from IMAC
TEAB:	Triethyl ammonium bicarbonate
TiO_2 :	Titanium dioxide

Contents

Chapter 1 Introduction

1.1 Background

Sulfolobus solfataricus P2 has been established as a model organism in the third domain of life: archaea [1, 2]. Its optimal growth occurs at temperature ranging from 75 to 80°C and at pH 2-3 [2, 3]. Research on the thermostable crenarchaea *Sulfolobus* spp. has attracted community's interest due to its growth in extreme high temperature and acidic growth conditions, which made them promising candidates for industrial applications. Some studies on re-construction of the central carbohydrate metabolic (CCM) pathways of *S. solfataricus* P2 have been intensively performed [4], however, these pathways have not been completely understood, since almost half of the genes/proteins have not been characterised yet [5, 6]. Research on amino acid metabolisms of *S. solfataricus* at a system level has not been investigated. Through a systematic proteomic study, most phosphoproteins identified were involved in the metabolism process [7]. However, until now, there was no report concerning on quantitative phosphoproteomic analysis in *S. solfataricus* at system-wide level has been reported yet.

1.2 Aims and objectives

The overall objective of this thesis is to fill a gap in amino acid metabolism and gain a good understanding of CCM of *S. solfataricus* at a system level by interpretating the -omics (proteomic and metabolomic) data from growth on different carbon sources.

In this PhD study, the first objective is to gain an understanding of changes in protein abundances and affected metabolites through comparasion of both proteomic and metabolomic aspects for assessing on carbohydrate pathways especially CCM and amino acid pathways of *S. solfataricus*. In addition, protein phosphorylation was proposed to be important due to its potential invoving in CCM regulation hence in responsing to change of carbon sources from glucose to tryptone [7]. I aim to optimise an efficient P-peptide enrichment strategy for further investigation of quantitative phosphoproteome changes in *S. solfataricus* P2 comparing different growth on glucose *vs* tryptone, from which contributions will be drawn to gain a better understanding of the regulation of CCM.

To address these aims, three different *S. solfataricus* strains including (P2; PBL2025, a spontaneous mutant; and PBL2073, a genetic modified strain) and different carbon sources (glucose, trypton, ethanol, acetone, propanol and iso-propanol) were used. Both global quantitative proteomic (iTRAQ expriments) and metabolomic (GC-MS) experiments were performed to generate in-depth data sets. Moreover, an efficient phosphopeptide enrichment strategy (SIMAC: combination of TiO_2 and IMAC) was optimised before a large scale phosphoproteomics study using iTRAQ was performed. Furthermore, integration of both proteomic and metabolomic data were applied to infer biological changes involved in carbohydrate and amino acid metabolic pathways in *S. solfataricus*. An overview of the contents in this thesis and their relationships are summarised in Fig 1.1. Aims and objectives are listed and contributions from PhD work are also summarised.

Chapter 1: Brief introduction of connections between different chapters in this thesis.

Chapter 2: Literature review of -omics (proteomic and metabolomic) studies in archaea especially in *S. solfataricus* with a focus on phophoproteomic studies.

Chapter 3: The first major task within this thesis is to do methodology optimisation for global proteome identification and characterisation of *S. solfataricus*. Quick test using both SDS-PAGE gel digestion and in-solution digestion compled with SCX fractionation were performed. Furthermore, P-peptide identification was also tested using different MS instruments, and a limit of detection of phosphopeptide using a UHR Q-TOF maXis instrument was determined with β-casein used as a standard.

Chapter 4: Based on above optimised protocols, 3 sets of 8-plex iTRAQ experiments were performed to investigate the responses of *S. solfataricus* strains (P2, PBL2025 and PBL2073) to different carbon sources including 0.4% glucose and 0.2% tryptone at a quantitative proteomic level. Bioinformatic analysis was applied to identify proteins that changed their abundances in various comparisons.

Chapter 5: This study focused on optimisation of P-peptide enrichment methodologies using a tryptic digest of whole cell lysates of *S. solfataricus* grown on 0.4% glucose media. Different affinity-based resins were tested including different TiO_2 formats (pipette tip, spin tip and magnetic beads), IMAC-Fe (PHOS-select iron affinity gel) and modified

SIMAC (a combination of both TiO_2 and IMAC-Fe). Optimisation processes were performed in aspects including the use of non-phosphopeptide inhibitors, peptides-to-beads ratio, and consecutive incubations for TiO_2 beads and elution buffers for IMAC. As a result, a modified SIMAC protocol was employed for a better phosphoproteome coverage.

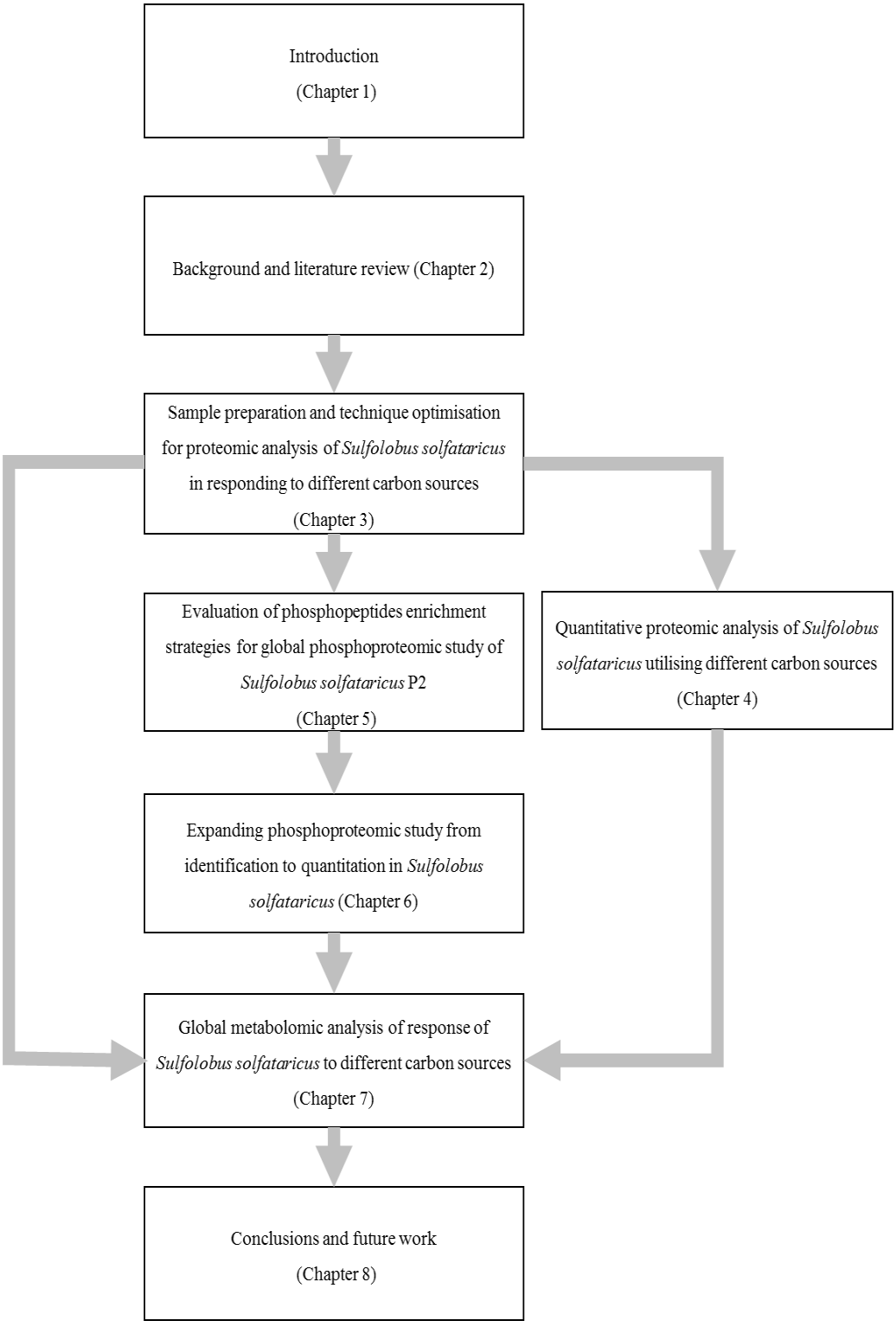

Fig 1.1 Overview of thesis contents and relationship of chapters are summarised.

Chapter 6: Global quantitative phosphoproteome study of *S. solfataricus* P2 grown on 0.4% glucose compared to 0.2% tryptone was performed to investigate whether the phosphorylated protein abundances were regulated or not by changing carbon source. In total, 109 unique P-peptides from 100 P-proteins were quantified with a determination of 147 P-sites on Ser/Thr/Tyr. Most of quantified P-proteins functioning in carbohydrate metabolisms showed no abundance change, whereas P-proteins involving in amino acids metabolisms regulated.

Chapter 7: GC-MS technique was applied to investigate the metabolomic changes of *S. solfataricus* strains (P2, PBL2025 and PBL2073) in responding to different carbon sources (0.4% glucose *vs* 0.2% tryptone). Global metabolomic data was submitted to an on-line XCMS software for statistical analysis. Identification of compounds was achieved using AMDIS and searched against NIST 2.0 library. Among 163 identified compounds from all experiments, 113 metabolites corresponding C number in KEGG pathway were found. However, only 47 of them were involved in metabolic pathways.

A brief comparison of -omics (proteomic data from Chapter 4 and metabolomics data from Chapter 7) data was interpreted to show the response of *S. solfataricus* strains to different carbon sources in system level.

Chapter 8: Conclusions and future work.

1.3 Contributions from PhD work

Responses of quantitative proteomic changes of different strains of *S. solfataricus* including P2, PBL2025 and PBL2073 to different carbon sources (glucose and tryptone) were reported. All experiments including proteomic and metabolomic work were performed using the same biological organisms grown on/treated under the same conditions, which provide valuable -omics information for future modelling work. Furthermore, information of regulated proteins may provide useful knowledge for further biological researches such as biofilm studies using PBL2025 and P2.

An optimised P-peptide enrichment strategy was established for further global quantitative phosphoproteomic study of *S. solfataricus*. Compared to enrichment-free precursor acquisition independent from ion count (PAcIFIC) technique [7, 8], the

advantages of enrichment strategy is a shorter MS sample running and data analysis time: 2 weeks *vs* 8 weeks. Also, it enables quantitative detection of phosphorylated proteins/peptides when coupled with label based proteomic tool such as iTRAQ or TMT.

The first global quantitative phosphoproteomic study in archaea using iTRAQ was performed in the model strain *S. solfataricus* P2. Future biological work can be performed based on the brief construction of protein-protein interaction especially quantified P-proteins. Moreover, promising target proteins were selected for PTMs cross-talk between phosphorylation and acetylation. Furthermore, target P-proteins may be selected for CCM regulation mechanism studies. As discussed below (Section 2.2.4), protein phosphorylation seems to play a major role in almost all the cellular process [9]. It is suggested that proteins undergo the rerversible phosphorylation/dephosphorylation process provide a tight reglulation on sugar metabolism in *S. solfataricus* [7]. For instance, enzymes at branch point such as 2-keto-3-deoxy gluconate aldolase involves in npED and spED metabolic branch pathway, isocitrate dehydrogenase involved in TCA and glyoxylate pathway [7]. The last but not the least, investigation on target P-protein show potential benefit for industrial applications. For instance, the successful expression of small heat shock protein (Sso-HSP20, encoded by gene *SSO2427* in *S. solfataricus*) provide protection of *Escherichia coli* on temperature shock at both 50 °C and 4°C [10]. In addition, it shed lights on the indusrial application of P-protein through reducing the energy consumption for temperature cooling down processes.

Chapter 2 Background and literature review

2.1 Introduction

2.1.1 Archaea

Archaea has been established as the third domain of life based upon ribosomal RNA sequence characterisation in late 1970s [11]. Since then, Eucarya, Bacteria, and Archaea are proposed as the three domains of life on the earth, while Bacteria and Archaea belong to the Prokaryotes [12]. It was termed as 'archaebacteria', but later changed to Archaea [11, 12]. Although there are a few common characteristics shared between Archaea and Bacteria, for instance, physical cell structures and basic metabolic pathways; there are still different attributes from biochemical and genetic aspects [12]. Interestingly, the genetic information processing system of bacteria-like archaea shows striking similarity to that of Eukaryotes. For instance, the chromosome replication-related proteins of *S. solfataricus* P2 were found to be more related to its eukaryotic counterparts but distinct from proteins in bacteria [3]. Gribaldo and Brochier-Armanet [13] summarised characteristics of Archea by stating that 'Archaea look like organisms that use eukaryotic-like proteins in a bacterial-like context'.

2.1.2 *Sulfolobus solfataricus*

Archaea can be divided into three kingdoms: Crenarchaeota, Euryarchaeota and Korarchaeota, based on their philosophy and particular habitats [14]. *S. solfataricus,* a hyperthermophilic Crenarchaeota, grows in sulfur-rich acidic hot springs at temperatures up to 90°C (optimum 75-80°C) and at pH values of 1-5 (optimum pH 2-3) [15] .

Sulfolobus spp. has been developed as a model organism for genetics studies [16], especially when the whole genome sequences became available for *S. solfataricus* [3], *Sulfolobus acidocaldarius* [17] *and Sulfolobus tokodaii* [18]. Research on thermostable enzymes of thermophilic crenarchaea of *Sulfolobus* spp. have generated considerable interests due to their activity at high temperature and extreme acidic conditions which are generally toxic to other organisms [19].

S. solfataricus can use various carbon sources such as tryptone, a variety of sugars and amino acids [20] as an energy source. *S. solfataricus* P2 was isolated in Italy and has been considered as a model organism representative of the archaeal domain [1, 2]. *S. solfataricus* 98/2 was isolated from United States [21, 22]. *S. solfataricus* PBL2025, which was the spontaneous derivative strain of *S. solfataricus* 98/2, has been widely used as genetic tool for molecular biology research. Compared to *S. solfataricus* P2, the strain PBL2025 lacks more than 40 genes (from *SSO3004* to *SSO3050*). The central metabolic pathways may be different between strains P2 and PBL2025 due to the genes not been present (please see Section 2.1.5 for details). *S. solfataricus* PBL2073 was the gene *SSO3117* disruption strain from *S. solfataricus* PBL2025 (from Professor Paul Blum's lab).

2.1.3 Application of archaea, especially *S. solfataricus*

Since their discovery and establishment as the third domain of life, the study of archaea has attracted much attention [23]. Usually, most of the microorganisms belong to archaea domain live in extreme environments [24], eg. halophilic, methanogenic, acidophilic and hyperthermophilic habitats. Their application seems to offer great opportunities for industrial biotechnology due to stability even under harsh pH or temperature conditions [25-28]. For instance, hyperthemophilic archaea survive at extreme acidic and high temperature conditions, and thus their proteins harbour unique properties, which made them promising candidates for industrial applications, e.g. thermostable enzymes [27], bioremediators [29], and biological waste water treatment [30]. Li *et al*.,[10] noticed the expression of small heat shork protein of *S. solfataricus* P2 (S.so-HSP 20) in an *E. coli* expression system was involved in the stress response to temperature change (high temperature up to 50 °C as well as cold shock: 4 °C treatments) based on structural and functional analysis. This study could shed light on cellular thermo tolerance study through investigating the expression of S.so-HSP 20 in different bacterial model system. The study of hyper thermal/acidic stable enzymes in archaea reveals new opportunities for the efficient production biofuel from renewable source, as reviewed by [31]. Take the characterisation of a bi-functional β-xylosidase/α-L-arabinosidase enzyme of *S. solfataricus* P2 as an example [32]. It gained particular interest due to the utilization for the pre-treatment of lignocellulose, before its further hydrolysis/conversion to bioenergy [32].

Isopropanol together with ethanol and methanol are mostly used as alcoholic solvents, which termed as volatile organic compound (VOC) with a total production about 6.5 million tonnes in 2013 (http://www.ceresana.com/en/market-studies/chemicals/solvents/). Atmosphere pollution from these VOCs resulted in numerous waste. Thus efficient control of these environmental unfriendly VOCs is necessary. Biological treatments of these VOC shows great potential benefits. There have been some reports on the biodegradation of isopropanol using thermophile bacteria: *Bacillus pallidus* [33], and hyperthermophilic archaeon *S. solfataricus* [34]. Moreover, there were few reports on the the utilization of alcohols (n-propanol or isopropyl alcohol) and ketones as carbon sources in *S. solfataricus* [34, 35]. Izzo *et al*., and colleagues [36] reported the biodegradation of widespread organic pollutant phenol using the thermoacidophilic *S. solfataricus* 98/2 under aerobic condition. Further study was performed in a fed-batch bioreactor by [37], which indicates it as an ideal bioremediation candidate for the chemical industrial application in waste management under extreme thermalphilic and acidophilic conditions.

There are three sections included in this Chapter:

- Section 1: Current proteomic study in *S. solfataricus* especially the response of *S. solfataricus* P2 to various carbon sources, with a brief review about proteomic techniques used in following studies including SDS-PAGE gel, shotgun proteome and iTRAQ.
- Section 2: Phosphoproteomic study and technical development. Due to limited investigation of the phosphoproteome in archaea, reviews of phosphoproteome methodologies with focus on enrichment techniques were mainly incorporated from other subjects.
- Section 3: Techniques applied in metabolomic studies in archaea.

2.2 Literature review

Section 1: Current -omics studies of S. solfataricus

2.2.1 -Omics studies of *S. solfataricus*

2.2.1.1 Proteomic studies of *S. solfataricus*

Reconstruction of central carbon metabolism (CCM) of *S. solfataricus* in responses to various carbon sources (as shown in Fig 2.1) and different living temperatures have been

performed in model strain P2 by incorporating -omics techniques in different levels including genomic, comparative genomics, transcriptomics and proteomics [38]. Quantitative proteomic analyses of P2 in responding to different carbon sources were successfully carried out using 2DE and ^{15}N metabolic labelling or iTRAQ peptides labelling techniques. For instance, Snijders *et al.*, [39] observed a minor change (3%) in transcript level and a significant change (14%) at protein level when yeast extract and tryptone were supplied as alternative carbon sources to glucose. Moreover, proteome level changes of the P2 strain, in response to alcohols or ketones (ethanol, acetone, n-propanol or iso-propropanol) with or without the presence of glucose were investigated through RT-PCR and iTRAQ by [35, 40]. Growth inhibitory effect was observed with the addition of alcohols or ketones compared to glucose control, but only 6.3% and 5.5% of quantified proteins showed up- and down- abundance changes respectively in proteome level [35]. The majority (88.2%) of identified and quantified proteins remain unchanged, which was speculated resulting from the presence of glucose. Chong *et al*., [40] noticed that *S. solfataricus* cannot survive in isopropanol or acetone only media, but it shows ability to utilise 0.199% or 0.498% (v/v) n-propanol. Further iTRAQ analysis revealed up or down regulations of 36% of detected proteins compared to glucose control [40]. The growth of *S. solfataricus* on various concentrations 0.101%-5.03% (v/v) of ethanol were evaluated with a determination of maximum consumption at 0.79% w/v ethanol [41]. Compared to glucose control, 21% and 31% of determined proteins showed their up and down regulations respectively, through using iTRAQ technique. Furthermore, CCM response of this P2 strain to optimal (80°C) and suboptimal (70°C) living temperatures under standard glucose media were investigated by incorporating the above -omic technologies as well as metabolomics and enzymatic techniques [38]. Standard operation procedures were proposed for future -omics study of *S. solfataricus* [38].

In contrast to the extensive studies of P2, no comprehensive quantitative proteome study has been investigated for commonly used genetic tool: strain PBL2025, not even in relation to its response to different carbon sources. Interest in quantitative proteomics study of these *S. solfataricus* strains was further inspired by the discovery of their attachments to various surfaces and the observation of different biofilm formation between PBL2025 and P2 [42]. Upon surface attachment, a strong induction of 8 genes including (*SSO3007, SSO3009, SSO3010, SSO3014, SSO3017, SSO3019, SSO3035* and *SSO3041*) were observed at transcript level through RT-PCR and quantitative PCR

analysis [42]. The expression of the other 10 tested genes are not affected at transcript level. Another concern on the large scale proteomics study is based on the identification of enzymes involved in extracellular polysaccharides formation in PBL2025, which are encoded by two of missing genes: *SSO3006* and *SSO3019* [43]. Further investigation efforts are required to have a better understanding of proteins presenting at different stoichiometry and the significantly affected pathways between P2 and PBL2025 grown on the same carbon sources as well as the response of one strain to different carbon sources.

2.2.1.2 Phosphoproteomic studies of *S. solfataricus*

Whilst CCM routes are partly well known in *S. solfataricus* (see Fig. 2.1), research on its regulation is relatively limited. Protein phosphorylation has been noticed to play important roles in cell processes as discussed in detail in the following Section 2. Glucose metabolism usually occurs via Entner-Doudoroff (ED) pathway with a non-phosphorylative branch [44]. In 2005, Ahmed and colleges found evidence for ED pathway with a semi-phosphorylative branch (sp-ED) in *S. solfataricus*. The activity of kinase: isocitrate dehydrogenase in the glyoxylate cycle was probably regulated through phosphorylation as proposed for *S. solfataricus* [45] and *S. acidocaldarius* [46]. At the proteome level, the possibility of the regulation of key enzymes involved in the CCM by protein post translation modification (PTM) was pointed out by Snijder *et al* for the first time [39]. Later, Esser and colleagues [7] observed a high number of phosphoproteins (540 phosphoproteins) in *S. solfataricus* P2 when carbon sources were changed from glucose to tryptone. By using a precursor acquisition independent from ion count (PAcIFIC) approach without incorporating any phosphopeptide enrichment technique, they found most of phosphoproteins were especially enriched in the metabolism process [7]. For instance, the observation of phosphorylation of enzymes involved in ED pathway: glucose dehydrogenase (GDH-1, encoded by *SSO3003*) on Tyr^{134} were observed when grown on both glucose and tryptone condition, while the phosphorylation of GDH-2 (SSO3204) on Tyr^{324}, Ser^{326} and $Thr^{333}$was only determined when tryptone was supplied [7]. However, the phosphoproteomic study was directed to phosphoprotein identification only in *S. solfataricus* [7] and *S. acidocaldarius* [47]. Until now, there is no report concerning quantitative phosphoproteomics in *S. solfataricus* and *S. acidocaldarius*, not even in archaea. The changes at quantitative proteome level involving in metabolic pathways especially regulation of CCM mechanism in phosphoproteome level of *S.*

solfataricus is therefore worth studying. Hopefully, this work will provide new information for extending knowledge on the pathways for regulation of carbohydrate metabolism in archaea.

2.2.1.3 Metabolomic studies of *S. solfataricus* and techniques applied

Compared to the relatively extensive proteomic studies, the investigation of *S. solfataricus* being performed at the metabolomic level is limited. One study carried out by Zaparty *et al* [38], which integrated quantitative metabolome changes and exometabolome analyses of *S. solfataricus* P2 in response to living temperature change from 80 to 70 °C were performed using GC-MS analysis. A total of 70 metabolites and 4 compounds (glycerol, glucose, inositol and erythritol) from exometabolome were detected, but glycerol may be from glycerol stock and glucose was used as the carbon source. Amino acid metabolism was found to be significantly affected, and to result from the change of living temperature by integrating datasets originated from proteomic, transcriptomic and metabolomics analyses [38]. Another metabolomic study of *S. solfataricus* investigated the comparison between strains PBL2025 and PBLΔ3195 (key enzyme 2-keto-3-deoxygluconate kinase deletion mutant using GC-MS technique [48] was carried out. Recently, in order to maximise the metabolome coverage, untargeted metabolome analyse by RP and HILIC fractionations prior to ultra-performance LC-MS (UPLC/MS) were applied in *S. solfataricus* and *Ignicoccus hospitalis* [49]. No reduced form glutathione (GSH) but low concentrations of oxidized form GSH were detected in both archaea [49]. It fills in a gap of the evidence or proof for the existence of GSH in *S. solfataricus*, which is supposed to result from the high sensitivity of the UHPLC-MS. Together with proteome data, it supports the existence of GSH biosynthesis in archaea [49].

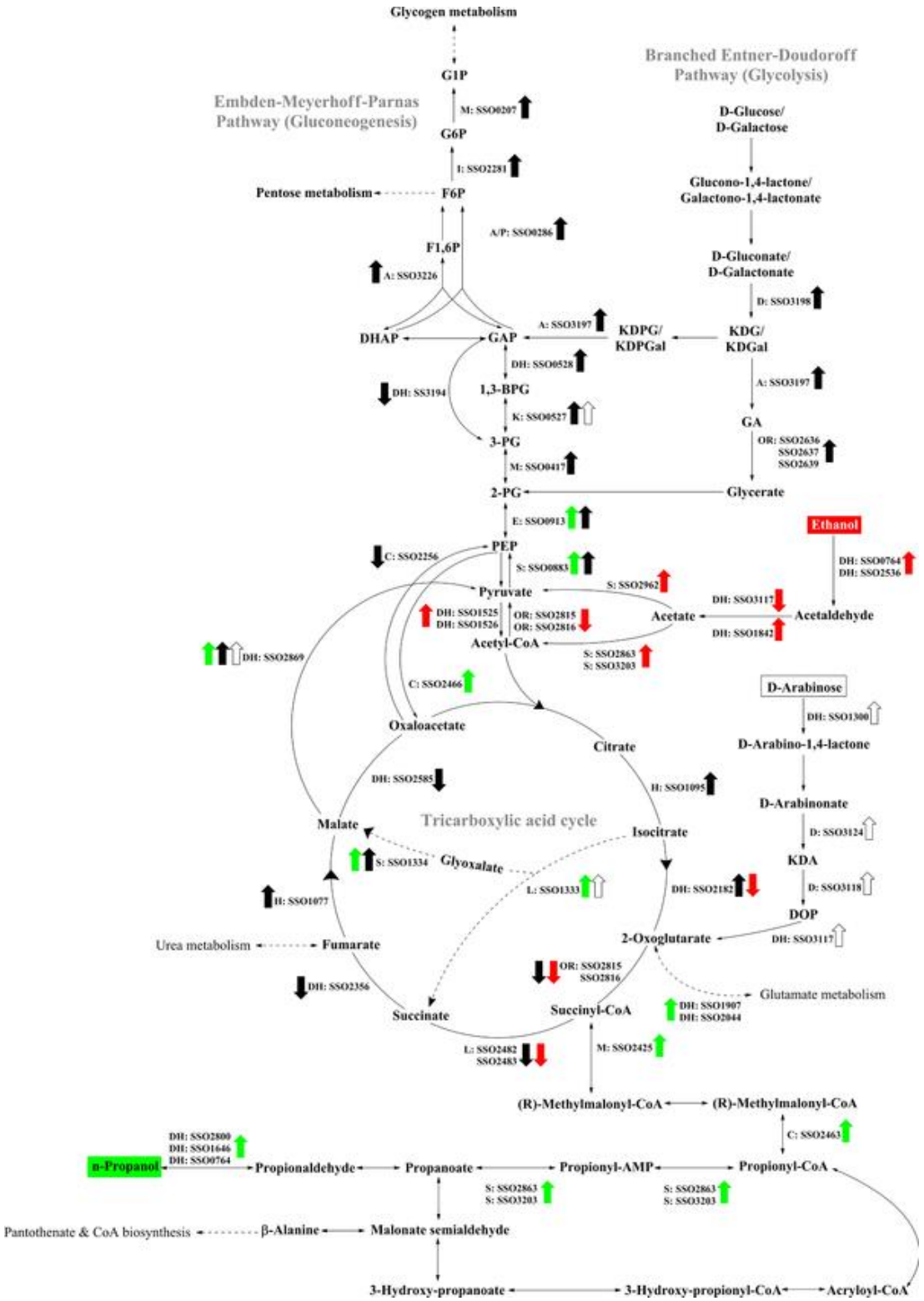

Fig 2.1 Reconstruction of CCM in *S. solfataricus*. [Reproduced from [4]]. The up and downregulated proteins compared to glucose grown cells are indicated by arrows: black, yeast extract and tryptone; white: D-arabinose; red: ethanol; green: *n*-propanoate. D: dehydratase; A: aldolase; DH: dehydrogenase; OR: oxidoreductase; M: mutase; E: enolase; S: synthase; K: kinase; I: isomerase; P: phosphatase; H: hydratase; L: ligase; C: carboxylase; G1P: glucose 1-phosphate; G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; F1,6P: fructose 1,6-bisphosphate; DHAP:

dihydroxyacetone 3-phosphate; GAP: glyceraldehyde 3-phosphate; KD(P)G: 2-keto-3-deoxy-(6-phospho)gluconate; KD(P)Gal: 2-keto-3-deoxy-(6-phospho) galactonate; 1,3-BPG: 1,3-bisphosphoglycerate; 3-PG: 3-phosphoglycerate; 2-PG: 2-phosphoglycerate; GA: glyceraldehyde; PEP: posphoenolpyruvate; KDA: 2-keto-3-deoxy-D-arabinoate; DOP: 2,5-dioxopentanoate.

2.2.2 Technique optimisation for proteomic/phosphoproteomic studies

The successful application of gel- and mass spectrometry (MS)-based proteomic techniques was recently reviewed by taking *S. solfataricus* as an example [4]. The following subsections explained some issues in detail.

2.2.2.1 Protein quantitation

Quantitative proteomics refers to the measurement of absolute or relative protein amounts under different biological conditions. MS-based quantitative proteomic techniques provides quantitative information either at the MS or MS/MS level, which is incorporated based on ion intensity or peak area. There are label-based and label-free quantitative methodologies. Isobaric and isotopic labelling techniques are widely used, such as iTRAQ (isobaric tags for relative & absolute quantitation) [50], TMT (tandem mass tags) [51], metabolic-based SILIC (stable isotope labelling with amino acid in cell culture) [52] and ICAT (isotope-coded affinity tags) [53]. iTRAQ was applied in this thesis since it allows to analyse up to 8 samples (4- or 8-plex iTRAQ) in a single experiment simultaneously.

2.2.2.2 iTRAQ (isobaric tags for relative & absolute quantitation)

4-plex and 8-plex iTRAQ reagents have been widely used in proteome studies of *S. solfataricus* and were reviewed recently [4]. iTRAQ tags are composed of reporter ions (114-117 for 4-plex, plus 113, 118, 119 and 121 for 8-plex), a balance group and primary amine specific peptide reactive group: N-hydroxysuccinimie ester (NHS-ester) as shown in Fig 2.2. Total mass of isobaric tags are the same at 145 Da for 4-plex. Upon MS/MS, reporter ions with various m/z are released; these ion will be used for relative quantification, and other ion information will be used for protein identification. Digested peptides from different biological treatments are labelled separately and pooled together before submitting to HPLC-based fractionation. Different iTRAQ tags are co-eluted from the HPLC due to similar physio-chemical properties. Protein identification and relative quantitation are thus characterised by using LC-MS/MS.

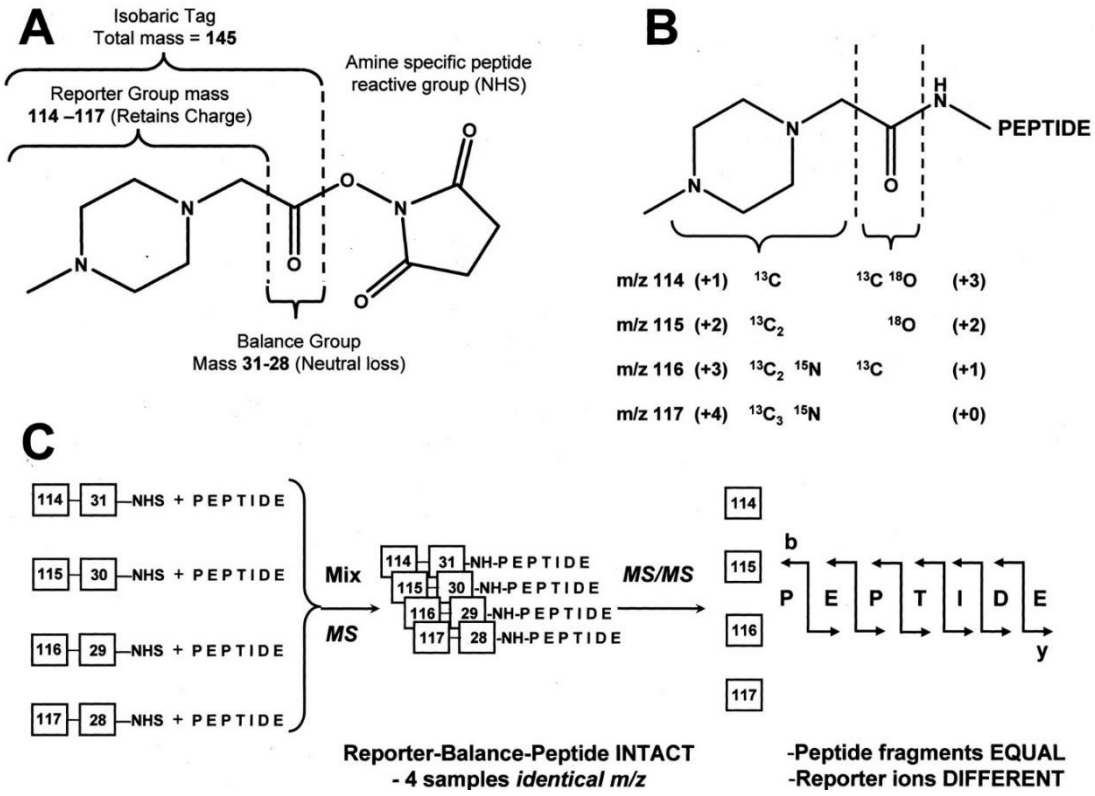

Fig 2.2 A. Diagram shows components of 4-plex iTRAQ tag. B. Combination between isobaric tag and peptide amine terminal. Total mass of reporter ion and balance group are kept constant. C. Fundamental principle explanation of isobaric tags used for identification and relative quantitation upon MS and MS/MS. [Reproduced from [50]]

2.2.2.3 Technology for phosphoprotein characterisation

MS techniques have been developed and are widely used for proteome research in areas such as the systematic analysis of proteins and their post translational modifications (PTMs). However, until now analysis of phosphorylation has always been a challenge for researchers: phosphorylation/de-phosphorylation occurs at sub-stoichiometry levels and phosphoproteins often occur at low abundance in cells as well as being an instantaneous phosphorylation/de-phosphorylation process [54]. In addition, the low sensitivity of phosphopeptides under positive ionization conditions due to the affect from the negative charge carried on phosphate group hampers the development of phosphoproteomic study. Further MS technical issues including the phenomenon of ionization suppression, in which the ionization of low abundance phosphopeptides is greatly suppressed due to the presence of high abundant non-phosphopeptides, the MS dynamic range can also limit the phosphoproteome studies if the dynamic range of protein/peptide concentration is

greater than that of MS [54]. Therefore, the enrichment of low abundance phosphopeptides seems to be vital for the identification and characterisation of phosphoproteins. A large number of reviews have emerged that examine the sample preparation process for phosphoproteomics research of which Kanshin *et al*., 2012 and Loroch *et al*., 2013 are more recent examples [55, 56] which focus on the enrichment of phosphopeptides. Furthermore, a single or combination of various MS dissociation techniques including collision-induced dissociation (CID), electron transfer dissociation (ETD), electron capture dissociation (ECD), generate cleavage at different sites and result in various product ions, benefits for PTMs study, and discussed in detail in Section 2.5.3 [57]. For example, a combination of complementary CID and ETD fragmentation techniques was proven to enhance proteome and phosphoproteome coverage than their sole use [57].

Besides different fragmentation techniques as mentioned above, neutral loss of H_3PO_4 (98, 49 or 32.7 Da, corresponding to +1, +2 and +3 charge state) or HPO_3 (80 Da), SRM (selective reaction monitor) /MRM (multiple reaction monitor) has been proposed to be beneficial for phosphopeptide characterisation [58]. Also, compared to the conventionally used positive ion mode, a strong relative signals of phosphorylated peptides was detected under negative ion mode [59]. However, the beneficial application of neutral loss dependent MS3 strategies for high mass accuracy instrument for global quantitative phosphoproteomics is not significant, due to the balance between time consuming and more information collection from complex samples [60]. However, not many reports were obtained for identification of the low abundant phosphopeptides and to date, no large scale quantitative phosphoproteome study has been performed in archaea. It is still a challenge. Further technical optimisation is unavoidable at this stage.

2.2.2.4 Techniques for phosphoproteomic studies

One of the issues for quantitative phosphoproteomic studies is the low abundance of phosphopeptides, as reviewed in detail recently [58]. It can be solved: take iTRAQ for an example, through the employment of phosphopeptide enrichment techniques, which is incorporated either before or after labelling. Some phosphoproteome studies suggested a use of large protein material (e.g. 10 mg) for phosphopeptide enrichment before labelling with iTRAQ reagents. The affected factors for phosphopeptide enrichment are discussed in details in Section 2. Besides the sample preparation and phosphopeptide enrichment

issues mentioned in Section 2.2.5.1, it is reasonable to notice that iTRAQ reagents are highly specific for primary amines, which require that the sample is free of amine. However, phosphopeptides are usually eluted by pure ammonium solvents in most of the enrichment techniques such as IMAC/MOAC, therefore, cleaning up samples such as using C_{18} material is necessary before moving to the labelling step.

One of factors that need to be considered for the design of phospohproteomic experiment is the available reporter ions upon ETD (electron transfer dissociation) fragmentation. Since upon ETD, only three of 4-plex [61], five of 8-plex iTRAQ labels [62] and four of 6-plex TMT tags [63] give rise to unique reporter ions that would allow for quantification, respectively. In ETD mode, both 113 and 114 tags appear at 101 Da; similarly, 104 Da for 116 and 117 tags, 106 for 118 and 119 tags for 8-plex iTRAQ reagents [61], while with 6-plex TMT, both 126 and 127 tags appear at 114 Da, 128 and 129 tags appear at 116 Da [63].

In summary, workflows for iTRAQ-based quantitative phosphoproteomics can be found in Fig 2.4, in which the labelling step could be performed either directly for peptides or after phosphopeptide enrichment.

2.2.3 Need for -omics studies of *S. solfataricus*

There are more than 40 genes (*SSO3004-SSO3050*) absent in PBL2025 compared to the model strain P2, some of which are involved in sugar metabolism, such as glycosyl hydrolase (*lac* S gene, coding for β-glycosidase) [64, 65]. Genes *SSO3006, SSO3019, SSO3022, SSO3032* and *SSO3036* in P2 were found to be involved in the CCM [66]. Three of glucose-1-dehydrogenases (dhg-1, dhg-2, dhg-3) genes have been reported in P2, among which dhg-2 (coded by gene *SSO3042*) has still an unknown function [66, 67].

In addition, the uptake of sugar substrates, for instance: glucose may be affected in PBL2025 because the transportation of glucose in *S. solfataricus* is via a high-affinity binding-protein dependent ATP-binding cassette (ABC) transporter [68]. A typical ABC transporter is composed by 2 permeases, 2 ATPases, and 1 membrane anchored substrate-binding protein. There are 37 putative ABC transporters and 15 binding proteins in P2 based on genetic information [69]. In contrast, the number of functional ABC transporters may be less in PBL2025 due to the absence genes as shown in Fig 2.3: genes *SSO3012,*

SSO3045 and *SSO3046*, produce ATP binding protein-ABC transporter; gene *SSO3043* codes for binding protein; and genes *SSO3047* and *SSO3048* code for permeases of ABC transporter; cellodextrin transporter component SSO3053 was proposed to be a homolog to an oligosaccharide transporter (TM1223) from *Thermotoga maritime* [70].

Several questions still remain: How does the absence of specific 'missed' genes affect the glucose metabolic pathway especially branched ED pathway (glycolysis), as show in Fig 2.3. Does dhg-2 possess similar function to asdhg-1and dhg-3? Does the lack of encoding genes from three of 37 ABC transporters show an effect on substrate uptake? Compared to P2, are there any other ABC transporters being activated resulting from the lack of some ABC transporter coding genes in strain PBL2025, which will provide carbon and energy source for cell growth? Is it because of the existence of promiscuous pathways or some iso-enzymes substitute for the missing functional genes so that *S. solfataricus* PBL 2025 could survive? Or just because the missed genes solely play a role in surface attachment and biofilm formation [71]. The investigation by comparing proteomic level differences between P2 and PBL 2025 can provide some answer.

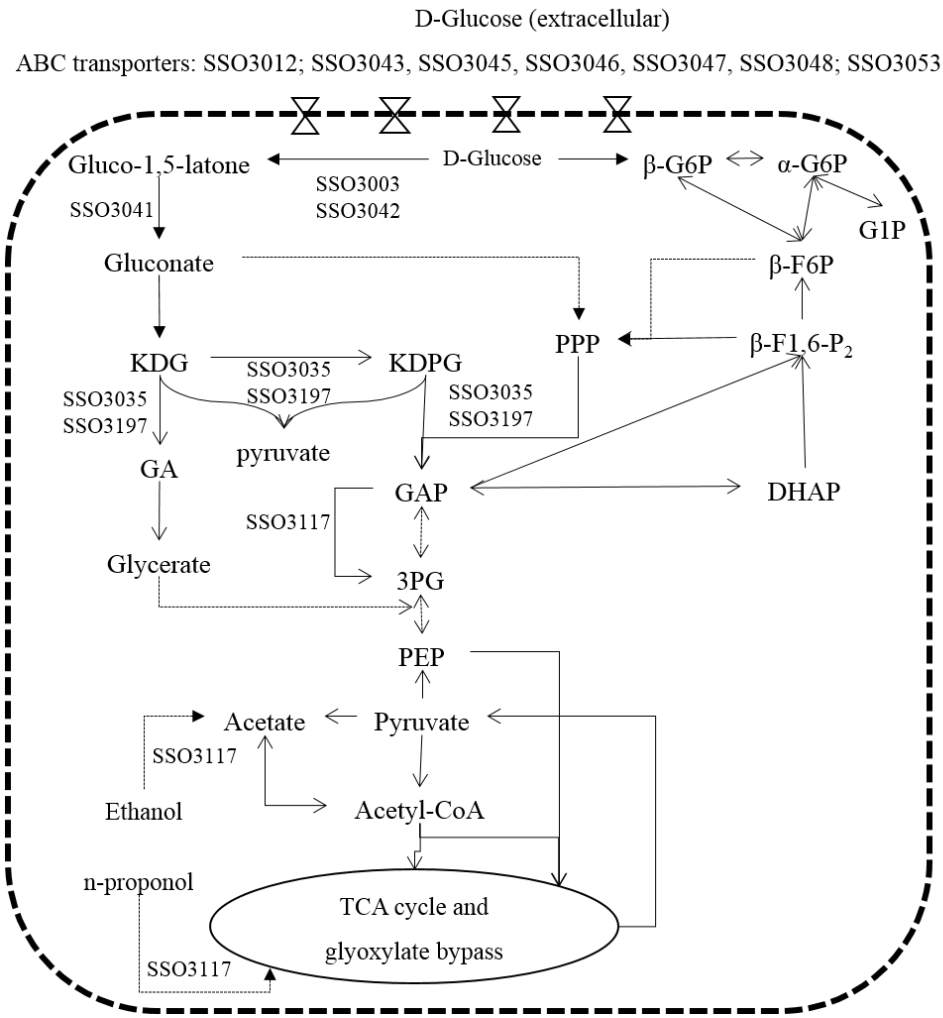

Fig 2.3 A modified branched ED pathway for the absence genes in PBL2025. G1P: glucose 1-phosphate; G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; F1,6P: fructose 1,6-bisphosphate; DHAP: dihydroxyacetone 3-phosphate; GAP: glyceraldehyde 3-phosphate; KD(P)G: 2-keto-3-deoxy-(6-phospho)gluconate; 3-PG: 3-phosphoglycerate; GA: glyceraldehyde; PEP: posphoenolpyruvate.

SSO3117 gene was found to catalyse acetaldehyde to acetate in ethanol catabolism pathway [45] and catalyse the reversible reaction from propionaldehyde to propanoic acid in the n-propanol metabolism pathway [72] in *S. solfataricus* P2, as indicated in Fig 2.3. *SSO3117* gene in P2 was highly expressed at protein level when it was grown in pentose D-arabinose compared to hexose D-glucose. The putative aldehyde dehydrogenase (gene *SSO3117*) has been re-annotated as 2, 5-dioxopentanoate dehydrogenase (DOP DH), also, its function has been characterised in pentose (non-natural isomer, d-arabinose) oxidation pathway, which was found to catalyse 2,5-dioxopentanoate, glyceraldehyde, and

glycolaldehyde [73]. Its gene function has also been confirmed in the natural pentose, D-xylonate and L-arabinonate [74] in *S. solfataricus* P2 and *S.acidocaldarius*.

However, it was revised to be aldehyde dehydrogenase (Lactaldehyde dehydrogenase, 'succinate semialdehyde dehydrogenase or other NAD-dependent aldehyde dehydrogenase') in *S. solfataricus* P2 [66]. Lactaldehyde dehydrogenase is one of the general aldehyde oxidation enzymes, which has been purified both in *Escherichia coli* [75, 76] and yeast [77]. This enzyme was oxygen induced and was plroposed to be involved in the oxidation aldehyde produced from L-1,2-propaediol, fructose and rhamnose etc metabolic pathways in *E.coli*. [75]. However, there is no report about the lactaldehyde dehydrogenase in archaea. The biological function of SSO3117 in *S. solfataricus* was unknown. What is the possible change after the disruption of the lactate dehydrogenase gene in *S. solfataricus* at the protein expression level? The comparisons of proteomics and metabolomics changes in system level between mutant and wild strains would benefit for elucidation of functional role of SSO3117.

Section 2 Phosphoproteomics studies in the archaea

2.2.4 Phosphoproteomic studies

2.2.4.1 What are post translational modifications (PTMs)?

Post translational modifications (PTMs) are the covalent modifications of proteins being followed by the synthesis process, where processed proteins undergo modification by chemical addition, for examples, phosphorylation, glycosylation [78-81], acetylation [82, 83], proteolytic cleavage [84, 85], methylation, etc. Protein properties and activities might change after these modifications. Each modification performs a distinguished function throughout the cell's life, e.g. phosphorylation plays an important role in signal transduction [9], protein function, activity and interaction as well as intracellular protein localization [86]. PTMs studies become fascinating when the cross talk between phosphorylation and other PTMs such as ubiquitination [87] and acetylation [88] were reported.

2.2.4.2 Bibliometric data analysis

To get a picture of recent studies of PTMs (last ten years), different PTMs keywords such as phosphorylation, methylation, glycosylation, acetylation and proteolysis as well as

human, archaea and *S. solfataricus* were used to search within the National Center for Biotechnology NCBI Pubmed (http://www.ncbi.nlm.nih.gov/pubmed/) database. The results are presented in Fig 2.4. From Fig 2.4A, we can see that study of phosphorylation is always one of the so called 'hot' research topics. However, few studies have been reported in archaea compared to the wealth of research numbers reported for study of phosphorylation in human material (Fig 2.4B).

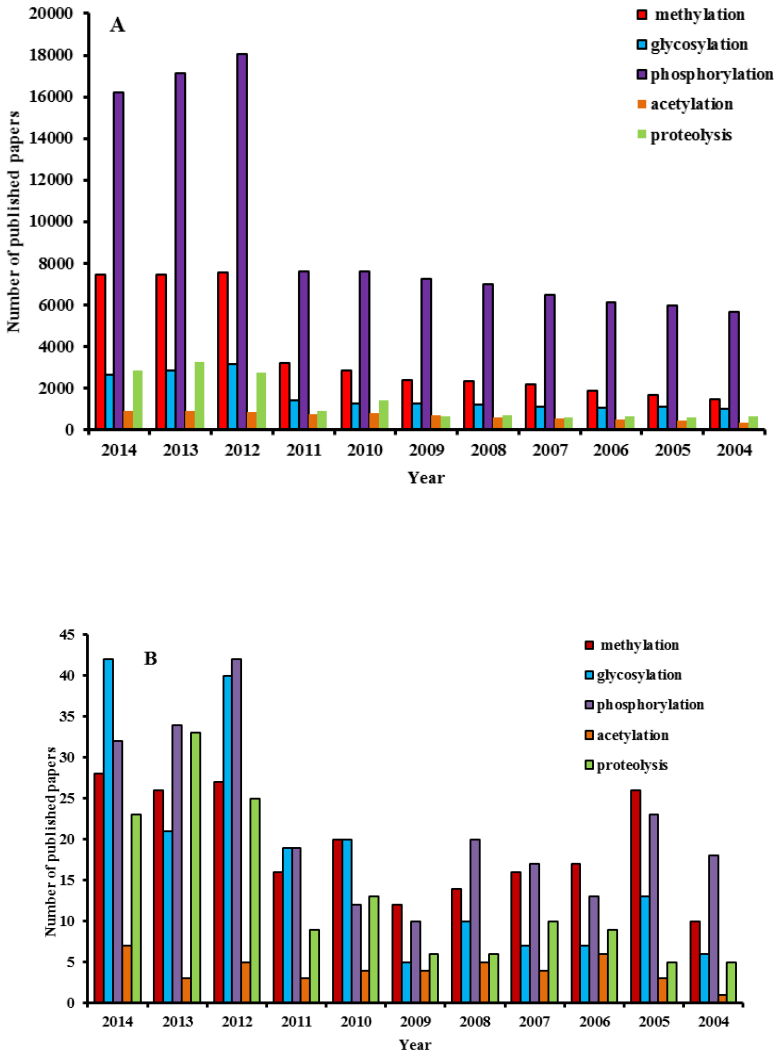

Fig 2.4 Numbers of post translational modifications studies reported for (A) humans and (B) archaea from 2004 until 2014.

It can be seen that although the number of published papers in archaea is lower than that observed for humans, the broad trend is similar. The study of *S. solfataricus and S. acidocaldarius* is still mainly focused on the identification of phosphoproteins. There is a great depth of knowledge to be acquired through the phosphoproteomic

study. In 2000, the identification and characterisation of enzymes mainly depended on PCR and other biological tools [89-91]. With the development of new techniques, the application of MS for protein identification and PTMs research, especially for phosphoproteomic studies attracted more attention [54, 92, 93].

2.2.4.3 What is protein phosphorylation?

Phosphorylation of proteins is one of the most important PTMs in prokaryotes [94, 95], eukaryotes [9] and archaea [96]. In this process, the γ-phosphate group transfers from ATP or GTP to the side chain of amino acids in a (polypeptide chain) protein [97, 98]. Phosphorylation has been found on 9 amino acids. It occurs not only on Ser, Thr, Tyr, Asp, His [99], but also on Cys, Lys, Gln [100] and Arg residues [101]. The activities and functional properties of protein, especially enzymes can be changed after phosphorylation. For example, phosphorylation on Ser^{309} of phosphopeptide [VGpSVDIAHK] from protein (encoded by gene *SSO0207*) has been shown to attenuate the catalytic efficiency of phosphohexomutase in *S. solfataricus* P1 [102]. This instantaneous and reversible process is well known to play an important role in signal transduction in almost all cellular processes, including cell division and metabolism [9], to cope with stress adaptation, eg. temperature, pH, nutrition deficiency [95, 103, 104], DNA damage response in chromosomes [105], and control of cell growth [106]. In *S. tokodaii*, phosphorylation of the fork head domain protein (ST0829) negatively regulated its DNA-binding activity, and it was postulated that this could be involved in the adaption of this organism to extreme living environments (high temperature, low acidic pH) [107].

2.2.4.3.1 Ser/Thr/Tyr protein phosphorylation

Sporadic studies on phosphorylated proteins have been done in archaea and were reviewed [96, 108]. Updated phosphoprotein identification information was collected and is listed in Supplementary Table SP2.1.

Although the phosphoserine was detected in egg yolk protein in 1932 [109]. It was over two decades later in 1954 that the first enzymatic phosphorylation of proteins was described and phosphothreonine was identified in casein [110]. However, phosphotyrosine was still unknown until 1979 [111]. Ser/Thr/Tyr phosphorylation has become known as one of the traditional PTMs in eukaryotic microorganisms [112, 113]. Also, Ser/Thr/Tyr phosphorylation has been found in bacteria such as *B. subtilis* [114]

and *E. coli* [115]. Interestingly, researchers noted that Ser/Thr/Tyr phosphorylation in some bacteria was linked to bacterial pathogenicity [116]. In the minimal organism *Mycoplasma pneumonia*, Schmidl [117] found a highly conserved phosphorylated Ser site from archaea to human for the first time, and suggest the ecological niche adaptation role of protein phosphorylation. The number of identified phosphoproteins, phosphopeptides and the assignment of phosphosites are increasing. The availability of genomic information and increasing maturity of high-accuracy MS techniques such as CID, ETD and ECD lead to a rapid and sensitive study of phosphoproteomics [57]. Furthermore, researchers are making progress to work on the dynamics of *in vivo* phosphoprotein and to reconstruct the systemic signalling networks [57, 118].

However, developments in phosphoproteomic studies in archaea is not as well investigated as those in eukaryotes and bacteria. The first report on phosphoproteins in the extreme halophile archaeon *Halobacterium halobium* was observed in 1980 through using (^{32}P) orthophosphate pulse-labeling by Spudich [119]. After that, phosphotyrosine proteins were detected in *S. sulfataricus* P1, *Haloferax volcanii* and *Methanosarcina thermophia* using antiphosphotyrosine antibody [120]. Later, phosphorylation on Ser/Thr/Tyr of protein kinases was reported using the traditional ^{32}P incorporation method with the aid of MS analysis, as listed in Supplementary Table SP2.1. Identification number of phosphoprotein in archaea increased significantly by incorporating MS based techniques. The first comprehensive Ser/Thy/Tyr phosphoproteome study was presented by Aivaliotis *et al*., [121] in the halophilic archaeon *H. salinarum* by employing SCX fractionation followed by TiO_2 beads enrichment, from which 90 unique phosphopeptides corresponding to 69 phosphoproteins were determined by LC-MS/MS. Later, a higher number of phosphoproteins were determined in model *Sulfolobus* species: 540 and 801 from *S. solfataricus* [7] and *S. acidocaldarius* [47] respectively by using LC-MS/MS with multiple injections and MRM and data analysis by a precursor acquisition independent from ion count method without incorporating any phosphopeptide enrichment technique.

2.2.4.3.2 Phosphorylation of His-Asp signal transduction system

The availability of genomic sequence information has ensured the development of protein phosphorylation research in all three domains of life [99, 101, 122]. Phosphoryl groups can be added on the 1-N or 3-N of iminazole ring in histidine [123]. The first instance of

phosphohistidine was reported by Boyer [124]. Phosphorylation sites in bacteria have been initially reported in His and Asp residues [94, 95, 125]. Gradually, the phosphorylation of His-Asp signal transduction system was discovered in yeast, fungi [106], *Arabidopsis* and rice [126, 127]. Kanamaru [106] found that this famous signal transduction system played an important role in controlling cell growth. In archaea, Rudolph detected phosphohistidine in kinase ChA [128] and Asp in ChY [129]. However, this well characterised two-component regulatory system (TCS) to date has not been identified in animals and is not encoded in the human genome [130]. With the discovery of the TCS in some eukaryotes, vertebrate cells and histones [105, 131-136], research on acid-labile phosphohistidine has attracted more attention. The phosphorylation of hydroxyl oxygen on Ser/Thr/Tyr residues (also referred to as O-phosphorylation) forms a phosphate ester, which has been shown to be acid-stable [134]. The phosphor-amidate bond formed during the phosphorylation of nitrogen on His, Arg and Lys is acid unstable [137]. Besant and Attwood [123] pointed out that acid/alkali-stability was a characteristic that is distinct from other phosphorylated residues. However, conventional MS techniques for Ser/Thr/Tyr detection are limited for the study of phosphohistidine due to the P-N bond being unstable under acidic conditions [134]. This potentially is one of the reasons why phosphohistidine has not been identified in most MS experiments.

2.2.4.4 Phosphoproteomic studies in archaea

Protein phosphorylation in archaea has been noticed to play important roles. Skórko observed that phosphorylation in *S. acidocaldarius* depends on the state of growth and the extent of phosphorylation reached its maximum in the late exponential phase [138]. Previous studies have demonstrated that proteasomal phosphorylation is regulated by growth in halophilic archaea [139]. Aivaliotis and co-workers found that phosphoproteins in *H. salinarum* mostly appeared in cell metabolism [121]. Humbard reported that phosphorylation of the α1 subunit of 20S core particles (components of proteasomes) in *Haloferax volcanii* was related to late-stationary phase growth [140]. Moll and Schäfer have proven that the plasma membrane of *S. acidocaldarius* can withstand low pH 2-3 while maintain internal cellular pH around 6.5 [141]. Van de Vossenberg, *et al.* demonstrated that the ion permeability of cytoplasmic membrane regulates the limit growth temperature in *S. solfataricus* [142]. Konings, *et al.*, reviewed the critical role of

cell membrane of archaea in surviving of extreme living environments [143]. Recently, the total phosphoprotein identification number in archaea increased significantly with the employment of MS techniques, the study of archaea using proteomics tool has gained some attention since its establishment for the third domain of life [4]. However, no quantitative phosphoproteome research has been carried out in archaea.

2.2.5 Technology for phosphoproteomic studies

Current phosphoproteomic studies (as shown in Fig 2.5) mostly employ enrichment techniques such as IMAC, MOAC in combine with LC-based fractionation to gain a better phosphoproteome coverage, as detailed in the following parts: 2.2.5-2.2.6.

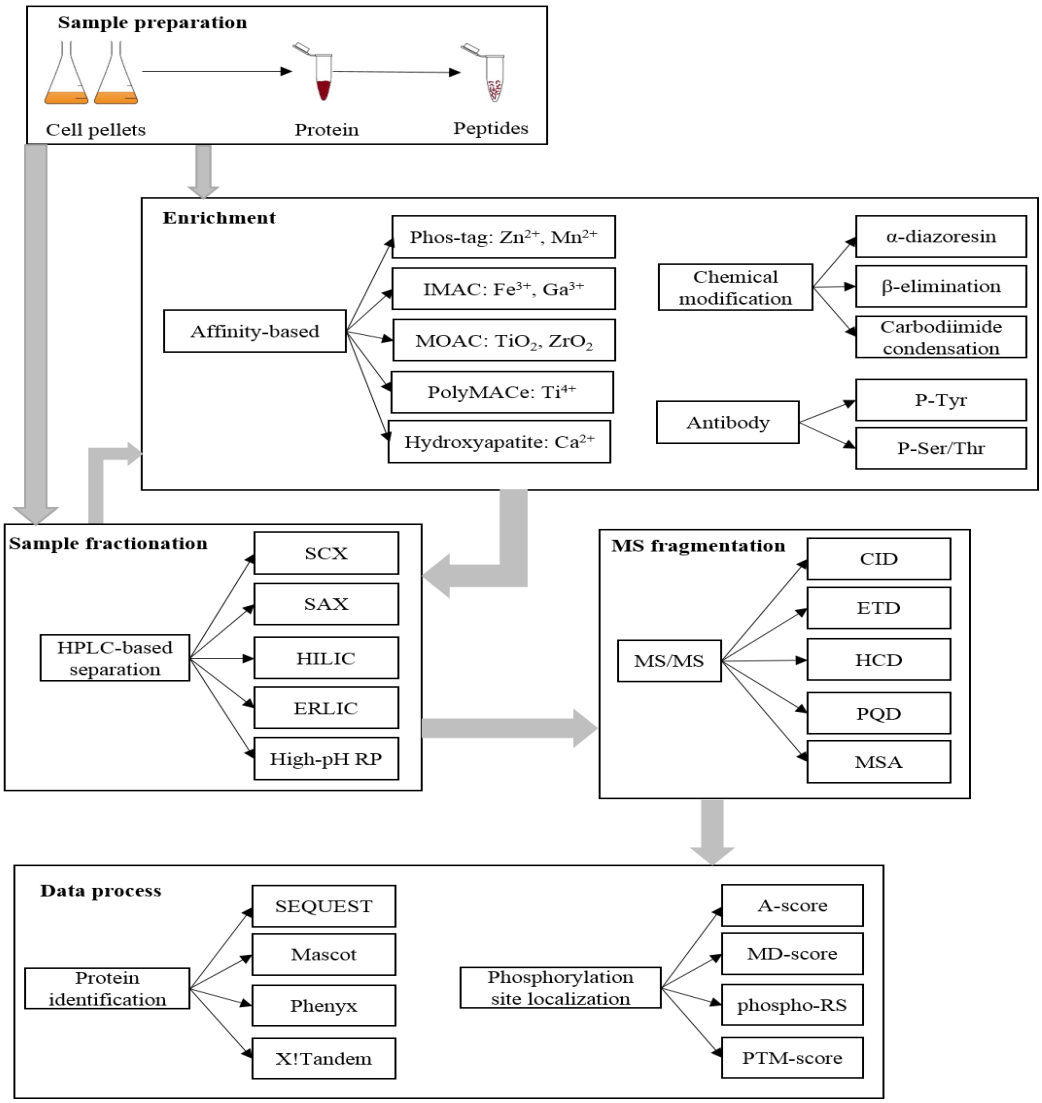

Fig 2.5 Phosphoproteomic strategy. [Figure modified based on [56, 144, 145]]. IMAC: immobilized metal ion affinity chromatography; MOAC: metal oxide affinity chromatography; SAX: strong anion exchange; SCX: strong cation exchange; HILIC: hydrophilic-interaction

liquid chromatography; ERLIC: electrostatic repulsion hydrophilic interaction chromatography. CID: collision-induced dissociation; ETD: electron transfer dissociation; HCD: high-energy C-trap dissociation; PQD: pulsed Q collision induced dissociation; MSA: multi-stage aciviation.

2.2.5.1 Sample preparation for phosphoproteomic studies

2.2.5.1.1 Matrix effects and limit of detection

As with many techniques, it is important to have standards from which to estimate the starting materials need to be used. In the case of phosphorylation, bovine α, β-casein can be used as standards for limit of detection of phosphopeptides determination and therefore the starting material information for phosphoproteome optimisation experiment can be calculated.

In LC-MS-based phosphoproteomics experiments, reproducibility and accuracy of analysis results can be affected by impurities. This has been termed the 'matrix effects (ME)' [146]. Minimisation of the ME would greatly improve the quantitative results. One of the key modern MS soft ionization techniques, electrospray ionization (ESI), is prone to be negatively affected by ME. Trufelli, *et al*. [146] have reviewed the potential mechanism of ion suppression/enhancement that results from ME interference. In addition, they made some useful comments on how to overcome ME in different aspects, sample preparation, chromatography, calibration, and MS, etc.. Some suggestions include the use of hydrophilic-interaction liquid chromatography (HILIC) (for high polar compounds analysis) or ultra-HPLC as alternatives to traditional reverse phase LC-MS, the modification of MS conditions, eg. ionization mode/technique, and discussed by Liu *et al*. [147].

2.2.5.1.2 Improvements for efficient digestion

While high-quality samples are the foundation for biological analyses, sample preparation methods are crucial steps for MS-based proteomics study, which will affect the subsequent data quality.

Enzymatic and chemical cleavage techniques, and developments focus on enhancing digestion efficiency and accelerating the laborious digestion processes, have been reviewed [148]. Different size peptides are generated using different enzymes. Small

phosphopeptide (residues 33-48, with phosphorylation on Ser 35) from 100 fmol of β-casein was detectable using Lys-C digestion, but no phosphopeptide was detected from trypsin digestion products [149]. For phosphoprotein digestion, enhanced digestion efficiency can be achieved through mild digestion or multiple enzymatic treatment or multidigestion by protease-immobilized microreactors [148]. Incomplete/mild digestion using trypsin or endoproteinase Lys-C was reported to aid the determination of multiply-phosphorylation sites from 20 recombinant *E. coli* proteins using MALDI-MS. It is inferred to be caused by ionization efficiency enhancement, resulting from positive charge of residues especially those contained Arg [148].

Compared to the single use of trypsin, a good digestion can also be achieved by using multiple digestion protocols, which incorporates a combination of trypsin and another non- or less specific proteases such as elastase, Lys-C, Glu-C or Lys-N. Compared to the single use of trypsin, the detection of unique phosphopeptides and phosphorylation sites of HeLa cells increased significantly from 3891 to 8062 and 4647 to 8507, respectively, by using firstly Glu-C and then trypsin [150]. However, in archaea, the effects on digestion efficiency by employing multiple digestion have been evaluated for membrane protein analysis but not yet in phosphoproteome study. For instance, protein quantifiable range was extended by incorporating both trypsin and Glu-C in *Halobacterium salinarum* [151]. A high number of unique peptides and proteins (75% more for total proteins) were determined resulted from a better digestion by a combine using of trypsin and chymotrypsin in *S. solfataricus* [152].

In addition, attention needs to be paid to the missed cleavage site resulting from negative interference of digestion for data analysis in a phosphoproteome study. Trypsin digestion efficiency could be negatively affected by amino acid residues, for example, if Pro is closely proximal to tryptic cleavage site; basic amino acids or negatively charged residues adjacent to phosphosites [145], and phosphorylation-modification [153]. Phosphorylation amino acids type, location and sequence information as well as the salt bridge formed between Arg and Lys has a negative effect on trypsin digestion [154]. Dickhut *et al.*, [154] observed a strong negative effect from phospho-Ser and phospho-Thr and a smaller inhibitory effect from phospho-Tyr, which is proximal to cleavage sites using 19 artificial miss-cleavage model peptides.

Impairment tryptic digestion resulted from phosphorylation-modification can be overcome to some extent by employing an increase of trypsin/peptides ratio from the commonly used 1/50 to 1/20 or 1/10 and addition of organic solvents such as ACN and trifluoroethanol (TFE) [154]. The positive improvements by using trypsin/peptides ratio 1/20 or 1/10 and the addition of organic solvents including 10% ACN or 5% TFE for phosphoproteome study were validated in large-scale quantitative phosphoproteome experiments using both label-free and iTRAQ label-based studies of human platelets samples [154]. Recently, a trypsin/peptides ratio of 1/1 was used by Liu *et al*., [155]. In this 'one-step protocol', interference from excessive trypsin and its autolysis product was removed since Ti^{4+}-IMAC enrichment was applied directly after digestion [155]. Although phosphorylation sites determination and quantification was significantly observed by the addition of TFE, considerations need to be taken for its use in high-complex biological samples due to its interference with the following TiO_2 enrichment and a reduced digestion efficiency for certain model peptide with the presence of TEAB [154].

In-solution digestion seems to be simpler than the others in terms of sample handling but has similar time-consuming problem due to overnight digestion, which is commonly used for trypsin digestion. Steps are moving forward from conventionally > 25h protocol to > 16h protocol, and recently, a more efficient 25 min digestion procedure was optimised and tested by using HeLa cells [155]. The 'one-step protocol', refers to the combination of cell lysis, protein extraction and digestion into one step, and ~25 min was consumed but with a compromise of a high amount of trypsin [155]. Cell pellets were re-suspended in lysis buffer together with a high amount of trypsin, cell disruption and protein extraction was performed using ultrasonication at 37 °C and trypsin digestion were simultaneously achieved [155]. In contrast to the 1:25 peptide to trypsin ratio, which was used in conventional > 28 h (Protocol A) and improved >16 h (Protocol B) procedures, the optimised ratio was 1:1 for 'one-step protocol'. Furthermore, no protease inhibitor was applied due to short time treatment. Interference from excessive trypsin and its autolysis product will be removed since Ti^{4+}-IMAC enrichment was applied directly after digestion [155]. The identification number of unique phosphorylation sites from new protocol, protocol A and B were 2987, 2682 and 2020 from HeLa cells (10^5 cells, around 40 µg protein) and 1452, 1167 and 965 from HepG-2 cells, respectively. Phosphosites quantification number is 1465, 1622 and 965 from similar amount of HeLa cells

determined by using heavy and light isotope dimethyl labels [155]. However, the limit factor to routine usage of this technique is the expensive cost of huge amount of trypsin especially for experiments requiring a large amounts of starting materials. Additional considerations need to be taken if followed by label-based quantitative phosphoproteome experiments such as TMT and iTRAQ, because the estimation of protein concentration seems impracticable.

Besides improving tryspin digestion efficiency, efforts have been put on reducing sample preparation time. Good trypsin digestion efficiency can be achieved under optimal pH and temperature conditions. Combination of microwave irradiation and acid hydrolysis, termed microwave-assisted acid hydrolysis (MAAH) technique, was reported to aid for protein sequence and phosphophorylation sites determination of standard proteins [148]. Tryptic digestion efficiency was improved without affecting PTMs, eg. phosphorylation sites remain intact and it shows promise for phosphoprotein characterisation study. However, MAAH assisted technique was only tested for phospophorylation study in standard proteins, and its application in archaea membrane proteome was reported for *Pyrococcus furiosus*, but not in a phospohproteome study [156].

2.2.5.1.3 In-solution / in-gel digestion and on-membrane digestion

Compatibility with the following enzymatic digestion or phosphopeptide enrichment need to be considered for careful selection of extraction solvents. As many detergents and surfactants may have positive or negative effect on enzyme activity or phosphopeptide enrichment. One example is the chemical denaturation reagent sodium dodecyl sulphate (SDS), the presence of SDS was proven to prevent phosphopeptides loss due to their binding to plastic surfaces, such as pipette tips and Eppendorf tubes prior to IMAC or TiO_2 enrichment [157]. However, the incompatibility of SDS with subsequent LC-MS/MS needs to be taken into consideration, specifically identification of strategies to remove SDS. Worthington *et al*., [158] proposed to avoid the usage of urea for protein denature, because the induced carbamylation modification may contribute greatly to the non-specific binding with IMAC or TiO_2 metal resins, as discussed in Section 2.6.1.

One of the solutions for efficient digestion is to immobilise proteins in PAGE gel or on membrane. It benefits for the separation of protein isoforms and PTMs and therefore increase the possibility of determination by downstream LC-MS/MS [159]. In-gel protein

digestion are widely used in archaea research, while in-column or on membrane (nitrocellulose or poly (vinyldifluoride) PVDF), filter-assisted sample preparation (FASP) digestion has not been applied yet. In-gel digestion is of high efficiency, but often laborious and time-consuming [160] for MS. Compared to in-solution digestion, there is a high-risk of losing 50-85% amount of protein samples by using in-gel digestion [161], due to the complex washing procedure (including stain, de-stain and in-gel digestion).

Benefits of employing on-membrane digestion of gel-resolved proteins includes efficient tryptic digestion, time saving (a half time) compared to in-gel digestion as well as a better proteome coverage especially suitable for study of membrane proteome [162]. Peptides from on-membrane digestion can be analysed directly by MALDI-based MS. Improvements (remove of membrane) made by [162] enabled the possibility that peptides from on-membrane digestion can be detected by (LC)-ESI-based MS. Lee *et al*. [163] tried the combination of polyurethane membrane digestion with in-solution digestion to detect tetra-phosphorylated peptides in β-casein.

In summary, sample preparation is of importance. Efficient digestion can be achieved by a selection of most appropriate extraction buffer and enzyme as well as optimised conditions.

2.2.6 Enrichment technology for phosphoproteomic studies

2.2.6.1 Phosphopeptide enrichment

Although protein phosphorylation plays an important role in almost all of the cellular processes detail can be found in Section 2.2.4, its study is still a challenge. The enrichment of low abundance phosphoproteins or phosphopeptides has been reported to be vital for the phosphoproteomic studies. Technical development enables the efficient quantitative phosphoproteomic studies. A large number of reviews have emerged and have been examined the sample preparation process for phosphoproteomics research with a focus on the improvements on phosphoprotein/phosphopeptides enrichment techniques as well as challenges for their application to large scale phosphoproteomics, as listed in Fig 2.5 [55, 144, 164]. Usually, technologies for phosphoprotein studies including P^{32} radio-labelling and 1-D or 2-D gel electrophoresis, chemical modification and antibody based immunoprecipitation, affinity chromatography, etc. The advantages and disadvantages of various methods have been reviewed in detail elsewhere [144, 145].

Phosphopeptide enrichment can be achieved through chemical modification, which include carbodiimide condensation, β-elimination, oxidation-reduction condensation, α-Diazo resin and so forth. Compared to IMAC/MOAC, the advantage of using chemical modification is the high specificity [144]. For instance, β-elimination is one of the conventional enriched techniques, in which process phosphate group of phospho -Ser or -Thr is eliminated through the addition of basic solvents e.g. $Ba(OH)_2$. This process was coupled with the addition of propanedithiol (named Michael addition), through which specific binding will be achieved using dithiopyridine resin. Thus, peptide with phosphorylation modification can be enriched through these specific bindings [144]. The benefits of employing phosphate elimination for MS analysis include avoiding neutral loss and thus retaining intact peptide sequence, which is useful for further identification and also, increasing of ionization efficiency for common used MS positive ion mode [145]. One of the obvious issues raised by the use of chemical reactions is the sample loss due to many reaction steps and increase of sample complexity resulted from incomplete reaction and side reactions [144]. Commercially available anti-tyrosine antibodies have high affinity for phospho tyrosine and they are suitable for immunoprecipitation, but the specificity of commercially available phospho- Ser and Thr antibodies currently limit their application for phosphoproteome [145]. Affinity-based enrichment techniques such as immobilized metal ion affinity chromatography (IMAC), Phos-tag chromatography (phosphate-binding tag molecular), metal oxide affinity chromatography (MOAC), polymer-based metal ion affinity capture (PolyMAC) enrichment are reported, as discussed by [144]. Phos-tag was similar to IMAC in that it is based on the binding between negative charge oxygen atom of phosphate and positive charge of Mn^{2+} or Zn^{2+}, which was anchored to agarose such as SDS-PAGE as a matrix and therefore act as Phos-tag molecular bound [165]. The major difference between Phos-tag and IMAC is the neutral working conditions around pH 7.5 for Phos-tag compared to low pH incubations for IMAC. Only few phospho-protein/peptide enrichment studies have been carried out to improve enrichment efficiency using Phos-tag in respect of testing different incubation and elution buffers. There are also a few reports on the use of using hydroxyapatite chromatography for phospho-protein or peptide enrichment [144]. Although these studies using Phos-tag, PolyMAC and hydroxyapatite techniques for phosphopeptide enrichment seems promising, they have not been widely tested in complex samples, from which no

solid conclusions are withdrawn in aspects of their superiorities to IMAC and MOAC [144].

A brief summary of various improvements on phosphopeptide enrichment efficiency by using IMAC-Fe and MOAC-TiO_2 is shown in Table 2.1. Furthermore, the observation of novel phosphorylation sites or phosphopeptides and moreover, phosphoproteome coverage seems to be significantly improved by a combination of different methodologies.

Table 2.1. Summary of improvements on phsphopeptide enrichment using IMAC-Fe and MOAC-TiO_2

	IMAC-Fe	**MOAC-TiO_2**
Loading buffer to reduce non-specific binding	High organic solvents (ACN) and acids to reduce binding of acidic peptides	
	Acetic acid > TFA, hydrochloric acid, formic acid	TFA together with glycolic acid, lactic acid, citric acid, 2,5-DHB, phthalic acid, glycerol
Consecutive incubations	Tandem IMAC	Multiple incubation cycles
Other factors	Matrix support: NTA or IDA resins.	Peptide/beads ratio; Incubation time; Physical properties, higher surface/volume ratio: mesopoous > smooth particles; nanoparticles > micro-sized beads.
Formats	Matrix-assisted laser desorption ionization (MALDE) plate, column,tips, gel	Magnetic beads, pipette tips, spin tips, columns
Elution buffer	Ammonium hydroxide	
	Phosphoric acid, addition of DHB or ACN	pH, amine nucleophilicity and structure of elution solvents; bis-Tris propane; pH step wise elution and pH step gradient elution

Note: NTA: nitriloacetic acid; IDA: iminodiacetic acid; 2,5-DHB: 2,5-dihydroxybenzoic acid; TFA:trifluoroacetic acid.

Until now, Fe^{3+} or Ga^{3+} and TiO_2 are among the most widely used materials for IMAC and MOAC, respectively. Principles for IMAC and MOAC are similar, negatively charged phosphates that is oxygen atom will combine with the positively charged metal ions, for instance, Fe^{3+} of IMAC or Ti^{4+} of TiO_2 through incubation. Also, similar procedures were incorporated by IMAC and MOAC. A wash step was carried out after incubation to remove non-specific binding. Finally, phosphopeptides were eluted by basic solvent, such as ammonium solution. A notorious issues for both IMAC and MOAC is non-specific binding of acidic peptides (Asp and Glu), which is caused by the affinity

between the carboxylate group of amino acid residues and metal ions materials. Thus the co-elution of acidic peptides and phosphopeptides are unavoidable, which significantly affect the enrichment efficiency. IMAC was also observed to have a higher affinity towards multiphosphorylated peptides, while the determination of monophosphopeptides from TiO_2 enrichment is obvious [166]. Therefore, improvements on phosphopeptides recovery and in-depth phosphoproteomic study is still on progress in terms of enhancing selectivity, specificity and sensitivity. For IMAC and MOAC, developments focus on testing different metal ions/oxide materials due to different binding efficiency and support resins such as magnetic beads, MALDI plate, column, tips, Poly-MAC as well as optimisation of reaction conditions [144, 164].

2.2.6.1.1 IMAC

As mentioned above, IMAC is based on the affinity reaction between the negatively charged oxygen of phosphates and positively charged metal ions. Fe^{3+}, Ga^{3+}, Zr^{4+}, Al^{3+}, Cu^{2+}, Co^{2+}, Ca^{2+}, Zn^{2+}, Ni^{2+} and Ti^{4+} have been introduced, and Fe^{3+} is mostly used [144, 145, 167]. However, their application has been hampered. Besides the above mentioned non-specific binding issues caused by acidic peptides, one of the disadvantage of IMAC is sample loss and possible contamination caused by Fe^{3+} leaching due to non-covalently binding with solid phase as discussed in [167]. Negroni *et al*., [168] also pointed out another limitation of 'rusty IMAC' caused by re-using POROS-Fe^{3+}IMAC packed columns, which is due to the accumulation of Fe^{3+}precipitation (Fe^{3+} and Fe_2O_3) after ammonium elution. Numerous efforts have been performed to improve the specificity and sensitivity of IMAC from every aspect including testing different support resins, various metal ions and optimisation of operation protocols.

For nonspecific binding from acidic peptide issues, the employment of other enzymes with one carboxyl terminus cleavage such as Glu-C and the esterification of carboxyl to methyl ester were reported [144, 145]. However, wide usage is limited by loss of sample and increase of sample complexity resulting from side reactions and incomplete reaction.

It was noticed that nitrilotriacetic acid (NTA) resins gave better results when combined with Fe^{3+} than Ga^{3+}, while iminodiacetic acid (IDA) resins worked better for Ga^{3+} than Fe^{3+} [145]. But still, no confident conclusions were drawn due to limited reporting.

The influence of incubation buffer on IMAC binding specificity as well as elution buffer are not ignorable [145]. It was observed that acid strength (optimal pH) and ironic composition and strength aids in reducing non-specific binding from acidic peptides with IMAC [144]. Moreover, an optimal concentration of acid and ions need to be determined, since a superoptimal concentration also prevents phosphopeptide binding and sub-optimal concentration will not reduce the non-specific binding of acidic peptide [144]. Compared to TFA, hydrochloric acid and formic acid, an optimal concentration of acetic acid was mostly used in IMAC loading buffer, because strong acids (such as TFA) may lead to phosphopeptide loss [144]. In addition, the presence of ACN in loading solutions shows superior positive influence on blocking non-specific hydrophobic interactions compared to acetone, ethanol and methanol [144]. Ye *et al.* noticed an affinity enhancement between IMAC resin and phosphopeptides when a higher concentration of ACN was applied to treat a tryptic digestion of α- and β- casein. The authors also observed a negative effect of ACN on the ionization of acidic peptides, but not for phospohpeptides [169].

As mentioned before, the ideal elution buffer is also to elute the bound phosphopeptides/proteins from IMAC resins. Ideal elution buffers need to have a good elution efficiency (ideally 100%) and be compatible with subsequent MS. IMAC-bound phospohpeptides can be eluted by phosphoric acid and ammonium hydroxide (pH 10.5); DHB or ACN has been introduced with phosphoric acid as to improve phosphopeptide elution [144]. A mixture of phosphoric acid and DHB is good for MALDI-MS while phosphoric acid with ACN is compatible with ESI-MS [144]. Stepwise elution may be carried out when one cycle elution is insufficient and is discussed in SIMAC part. The ideal protocol for Fe^{3+}-NTA-silica IMAC resin is incubate 1 h and wash with the same 60% ACN in 100 mM acetic acid and finally elute by 5% ammonium hydroxide [169]. An optimal loading and washing solvents as well as efficient elute solution need to be determined for small or large starting materials.

Furthermore, improvements on the application of IMAC to phosphoprotein/phosphopeptides enrichment are still in progress from different aspects, which include the testing of different materials for efficient binding and various support resins as well as the optimisation of enrichment protocols should reduce non-specific binding especially of acidic peptides. Efficient elution of phosphopeptides for a high

phosphoproteomics recovery is also required [164]. Furthermore, the application of phosphopeptide precipitation by $CaCl_2$ coupled with Fe-IMAC, phosphotyrosine immunoprecipitation coupled with methyl esterification and Fe-IMAC, β-elimination with Fe-IMAC, tandem IMAC (IMAC-IMAC), combining of different IMAC metals, have been proven to be superior to the one step IMAC enrichment [55, 144, 164]. Multiple TiO_2 incubations or a combined use of IMAC and TiO_2 (termed SIMAC) will discussed in following part.

2.2.6.1.2 MOAC

Phosphopeptide enrichment using metal oxide such as TiO_2, zirconium dioxide (ZrO_2), aluminium oxide (Al_2O_3), niobium oxide (Nb_2O_5) and aluminium hydroxide was discussed in detail by [144, 145, 167]. A number of different configurations exist for operating this technique. One example was that ZrO_2 packaged tips were shown to be superior to TiO_2 for monophosphopeptide enrichment, while TiO_2 was better for the enrichment of multi-phosphopeptides [170].

Advantages of MOAC over IMAC lies in the stability of oxide metal than irons with regard to pH and temperature, and better tolerance to detergents and solvents, as well as good sensitivity and selectivity [145, 171]. Compared to IMAC, TiO_2 does not show the Fe^{3+} leaching issues as mentioned above by Negroni *et al*. [168]. They appear in many formats including nanoparticles, magnetic beads and phosphopeptides-affinity MALDI plate. Nanoparticles work better than microparticles due to higher surface area [168]. The commercially available TiO_2 beads has been widely applied for phosphopeptides isolation due to its higher selectivity for phosphopeptides, as well as their robustness and tolerance towards many regents [157].

Physical properties of TiO_2 beads have certain effect on its enrichment capacity. One of the possible explanations is more binding sites resulting from a higher surface/volume ratio, which benefits for beads enrichment characteristic [144, 172]. For instance, the advantages of mesoporous over smooth particles [144] and higher selectivity of nanoparticles over micro-sized beads for phosphopeptide [172].

As mentioned before, the non-specific binding of acidic peptides was notorious for both IMAC and TiO_2 enrichment. For TiO_2, high acetonitrile and TFA concentrations in

loading buffers help to reduce non-specific interactions between acidic peptides with the beads and are mostly used in incubation solution [145]. Moreover, a variety of organic acids such as 2,5-DHB [173], phthalic acid [174], glycolic acid, ammonium glutamate, lactic acid, β-hydroxypropanoic acid [175], citric acid [176], glycerol [177] and some other acids have been proposed as non-phosphopeptide excluders. The binding interactions with metal oxides of TiO_2 was: phosphate group > organic acid > carboxyl group, therefore, the addition of organic acids competed for binding sites with acidic residues. As a result, the non-specific binding was reduced and TiO_2 enrichment specificity was enhanced. However, they all have advantages and disadvantages. Due to the co-elution of some phosphopeptides by using DHB and its high hydrophobicity, alternative reagents were applied to avoid the potential contamination from 2,5-DHB and phthalic acid on LC system and the inlet of MS as proposed by [157]. Monophosphopeptide enrichment is less efficient through using 1 M citric acid, which is possibly caused by similar binding to TiO_2 beads as monophosphopeptides [157]. Hydrophilic and soluble glycolic acid and lactic acid were preferable because of better compatibility with LC-MS than 2,5-DHB [167].

Continuous improvements are still on to solve the problem of different acids/TiO_2 enrichment. Zhao *et al*. [176] developed CATSET (citric acid-assisted two-step) strategy to overcome the problem of citric acid/TiO_2 enrichment as mentioned above [157]. Compared to the traditionally used DHB/TiO_2, more than 37% of the total phosphopeptides and 2.6-fold more of the multiphosphorylated peptides were identified by using CATSET strategy [176]. Furthermore, Kanshin and colleagues [178] noticed that peptides containing multiple glutamine and asparagine residues containing peptides (named N/O-rich peptides) are predominantly co-enriched with phosphopeptides by TiO_2 beads *via* examining the amino acids distribution patterns in both phosphopeptides and non-phosphopeptides in mammalian cells (HeLa and HEK293 cells), yeast (*Saccharomyces cerevisiae* strain YAL6B) and insect (*Drosophila melanogaster* Schneider S2 cells). The portions of poly-N/Q peptides of non-phosphopeptides varies from 11% to 17% in these examined species, and are most typical in yeast cells [178]. Elution of N/Q-rich peptides is not reported in IMAC. Addition of 125 mM asparagine and glutamine in wash buffer (70% ACN, 3% TFA) and an optimum peptide/TiO_2 ratio (408 μg/mg) were successfully applied to mitigate N/Q-rich peptides for an efficient TiO_2 enrichment [178]. Finally, a 30% increase in phosphopeptides enrichment was achieved

as well as a 5-fold decreased in the intensity of non-phosphopeptides without the obvious change of phosphopeptides intensities [178].

Other factors affecting phosphopeptide enrichment by TiO_2 beads include ratio of peptide-to-beads, incubation time and consecutive incubations have been investigated by [179] and [180]. The amount of beads shows a significant effect on enrichment selectivity. Optimum peptide/bead ratio of 1/2 to 1/8 (mass/mass) was observed for HeLa cell lysates [179]. Optimum peptide/beads ratio and incubation cycles need to be determined for different biological samples [179]. There was a slight increase of the phosphopeptide identification number with extend of incubation time and it reached a plateau at certain time (30 min) [179]. Singly phosphopeptides were predominantly determined when larger beads volume was applied, while deficient beads tended to favor the identification of multiplyphosphopeptides [179, 180].

Efficient elution of phosphopeptides from TiO_2 resins is of paramount importance. One of the big issues for using TiO_2 is the incomplete elution of multiphosphopeptides that are very tightly bound with beads. In addition to pH-dependence of phosphopeptides elution solvents, amine nucleophilicity and structure of elution solvents also play important roles in phosphopeptide recovery, especially the elution solvent structure [177, 181, 182]. pH step wise elution and pH step gradient elution were reported to be beneficial for TiO_2 enrichment specificity enhancement. Four novel phosphorylation sites of human cyclin-dependent kinase 2 were characterised using methyl esterification together with a pH step-wise elution by using firstly 0.1 M triethylammonium bicarbonate (pH 8.5) and then 3% NH_4OH pH gradient (11.5) [182]. Later, 15% more phosphopeptides and 53 more multiple phosphopeptides were identified from tryptic human neuroblastoma SH-SY5Y cells by using pH gradient elution using NH_4HCO_3 (adjust from pH 9.2 to 11.3 using NH_4OH) compared to one-step elution [181]. Besides conventionally used NH_4OH, other phosphopeptides elution solvents including NaOH, TEA, bis-Tris, bis-Tris propane and NH_4HCO_3 were compared, composition of elution solvents was proposed to play a major role in phosphopeptide recovery [177]. NH_4OH in favours of eluting shorter phosphopeptides (1-1.5 KDa); bis-Tris propane seems to greatly recover longer phosphopeptides (1-4 KDa), which contained more hydrophilic and/or more acidic residues. A total 1.4-fold more phosphopeptides were detected by

using two-step elution (NH_4OH and bis-Tris propane) from PC3 prostate cancer cell lysates [177].

Similar to IMAC, the enrichment specificity is still a bottleneck for MOAC enrichment to be dealt with.

2.2.6.1.3 SIMAC

SIMAC (sequential elution from IMAC), in which gallium-IMAC and TiO_2-MOAC treatment were combined, was performed to improve the phosphoproteome coverage especially to increase the identification of multiply phosphorylated peptides by [166]. Later, Sun *et al*., [183] found the limitation of SIMAC in respect of its reproducibility and linearity of the detected phophopeptides by using casein and standard phosphopeptides. Thus, SIMAC was proposed as a semi-quantitative method for large scale phosphoproteomics studies [183].

In conclusion, the enrichment of phosphopeptides is still a challenge, and the combined utilisation of different techniques is unavoidable at this stage. Thus, it is important to decide what the target of the research particular, since in terms of methodologies, "one size" may not fit all [184].

2.2.6.2 LC fractionation

Various LC techniques have been applied to reduce the sample complexity and thus benefit detection of low abundance phosphopeptides by downstream MS analysis. LC separation can be regarded as one of the enrichment techniques, but it's widely used in couple with IMAC/MOAC. Among these methodologies, strong cation exchange (SCX) is widely used, however, other techniques such as HILIC, ERLIC, RPLC and capillary electrophoresis (CE) -based separation were also applied.

2.2.6.2.1 Strong cation and anion exchange (SCX and SAX)

SCX and SAX techniques separate peptides based on peptide charge. For SCX, peptides carry different positive charge in acidic SCX loading buffers (usually pH ~3): an average net charge of +2, +1 and 0 for tryptic non-phosphopeptide, singly phosphopeptide and multiple phosphopeptide, respectively [144]. These peptides will firstly retain to negative charge stationary phase and will be eluted by increasing ionic strength in mobile phases. Principles for SAX are similar to SCX, but with reverse charge state materials as

stationary phase. Phosphopeptides are mostly eluted in early SCX fractions. However, acidic peptides cannot be separated efficiently from phosphopeptides by employing SCX due to similar net charge. Sample loss from multiple phosphopeptides with an average net charge of 0 [144] was another limitation. In addition, Worthington *et al*., [158] observed that reverse-phase chromatography is more preferable for phophopeptide separation than SCX, due to the enrichment of these modified peptides by the latter.

2.2.6.2.2 Hydrophilic interaction chromatography (HILIC)

Peptide separation using HILIC is based on hydrophilic interactions between peptide residues and mobile phases. Elution was achieved by an increase of mobile phase polarity. Basic peptides are retained in stationary phase, followed by the phosphopeptides and highly hydrophilic peptides. It has been used for the fractionation of complex phosphopeptides from cell lysates [185]. Similar to SCX, the co-elution of multiphosphopeptides and multi acidic residues containing peptides is one of the disadvantages.

2.2.6.2.3 Electrostatic repulsion hydrophilic interaction chromatography (ERLIC)

Research on the application of ERLIC to investigate phosphoproteomic studies were reported by Alpert [186], peptides were separated based on both charge and polarity by using a Poly weak anion-exchange (WAX) material column. Compared to SCX, a good separation of tryptic digest of rat kidney tissue was obtained by using ERLIC [187]. A significantly high number of both phosphoproteins (158% higher) and glycoproteins (0.1% higher) were detected from a single ERLIC fraction (ERLIC04) than which from SCX (SCX01) [187]. A maximum identification number (338 phosphoproteins and 583 phosphosites) was determined from ERLIC (ERLIC04). The significant phosphopeptides loss through SCX might be resulted from the C_{18} desalting and co-elution with high-abundant peptides without PTMs [187]. In addition, modified pH and concentration of ERLIC buffers benefit the retain of phosphopeptide and glycopeptides and therefore subsequent identification [187].

2.2.6.2.4 HPLC fractionation combine with IMAC/MOAC enrichment

Zarei, *et al.* [185] investigated three LC fractionation techniques: ERLIC, HILIC and SCX by combination with consecutive TiO_2 enrichment for phosphoproteomics analysis of tryptic digested HeLa cells. They found a maximum phosphopeptides identification number (6013 and nonredundant number is 3913) using SCX-TiO_2, while ERLIC-TiO_2

was preferable if one was focussed on selection of multi-phosphorylated peptides. Zarei, *et al.* [188] evaluated the combined use of SCX-ERLIC *vs* ERLIC-SCX *vs* SCX following by TiO_2 enrichment for phosphopeptides analyse of tryptic cell lysate of HeLa cells. A maximum number of nonredundant phosphopeptides (6433) was determined from ERLIC-SCX. However, 49% (*vs* ERLIC-SCX) and 15% (*vs* SCX) more multiple phosphopeptides (≥ 3 phospho sties) were detected using SCX-ERLIC [188].

Engholm-Keller, *et al*. [189] identified more than 6600 unique phosphopeptides from 300 µg of tryptic digestion of insulinoma cells by incorporating TiSH strategy (a combine use of TiO_2, SIMAC and HILIC), which resulted in 95% enrichment specificity.

Zhou, *et al.* [190] evaluated the performance of Ti^{4+}-IMAC, Ti^{4+}-IMAC-HILIC and SCX-Ti^{4+}-IMAC-HILIC that were termed 1D, 2D and 3D respectively, for efficient phosphoproteomics analysis using Lys-C and trypsin tandem digestion of human cancer cell lines: HeLa and K562. For HeLa cells: 2D outperformed 1D in terms of phosphopeptides identification and phosphorylation site localisation (143% and 131% more) when using the same starting material. It is potentially attributed to the reduction of sample complexity by HILIC separation and multiple MS runs in 2D. A better performance of 2D over 3D was also observed for both phosphopeptides (38% more) and phosphorylation sites (36% more) determination of HeLa cells, which was possibly caused due to sample losses during additional separation processes of 3D [190]. The author noted only 70% overlap of unique phosphopeptides between 2D and 3D strategies. In general, enrichment efficiency varied according to the differences of sample complexity and phosphorylation level between HeLa and K562. To achieve maximal results, 750 µg of starting material was applied for K562 while only 125 µg was needed for HeLa cells. Furthermore, 3D technique gave a better result than 2D for K562 cells, which resulted in the identification of 1.32 times more unique phosphopeptides. A round 60% of detected phosphopeptides was found in both 2D and 3D techniques.

It is therefore noteworthy to use optimised starting material, LC separation and enrichment techniques to gain enough phosphopeptides before submitting to downstream MS.

2.2.6.3 Phosphoproteome detection using MS

Traditional techniques used to detect phosphoproteins include traditional biochemical methods: radioactive ^{32}P-labelling, phospho-specific antibodies, and Western blot analysis. The ^{32}P-labelling method is highly sensitive but also highly stressful to deploy due to radioactivity and the associated follow-up procedures [191]. The accuracy and speed to identify the phosphoprotein using antibodies is excellent, however, the lack of wider commercialisation of phospho-antibodies limits their application [106]. Antibodies are not always available for the protein of interest, limiting their application.

2.2.6.3.1 MS fragmentation

MS has developed as a powerful tool for protein identification and study of PTMs at the proteome-scale [54, 93, 192]. The use of multiple approaches could increase coverage in phosphoproteomics research. CID (also known as collisionally activated dissociation, CAD), Electron capture dissociation (ECD) and ETD [193] have been extensively used for the study of PTMs and for a detailed review see elsewhere [194].

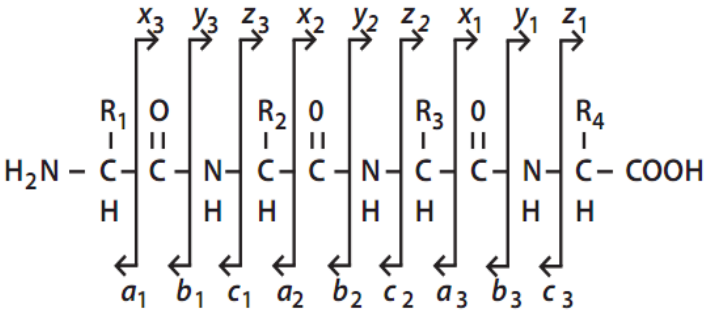

Fig 2.6 Nomenclature for fragment ions (Picture reproduced from http://www.sepscience.com by Klink [195].)

Dissociation occurs at C_α–$C_{carbonyl}$, $C_{carbonyl}$–N (i.e.,amide), or N–C_α bonds to produce a-, b-, or c-type ions if the charge is retained on the N-terminal segment and x-, y-, or z-type ions if the charge is retained on the C-terminal segment [57]. Fig 2.6 showed the common nomenclature for the MS fragment ions has been proposed by Roepstorff and Fohlman [196] and modified later by Johnson *et. al* [197].

CID is one of the most widely used peptide fragmentation techniques, it is a process where amide bonds cleave, with the result that b and y-ions dominate the product ion spectrum. CID is suitable for doubly charged and small ($\leq$ 7 amino acids) to middle sized peptides (8-11 amino acids) [198]. A difficulty is that protein PTMs such as

phosphorylation obscure the tryptic cleavage sites on the peptide backbone and the presence of basic amino acids (such as histidine) in a peptide prevent the protonation efficiency being sufficient and therefore influence modified peptide identification results [198]. The charge of the peptide precursor affects its ionization efficiency (for soft ionization approaches uses in modern MS instruments), which limits the use of CID for the study of labile PTMs (if the PTM doesn't remain on the residue during fragmentation in the MS, this reduces the ability to identify the site of modification). Alternative approaches such as ECD are usually confined with the most expensively Fourier transform ion cyclotron resonance (FTICR) MS and its wide application has been also limited by the requirement of dense near-thermal energical electrons, for which as Syka, *et al*. [193] pointed out, was still technically challenging. Because commonly used radio frequency electrostatic fields in instrument for trapping ions would made the thermal electrons kinetically excited, and not available for ECD [193, 198].

Compared to the traditional b and y-type ions observed in CID, ETD fragmentation typically produces c and z-type ions by cleavage of the C_{α}-N bond and provided more sequence information, which was more useful for peptide identification [199]. In 2007, ETD was employed by Molina, *et al*. [200] as a complementary technique for high throughput phosphopeptide localisation study and proved to be a useful tool. ETD has been shown to be efficient for highly-charged (usually +3 to +6 charge) and long peptides ($\geq$14 amino acids), and a modified supplemental low-energy method called electron transfer with collisionally activated dissociation (called ETcaD) [201, 202] was designed to improve fragmentation efficiency for doubly-charged phosphopeptides. The combining of ETD with CID for phosphopeptide identification and localisation [203] has been recommended as it can increase the identification coverage, although ETD was seen to be superior to CID for the identification of a number of phosphorylated peptides and sequence coverage. The application of CID and ETD fragmentation has been reported in many phosphoprotein studies, with more details available elsewhere [198, 204]. Kim, *et al*., [205] systematically evaluated the application of CID and ETD in large scale phosphoproteomic studies. Complementary information was observed, but still lack of proper searching algorithm when merge CID and ETD data before searching into database [205].

2.2.7 MS data analysis

2.2.7.1 Protein identification

MS/MS data sets obtained from low abundance proteins, for instance, phosphoproteins, generally has medium quality compared to that observed for high abundance proteins. Thus phosphoproteomics study requires a highly sensitive search algorithm. Widely used database search algorithms include SEQUEST [206], Mascot [207], Phenyx [208] and X!Tandem [209]. These are the most popular search engines for protein and peptide identification [93]. All of them have both advantages and disadvantages. The correlation based on SEQUEST (correlate experimental and theoretical data without a model) algorithm is famous for its high sensitivity, but the speed is slow compare to X!Tandem to process data, especially when PTMs are present [210]. It has been reported that the open-source X!Tandem algorithm is fast, but has problems of lower sensitivity and high false discovery rates [209, 211]. Some algorithms perform better at identifying certain types of proteins than others. Kapp, *et al.* [212] researched on blood specimen samples, and found Mascot and X!Tandem to be superior to SEQUEST in specificity. However, this conflicted with Searle's *et al.* working on a human plasma protein mixture [213]. Phenyx is an OLAV (a family of scoring schemes based on signal detection) scoring-based method, which was shown to be superior to Mascot in a high-throughput protein identification project [208]. Dagda and co-workers [211] found Mascot performed better individually than SEQUEST, X!Tandem and Phenyx for sensitivity and specificity. There is still an argument to the combine use of different search algorithms, called a consensus method-strategy, that could improve specificity, sensitivity and coverage [211]. In addition, the phosphopeptide identification number can be positively affected by searching parameters. For instance, 15% more phosphopeptides were identified with the incorporation of carbamylation and deamidation as variable peptide modifications, which are induced by routinely used phosphoproteomics strategies in sample preparation such as urea denature (carbamylation) and basic elution of IMAC or TiO_2 enrichment (deamidation) [158].

2.2.7.2 Phosphorylation site localisation

Various algorithms have been developed for precise phosphorylation site localisation. Since the estimation of phosphosite determination accuracy cannot directly obtained from any search engine such as the commonly used 1% or 5 % of false discovery rate (FDR)

in target/decoy database search for data estimation in peptide/protein identification [214]. False localisation rate (FLR) is proposed as a data quality indicator for phospho sites assignment. A probability based-score named Ascore, which was coupled with SEQUEST search engine [215], was proposed as an accuracy indicator for phosphorylation site localisation. Mascot delta (MD) score, as the name suggests, is from Mascot Daemon. It reflects the difference between the highest ion score and the 2nd highest ion score [216]. Ascore is proposed to benefit for phospho- Ser and Thr sites, while MD score works well for phosphorylation Tyr sites localisation. With technique developments, the application of a new algorithm provides improvement on the detetermination of more phosphorylation sites, for instance, PhosphoRS [217]. The choice of softwares for data analysis based on the MS data formats and searching engines used for protein/peptide identification. For instance, in terms of data format, the conversion of mass spectra raw data into .mgf file (Mascot generic file) and search against target database using Mascot search engine. Thus .DAT file can be obtained and will be used for MD score calculation without further process.

Section 3 Metabolomic studies

2.2.8 Technology for metabolomic studies

Metabolomics refers to the investigation of whole set of metabolites (the small molecular compounds) under specific conditions, from which differs the genetic information remains. Microbial metabolomic studies have attracted interest from the research community and the applied technologies were reviewed recently by [218].

The widely used and extensively developed nuclear magnetic resonance spectroscopy (NMR) and MS technology platforms have accelerated the field of metabolomic studies, as discussed in detail in [219]. Each method has its advantages and limitations. Briefly, the advantages of MS lies in its high sensitivity, selectivity and wide dynamic range over NMR. The combine use with gas chromatography (GC), liquid chromatography (LC), or UHPLC separation techniques prior to MS detection is greatly beneficial to reduce sample complexity [219] and therefore to the downstream identification using MS. The structure elucidation and absolute quantitation especially for volatile metabolites is one of the difficulties by using LC-MS [219] since non-volatile compounds cannot detected

by GC-MS. To make samples volatile and thus to be detected by GC-MS, chemical derivatization reaction is necessity, from which sample lose during this process is a shortage. It seems there is no single technique that can be used for universal detection of all metabolites. Therefore, combined use of NMR, GC/LC-MS and other robust analytical technologies are necessary at this stage.

Usually, metabolomics studies can be conducted as target and unknown/global metabolome. For target compound analysis, it requires extensive knowledge about the specific target and optimisation of methodology, but data statistical analysis is easier than for un-target metabolomic study. On the contrary, sample preparation is simple for global metabolome, but the later on high-through data processing and interpretation is still a bottleneck. Manual interpretation of large amount generated data is time-consuming, low-effectiveness, and experience dependent. Therefore, automatic bioinformatics tools are imperative. Comprehensive algorithms and various softwares for data processing and efficient metabolite annotation and quantitation are still in progress for untargeted metabolomics.

In terms of archaea, not many metabolomic and phosphoproteomic analyses have been carried. There are several contributing factors to this. On one hand, technical limits such as MS dynamic range is still bottleneck for low abundance phosphopeptide detection. Also, PTMs properties such as basic histidine phosphorylation is not favour for current MS technology, most of which are operated using acidic conditions. On the other hand, lack of proper methodology e.g. software for efficient interpret large scale information especially the combine -omics data are important obstacles. In addition, the incomplete genomic information limits the comprehensive interpretation of by incorporating -omics data, for instance, almost a half of genes in *S. solfataricus* P2 have no known function until now [3, 6].

2.3. Conclusions

There are some reports about re-construction of central carbohydrate metabolism (CCM) in *S. solfataricus* as recently reviewed [220] due to unique and unusual enzymes characteristics, which is partly resulted from modified metabolic pathway and hypothermal living conditions. However, these studies have not been fully completed yet.

Our study aims to provide complementary information to the existing CCM pathway and extend knowledge on how *S. solfataricus* utilise various carbohydrates through proteomic and metabolic studies of different *S. solfataricus* strains including P2, PBL2025 and PBL2073.

Additionally, phosphopeptide enrichment techniques need be optimised for specific biological microorganisms. It will aid phosphoproteome study, which is hampered by the substoichiometric of phosphorylation and low abundance of phosphoproteins. This thesis research will provide foundations for biotechnology and practical applications through further investigation of their carbohydrate metabolic pathways and potential regulation using phosphorylation mechanism in *S. solfataricus*.

Last but not the least, an -omics study which integrate genomics, transcriptomics, proteomics and metabolomics data at the system level is still not mature. There is a long way to go for this kind of application in the study of archaea.

Chapter 3 Sample preparation and technique optimisation for proteomic analysis of *Sulfolobus solfataricus* in responding to different carbon sources

3.1Abstract

The acidic thermophilic archaeon *Sulfolobus solfataricus* has been used as a model organism in archaeal research. This microorganism grows typically at 80 ºC and in low pH conditions around 2-3. It is well known that *S. solfataricus* can use different compounds as carbon sources such as D-glucose, tryptone and amino acids. For example, D-glucose can be converted into different metabolites using the Entner-Doudoroff (ED) pathways with non-phosphorylative and semi-phosphorylative branches. However, the metabolic pathway of D-glucose is not completely understood and information about their regulatory mechanisms is quite limit. Therefore, we would like to investigate the effects of different carbon sources to different *S. solfataricus* strains (P2, PBL2025 and PBL2073). Our second goal is to optimise a protocol for quantitative proteome (iTRAQ) experiments. An estimation of the limit of detection of phosphopeptides was performed for setting up a framework for further global quantitative phosphoproteome studies.

In order to achieve these goals, firstly, the effect of different carbon sources on *S. solfatarcius* growth was monitored to determine the sampling times. Secondly, different strategies were used to perform proteomic analysis such as: proteolytic trypsin digestion (in-gel and in-solution), peptide separation (using both strong cation exchange (SCX) and reverse-phase chromatography) as well as MS techniques: ion trap and Q-TOF instruments (using an electrospray HCT Ultra and maXis-UHR-TOF). In total, 13 unique phosphopeptides from 12 distinctive phosphoproteins were identified using HCT Ultra MS, with the determination of 19 phosphosites without incorporating any enrichment technique in *S. solfataricus* strain P2 grown on 0.4% D-glucose. Moreover, the minimum detectable level of phosphopeptide was determined using a maXis-UHR-TOF with trypsin digested beta casein. At least 240 µg of *S. solfataricus* is required for a single mass spectrometric run.

3.2 Background

S. solfataricus P2 is a model organism in the archaeal domain of research [1, 2]. Optimal growth occurs at pH 2-3 and at temperatures ranging from 75 to 80°C [2, 3]. *S. solfataricus* can use various sugars and tryptone as carbon sources [20]. The development of MS-based proteomic techniques has contributed greatly to unveiling the unique metabolism of this hyperthermophilic archaeon, as reviewed by Kort, *et al*., [4]. Quantitative proteomic techniques (especially iTRAQ) were widely applied to investigate the growth of *S. solfataricus* P2 on different carbon sources, such as ethanol, acetone, n-propanol or iso-propanol with the presence or absence of glucose [35, 40]. *S. solfataricus* 98/2 is geographically different from P2 [21, 22]. *S. solfataricus* PBL2025 is a widely used genetic tool in molecular biology research, which is the spontaneous derivative strain of strain 98/2. It lacks genes from *SSO3004 -SSO3050* compared to *S. solfataricus* P2. Detailed roles of the absent genes can be seen in Chapter 2. *S. solfataricus* PBL2073 is the gene *SSO3117* disruption strain from *S. solfataricus* PBL2025. In strain P2, the *SSO3117* gene was found to be involved in ethanol catabolism [35] and n-propanol metabolism pathways [40]. As a result, the metabolic pathways might be different among these three *S. solfataricus* strains (P2, PBL2025 and PBL2073), due to the absence of certain genomic regions. By comparing proteins involved in various pathways especially carbohydrate pathways at the proteomics level, the aim was to analyse proteins at the proteomic level to gain further understanding of the response of *S. solfataricus* to a change of different carbon sources and to access global changes between P2, PBL 2025 and PBL2073.

Glucose is proposed to be metabolised *via* a Entner-Doudoroff (ED) pathway with non-phosphorylative [44] and semi-phosphorylated [221] branches. A larger number of 540 identified phosphoproteins was identified in *S. solfataricus* P2 in response to changing carbon source [7]. Furthermore, 809 phosphoproteins were detected in 3 different *S. acidocaldarius* strains (including wild type and mutant strains) [47]. However, until now, the global quantitative phosphoproteome in *S. solfataricus* has not been reported. Therefore, we aim to study the phosphoproteome of *S. solfataricus* in order to investigate its importance and contribution in the regulation of the carbohydrate mechanism. In principle, the analysis of *S. solfataricus* will provide new information for extending knowledge on the pathways for regulation of carbohydrate metabolism in archaea.

MS has developed as a powerful tool for phosphoproteomic research [93]. However, this topic is still challenging, due to biological low abundance and substoichiometric nature of phosphorylation and technical limitations such as dynamic range of MS and loss of phosphate under the conventionally used positive ion MS mode as discussed in Chapter 2 [54]. Under these circumstances, the determination and handling of the amount of sample is important.

Basic proteomic techniques such as protein extraction, in-solution digestion and SCX peptide separation will be achieved easily. Phosphoproteomics studies are still a challenge, and determination of phosphopeptides is a key step.

As with many techniques, it is important to have standard operating procedures (SOPs) to ensure that all experiments are reliable and reproducible. One of the aims in this Chapter is to provide fundamental techniques for further quantitative proteomic experiments (in Chapter 4) as well as phosphoproteomics study (in Chapter 5 and 6). To satisfy the above aims, three different strains of *S. solfataricus* (P2, PBL2025 PBL2073*)* and different carbon sources (glucose, tryptone, ethanol, n-propanol, iso-propanol and acetone) were investigated.

3.3 Materials and methods

Fig 3.1 shows the workflow used for this Chapter. Cells obtained from stocks were cultured in standard glucose media, an initial OD_{650} = 0.2 was applied in all the experiments. *S. solfataricus* were collected in the late exponential growth phases based on the growth curves. Cell growth curve was determined *via* absorbance of cells (OD_{650}) using a spectrophometer (Ultrospec-2100 Pro UV/Visible, Amersham Biosciences, USA). Specific growth rates (μ) were calculated (biological triplicates) graphically by plotting OD_{650} *versus* time. Proteomic analysis was carried out as described in the following workflow.

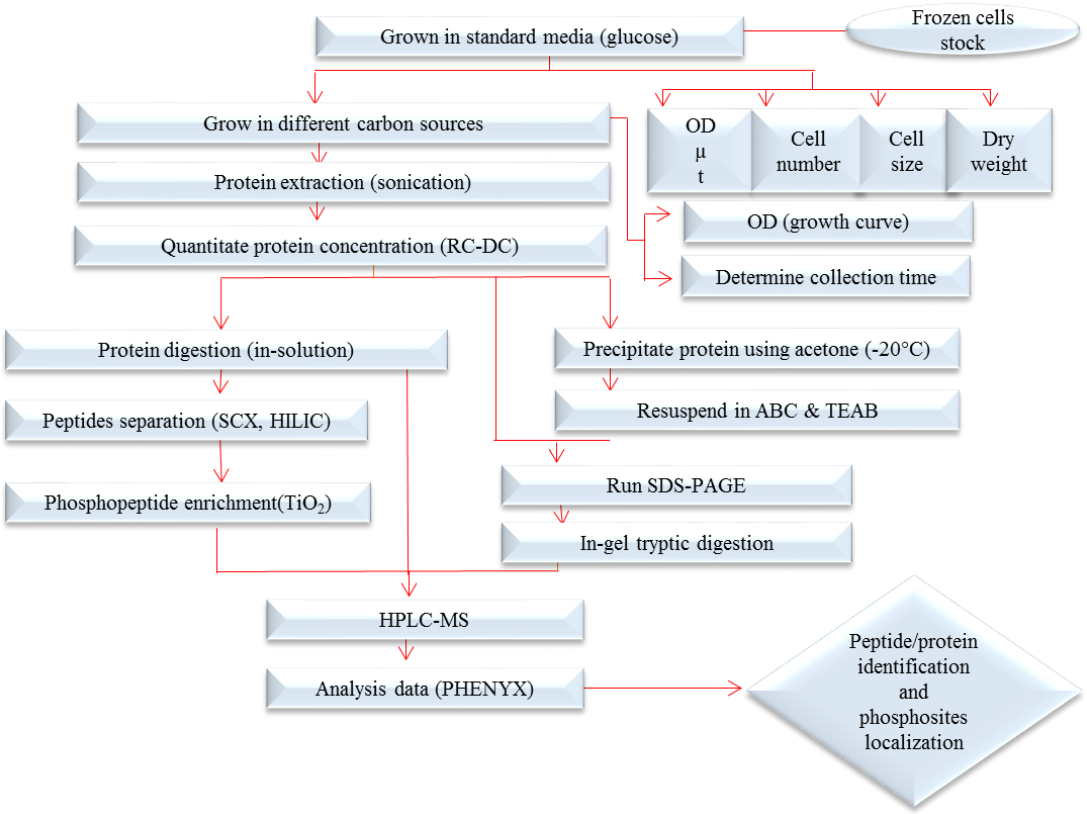

Fig 3.1 Workflow used for comparison analysis of *S. solfataricus* strains.
OD is optical density. μ is the specific growth rate and t is the doubling time. SCX is strong cation exchange. HILIC is hydrophilic interaction chromatography. TiO_2 is titanium dioxide. ABC is ammonium bicarbonate and TEAB is triethyl ammonium bicarbonate buffer. SDS-PAGE is sodium dodecyl sulfate polyacrylamide gel electrophoresis.

3.3.1 Cell culture and protein extraction

3.3.1.1 Cell culture and sampling time

S. solfataricus grows on a variety of different carbon sources such as various sugars and tryptone [1]. It also could utilize alcohol and acetone as reported by Chong, *et al*. [35]. To determine whether the *S. solfataricus* strains PBL2025 and PBL2073 (deleted gene *SSO3117*) could metabolise tryptone, acetone, ethanol and isopropanol as carbon sources, cells were cultured with tryptone (0.2%), acetone (0.4%), ethanol (0.8%), isopropanol (0.8%) and glucose (0.4%) separately, at 80°C, 120 rpm in a horizontal shaking thermal incubator (Thermotron, Infors, UK). Glucose was used as a positive control. The reference strains were sub-cultured to make stock cultures and were then stored at -80°C. *S. solfataricus* strains were activated at optical density (OD) at wavelength of 650nm (OD_{650}=1.0$\pm$0.1) using 0.4% glucose. All the experiments were started (initial OD_{650}=0.2)

in 150 ml of standard media as described by Zaparty, *et al*.[222]. The OD_{650} values were corrected against evaporation (negative control with media only). Cells were collected in the late exponential growth phases based on the growth curve. A volume of 50 ml of cultured cells was harvested and centrifuged at 6,000 x g for 10 min at 4°C. Cells pellets were then collected and stored at -80°C until required.

3.3.1.2 Haemocytometer for cell counting

To investigate if the cell number of *S. solfataricus* and OD_{650} share a linear relationship, we counted the number of cells using a haemocytometer and an optical microscope. The haemocytometer was initially invented for counting blood cells and gradually was widely used for determine other cell counts [223]. It consisted of two counting champers with ruled grids and a cover slip. Cell number counting was done by following the counting rules (left and top count, right and down not count) and calculation of cell numbers per millimetre followed Strober's [223] recommend protocol.

The surface area of the central 25 squares was $1mm^2$ (1.0 mm x 1.0 mm), and the depth was 0.1 mm for each square, thus the total volume was 1×10^4 mm^3. Cell numbers per ml = average numbers of 1.0 mm^2 square x dilution times x 10^4 (since 1ml = 10^3 cm). The samples were diluted to different times according to the growth phase to make sure the cell number within the range from 50 to 200 in each 1/25 central square.

3.3.1.3 Protein extraction and trypsin digestion

Cell pellets were collected as described in Section 2.1.1. Protein extraction was performed as described previously [224]. Briefly, protein was extracted using an ultra-sonicator (Branson SONIFIER 450, UK) for 8 cycles with alternative sonication for 45 seconds and incubation in ice for 90 seconds at 70-80% duty cycle. Lysis buffer that contains 50 mM Tris-HCl (pH 7.5) and a 5 mM concentration of each of the following chemicals: sodium vanadate, sodium fluoride, 2-glycerol phosphate and sodium pyrophosphate [225] was added to inhibit the activity of phosphatases. Supernatants containing soluble crude cell extracts were centrifuged at 12000 × g for 30 min at 4 °C to reduce potential protein degradation. Protein concentration was determined by the Bio-Rad RC-DC Protein Quantitation Assay (Bio-Rad, UK), and protein was then stored at -20°C until needed.

500 µg of protein sample was dried in a vacuum concentrator (Eppendorf Concentrator 5301, Germany) before resuspending in either 50 mM ammonium bicarbonate (ABC) or

0.5% sodium dodecyl sulphate (SDS) or denaturing with 8 M Urea. Samples were reduced by 10 mM DL-Dithiothreitol (DTT) for 60 min at 56°C, before alkylated by 55 mM idoacetamide in 50 mM ABC for 30 min in dark. Samples were then diluted with 40 mM ABC in 9% acetonitrile (ACN) to ensure that the final concentration of urea was less than 1 M before trypsin was added with a ratio of 1/40 (trypsin/protein) and incubated at 37°C for overnight. Digested peptides were dried in a vacuum concentrator.

3.3.1.4 SDS-PAGE and in-gel tryptic digestion

To test the efficiency of protein extraction, a range of (1, 5, 10 and 20 µg) protein amounts was run on 0.75 mm thick acrylamide SDS-PAGE composing of 12% of resolving gel (30% Acrylamide/Bis, 10% SDS and 1.5 M Tris-HCl pH 8.8) and 4% stacking gel (30% Acrylamide/Bis, 10% SDS and 0.5 M Tris-HCl pH 6.8). Protein sample preparation for SDS-PAGE was performed as described elsewhere [160]. Briefly, proteins were resuspended in laemmli-sample buffer containing β-mercaptoethanol, then incubated at 95°C for 5 min to denature protein. Gels were run on a Mini-PROTEAN Tetra Electrophoresis system (Bio-Rad, UK). The running programme was applied as follow: 80 V for 20 min, then 180 V for 50 min. Subsequently, gels were stained with coomassie brilliant blue 250 (Fisher scientific, UK). Gels containing 20 µg of proteins were used for in-gel digestion [160]. These gels were cut into 24 pieces before being washed and de-stained with 200 mM ABC in 40% v/v ACN. Then, the reduction and alkylation processes were performed in a similar way with in-solution digestion. Gel pieces were washed several times using 50 mM ABC before adding trypsin, and then were kept overnight at 37°C.

3.3.1.5 Acetone precipitation

Acetone precipitation has been used for soluble protein purification [226] and remove of impurities such as salts and detergents which would interfere with MS analysis [227]. Five time volumes of ice-cold acetone (-20°C) were added to sample. The mixtures were then kept overnight at -20°C, then supernatants were removed post centrifuge 12000 x g for 20 min. Different resuspension buffers including 50 mM ABC at pH 7.5 and triethyl ammonium bicarbonate (TEAB) were tested. In total, 100 µg protein was used for precipitation and then 20 µg protein (in different resuspension buffers) used to run on SDS-PAGE to test the difference between with and without acetone precipitation as well as ABC and TEAB re-suspension.

3.3.2 Peptide separation and purification

3.3.2.1 Strong cation exchange (SCX)

Limit of detection (LoD), dynamic range, and sample reproducibility of MS would be negatively affected by sample complexity. To get a better sample characteristic in a proteomics/phosphoproteomics experiment, the effects of high sample complexity and background need to be minimised by employing approaches such as SCX/HILIC chromatography, to reduce sample complexity by fractionation.

Ion exchange chromatography has proven to be a powerful tool for protein separation and purification [185, 228, 229]. SCX has been widely employed for peptide separation in PTMs study especially phosphoproteomics study, which was based on charge interactions between peptides and stationary phase [230]. Digested peptides (200 µg) were re-dissolved in 70 µl of SCX buffer A (20% ACN, 10 mM KH_2PO_4, pH 2.85) prior to loading onto a 2.1 x 200 mm Poly SULFOETHYL-A column (Hichrom Limited, UK) connected with a DIONEX BioLC GS50 Peek System pump (Dionex, USA). The fractionation of peptides was performed at a flow rate of 0.2 ml/min using a gradient as follow: 0% of buffer B (20% ACN, 10 mM KH_2PO_4, 0.5 M KCl, pH 2.85) for 5 min, ramped up to 40% of buffer B for 20 min, then ramped up to 100% of buffer B and kept for 10 min, and finally 0% of buffer B for 10 min. Peptides were detected by an UV detector (Dionex UVD170U, USA) at wavelength at 214 nm. Fractions were collected every minute, and dried in the vacuum concentrator.

3.3.2.2 Hydrophilic interaction chromatography (HILIC)

Peptide fractionation using HILIC based on hydrophilic interactions technique has been studied for complex phosphopeptides from cell lysates [185]. A total of 200 µg of digested peptides were re-suspended in 90 µl HILIC buffer A (90% ACN, 10 mM ammonium formate, pH 3) prior to loading onto a 4.6 x 200 mm Poly HYDROXYETHYL-A column (Hichrom Limited, UK) connected with an Agilent 1100 Series HPLC system (Agilent, US) which consists of a G1311 QuatPump, G1379 DEGASSER and G1314A UWD UV detector at wavelength of 280 nm. Peptides were fractionated by increasing percentage of buffer B (10% ACN, 10 mM ammonium formate, pH 4) at a flow rate of 0.5 ml/min. The HILIC gradient started with 0% of buffer B for 10 min, ramped up to 60% of buffer B for 15 min, then ramped up to 100% of buffer B

for 5 min, and finally 0% of buffer B for 5 min. Fractions were collected every minute and dried in the vacuum concentrator.

3.3.3 Technique optimisation for phosphoproteomic studies

3.3.3.1 Limit of detection (LoD) curve

Compared to high abundant non-phosphopeptide, low abundance phosphopeptides are not easy to detect by MS analysis as mentioned in Chapter 2 (Section 2.1.4). To determine the LoD of phosphopeptides using TiO_2 enrichment for MS (Bruker maXis UHR-TOF instrument) analysis and to estimate amount of *S. solfataricus* proteins used for phosphoproteome experiment, the commercialized β-casein, with five known phosphoserine sites was chosen as a standard for method development. Larsen, *et al.* [173] reported four different phosphopeptides in bovine β-casein, the monophosphorylated peptide ^{48}FQ[pS]EEQQQTEDELQDK^{63} (M_r 2060.82), and three tetraphosphorylatedpeptides
^{22}NVPGEIVE[pS]L[pS][pS][pS]EESITR^{40} (M_r 2352.85),
^{17}ELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR^{40}(M_r2966.16), and
^{16}RELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR^{40} (M_r 3122.27).

A range of concentrations of β-casein start from 5 to 2.5, 0.5, 0.1, 0.02, 0.004 pmol and 8 fmol were used. We also investigated the interference of complex sample to TiO_2 enrichment of phosphopeptides from β-casein by using 500 µg of trypsin digested cell lysates of *S. solfataricus*. Peptides mixture was enriched by TiO_2 (50 µl beads slurry, ratio of 1/10 of beads/protein) and C_{18} desalting was performed before its submission to LC-MS/MS. A signal to noise (S/N) ratio was obtained automatically by using Analyst QS (V.1.1). A S/N ratio correspond to three times the noise level for targe P-peptide was choosed [221]. In addition, the relationship between the signal changes of targeted phosphopeptide with the changes of injection concentration of β-casein was obtained graphically by plotting β-casein concentration *versus* S/N ratio of target phosphopeptide.

3.3.3.2 Phosphopeptide enrichment by TiO_2

TiO_2 enrichment was applied as it has been widely reported for phosphopeptide enrichment [185, 231, 232]. Certain amounts (500 µg, calculated based on LoD curve) of trypsin digested peptides from whole cell lysates of *S. solfataricus* grown on glucose were desalted using C_{18} material. Then, phosphopeptide enrichment was carried out using

magnetic TiO_2 beads (GE Healthcare TiO_2 Mag Sepharose, UK). The loading, washing and elution of samples were performed as recommended from the manufacturers (GE Healthcare, UK). Briefly, peptides were re-suspended in 100 µl of loading buffer containing 80% v/v ACN, 1% v/v TFA and 1 M glycolic acid to inhibit binding of non-phosphopeptide with 50 µl of TiO_2 beads. Then, the non-phosphopeptides were removed twice by washing buffer containing 80% ACN and 5% FA, while the elution of phosphopeptides using 50 µl of 5% ammonia, pH 12. The enriched phosphopeptide sample was dried to completion in a vacuum concentrator.

3.3.4 Mass spectrometry (MS) and associated data analysis

All the dried peptides were desalted using a C_{18} column 200-mg C_{18} Sep-Pak SPE column (Sigma, UK) and dried with a vacuum concentrator before submitting to a nano-HPLC (Dionex, UK) coupled to Esquire HCT Ultra ion trap mass spectrometer (Bruker, Germany). Peptides from gel slices (20 µg sample) and SCX fractions (200 µg sample) were re-dissolved in 10 µl of MS buffer consisting of 0.1% FA and 3% ACN, then submitted into an ESI-ion trap (HCT Ultra) MS.

In addition, to access the performance of different MS instruments, one of the high intensity SCX fractions was submitted into high resolution and sensitivity MS: maXis UHR-TOF (Bruker, Germany) using the same LC gradient for HCT Ultra ion trap MS: A [0.1% FA in 3% ACN] and B [0.1% FA in 97% ACN]. The increasing organic proportion was used for peptide separation (separation ramp from 3% to 45% B within 40 min, followed by a 5 min ramp to 95% buffer B, and then holding at 5% buffer B for 10min). The pump flow rate was constant at 300µl/min. The ESI detector was set at a mass range of 250-2000 m/z for data acquisition in full MS scan under the positive ion mode. During 1 s MS scan, peptides with a +2, +3, and +4 charge state were selected for fragmentation. Three precursors were selected for MS/MS.

A phosphorylation specific neutral loss scan of H_3PO_4 (97.97, 48.88 and 32.66 Da) or HPO_3 (79.97, 39.99, 26.67Da), which correspond to +1, +2 and +3 charged peptides are predominantly occur upon CID in LC-MS/MS. This neutral loss program was incorporated to solve ion suppression phenomenon using β-casein as a standard. Ion suppression is the detection of low abundant phospho-peptides that are strongly

suppressed by the presence of high abundant non-phosphopeptides [58]. Usually, upon neutral loss of phosphoric acid (98Da), phospho-serine and phospho-threonine were converted into dehydroalanine (69Da) and dehrdroaminobutyric residue (83 Da) respectively in MS/MS spectra. But most of phospho-tyrosine containing peptides stay stable due to the structure properties.

Usually, the neutral loss of H_3PO_4 (98Da) and HPO_3 (80 Da), especially the neutral loss of phosphoric acid on Ser and Thr caused by β-elimination was dominant upon CID fragmentation for the commonly used LC-MS/MS program. As a result, it reduces the product b- and y- ions intensity and caused the lacks of the informative fragmentation on backbone for phosphopeptide identification and precise site localization [233]. Therefore, MS/MS/MS (MS^3) spectrum fragmentation was required to provide informative sequence fragmentation for effective phosphoproteomics study as described [60]. MS^3 program was initiated only if neutral loss peak was dorminant in MS^2 spectra to save time [60]. The MS^3 program was then used for analysing phosphopeptides from *S. solfataricus* samples.

All MS data were were converted to MGF files using Bioanalyst version 4.0 (Bruker Daltonics, Germany) before submitting to an in-house Phenyx searching engine (v.2.6, Geneva Bioinformatics, Switzerland) as well as Mascot Daemon (V2.5). *S. solfataricus* P2 protein database was downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes) in Oct 2013. The .mgf files was used to search against both forward and reversed decoy databases to estimate data quality by calculation of false discovery rate [214]. The search parameters were used as follow: trypsin as digestion enzyme with up to one missed cleavages, carbamidomethyl (C) as a fixed modification, oxidation (M) and phosphorylation (STY) as variable modifications. MS and MS/MS tolerances were set up at 1.2 and 0.6 Da respectively. The MS and MS/MS for one the SCX fractinonation run on MaXis were used 0.1 and 0.1Da respectively. Phenyx/Mascot data were then exported into excel (Microsoft 2010, USA) for further analysis.

3.4 Results and discussions

3.4.1 *S. solfataricus* growth files

3.4.1.1 The growth of *S. solfataricus* on different carbon sources

Different *S. solfataricus* strains were grown on different carbon sources: glucose (0.4%), tryptone (0.2%), ethanol (0.8%), acetone (0.4%) and isopropanol (0.8%). Biological triplicates of growing cells were applied and results are summarised and shown in Table 3.1. It was reported by Chong *et al*. [35] that *S. solfataricus* P2 could grow on isopropanol, ethanol and acetone. It reached a maximum consumption rate when grown on 0.79 % (w/v) ethanol [35]. Further pre-culture of both *S. solfataricus* PBL2025 and PBL2073 strains on acetone, isopropanol and ethanol were applied but both strains were demonstrated to be unable to grow on tested culture as sole carbon sources (data not shown). Therefore, these strains were grown on 0.4% glucose and 0.2% tryptone and will be used for further experiments.

Table 3.1 Growth of *S. solfataricus* strains on different carbon sources

Carbon source / Strain	0.4% Glucose	0.2% Tryptone	0.8% n-propanol	0.8% Ethanol	0.4% Acetone	0.8% Isopropanol
P2	+	+	+	+	+	+
PBL2025	+	+	-	-	-	-
PBL2073	+	+	-	-	-	-

Note: +: Grow. -: Not grow. Data of strain P2 grown on 0.8% ethanol, 0.4% acetone and 0.8% isoproponal were obtained from Chong, *et al*.,[35, 40, 41].

Their growth curves were obtained by plotting the optical density at a wavelength of 650 nm (OD_{650}) against time (Figs 3.2A and B). Specific growth rate (μ) and doubling time (t=Ln2/t)) were calculated based on these data, as can be seen in Table TS3.1 (in Appendix). From Figs 3.2A and B, we can see that growth of P2, PBL2025 and PBL2073 on standard glucose media are slightly different. Compared to strain P2, the lack of some ABC transporter encoding genes in PBL2025 and PBL2073 results in a longer lag phase, as shown in Fig 3.2A.

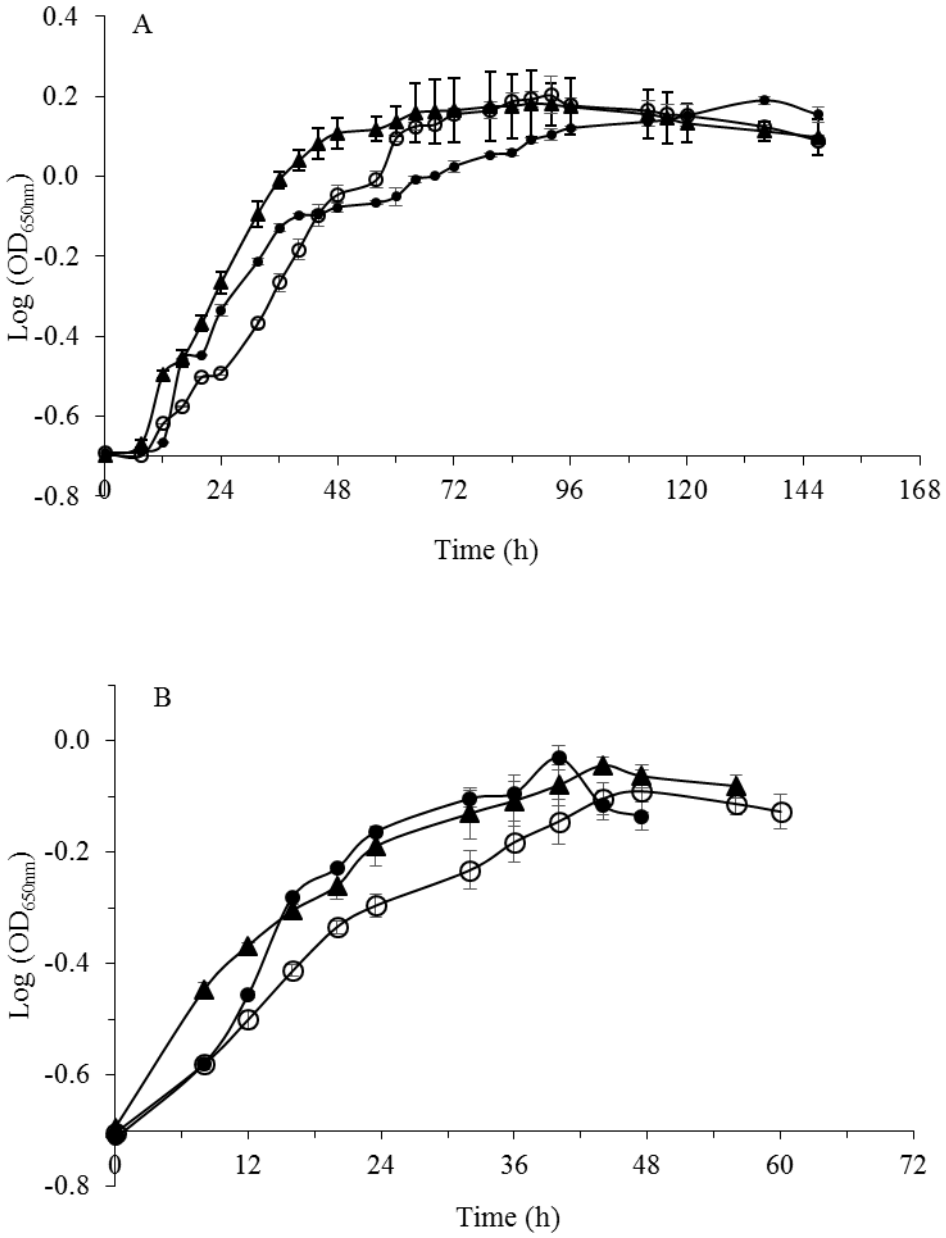

Fig. 3.2 Standard growth curve of *S. solfataricus* P2 (●), PBL2025 (○) and PBL2073 (▲) grown on 0.4% glucose (A) and 0.2% tryptone (B). Data were performed from biological triplicate cultures.

However, the average specific growth rates of both mutant strains are higher. PBL2073 grows slightly faster (15% higher than PBL2025 and twice more than P2) than the others on 0.4% glucose (0.0106 $\pm$0.002 h^{-1} and 0.0195 $\pm$0.001 h^{-1}, 0.0214 $\pm$0.0001 h^{-1} for P2, PBL2025 and PBL2073 respectively). It is inferred that some other ABC transporters may be activated, which will contribute a faster growth of both mutant strains. In addition, the absence of some genetic information in mutant strains may explain the growth difference. *SSO3117* gene in *S. solfataricus* P2 has recently been named as aldehyde dehydrogenase [66]. The wide substrate specificity of SSO3117 might provide some hints

to explain growth differences between PBL2025 and PBL2073. Furthermore, the activity of aldehyde dehydrogenase in *E.coli* provide some information for the role of the gene *SSO3117* deleted in *S. solfataricus* PBL2073 [75]. We inferred that some metabolic pathways in mutant strains might need to be activated, therefore it took longer time for these mutant strains to enter exponential phase than the wild type strain. This hypothesis will be elucidated *via* both proteomic and metabolomic analyses in Chapters 4, 7 and 8. The maximum OD_{650} of these strains were not different (based on student t-tests): 1.55 $\pm$ 0.01 for P, 1.59 $\pm$ 0.05 for PBL2025, and 1.52 $\pm$ 0.08 for PBL2073.

Tryptone is a mixture of free amino acids, which was obtained from a pancreatic enzyme digestion of casein in milk. When 0.2% tryptone was supplied, all *S. solfataricus* strains grew slowly and reached the stationary phase earlier (less than 40 h) than those grew on 0.4% glucose (more than 60 h) (Figs 3.2A and 3.2B). Although the maximum OD_{650} values were almost the same when these strains were grown on tryptone (0.75 $\pm$ 0.03), the growing time to reach these values was different: 34, 42 and 33 h for P2, PBL2025 and PBL2073 respectively. We postulate the lack of usable specific amino acids might affect the growth of *S. solfataricus* on 0.2 % tryptone media as indicated by the work of Sezonov and team [234]. They [234] found that cells physiology and size changed significantly when the OD_{600} reached up to 0.3 and when L-Ser and L-Thr amounts became limited.

To estimate the correlation between OD_{650} and number of cells, cells were collected at different OD_{650} and cell number were counted using an optical microscope with a haemocytometer. Cell cultures were diluted with distilled water to ensure numbers of cells were in a range for counting as mentioned in method section. As a result, a linear relationship between cell number ($x10^7$) and OD_{650} was observed and linear relationship equations were shown in Table TS3.2 in the Appendix.

3.4.2 Sample preparation

3.4.2.1 Protein extraction and SDS-PAGE

Cells were collected in the late-exponential phase for P2, PBL2025 and PBL2073, respectively, which corresponds to an OD_{650} range of 1.01 $\pm$ 0.03 when grown on 0.4% glucose or a range of 0.70 $\pm$ 0.05 with the presence of 0.2% tryptone. To investigate

protein extraction process, we carried out SDS-PAGE analysis with known amounts of protein, the evaluation process was performed based gel intensities, as showed in Fig 3.3 A and B. As expected, the highest amount of proteins resulted in high gel intensities in Fig. 3.3 A and B, corresponding to 20 µg amount of protein. Gel bands of corresponding to 20 µg protein (in Fig 3A) were cut and digested with trypsin before submitting to an LC coupled with HCT Ultra MS. Data were then used for a quick assessment of digestion efficiency.

Acetone has been used in proteomic analysis as a purification of proteins [226] and quantitative research based on the selective tagged peptides using acetone modification [227]. There was a report that acetone precipitation and ultracentrifugation could provide complementary protein identification data for human urinary proteins [235].

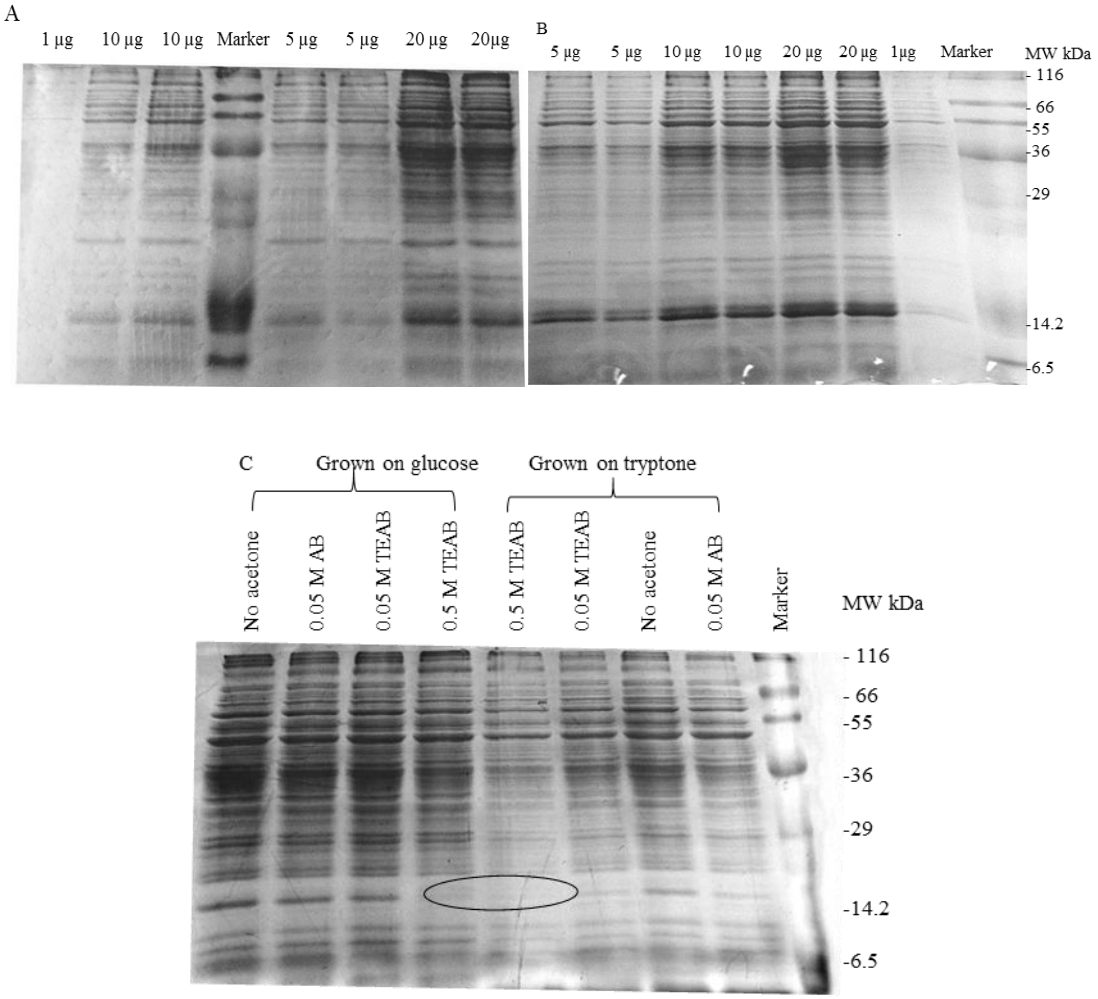

Fig 3.3. SDS-PAGE of protein from *S. solfataricus* (A) P2 grown on 0.4% glucose. (B) Grown on 0.2%tryptone. (C) SDS-PAGE of acetone precipitated protein of *S. sulfataricus* grown on 0.4% glucose and 0.2% tryptone, and re-suspend in pH7.5, 0.05 M AB, 0.05 M TEAB and 0.5 M TEAB.

Different re-suspension buffers were used after acetone precipitation of protein extracted from *S. solfataricus* and results were shown in Fig 3.3C. We can see that the gel bands intensity were reduced after precipitation treatment by acetone. It is inferred that impurities may be removed after acetone treatment as the band sharpness was different compared to which without acetone treatment. It might indicate deep desalting or ionic detergents washing out, which could be an effective method to remove impurity particles for downstream MS analyse. Therefore, it is recommend to perform protein acetone precipitation before further operation. Fig 3.3C shows similar a gel band pattern when precipitated proteins were re-suspended in either 0.05M AB or 0.05 M TEAB. Compared to 0.05 M of re-suspension buffer, a gel band (as indicated in Fig 3.3C) 'disappearance' was observed when protein was resuspended using 0.5 M TEAB buffer. By considering TEAB buffer is most widely used in labelling-based quantitative proteomic techniques such as iTRAQ, 0.05 M TEAB was recommended for future work.

3.4.3 Technical optimisation

3.4.3.1 Sample fractionation by high performance liquid chromatography (HPLC)

A single or combination employment of different LC fractionations techniques was beneficial for reducing sample complexity and resulted into improvement on proteome coverage. It shows special importance for detection of low abundance phosphopeptides [58].

3.4.3.1.1 SCX chromatography

The widespread use of SCX separation of complex peptides was performed [228] and results are shown in Fig 3.4 A-C. To test the performance of the MS instruments, trypsin-digested bovine serum albumin (BSA) was used as a standard. A technical replicate was applied to test the reproducibility of instrument as well as efficiency of trypsin digestion. From 20 µg of digested BSA, more than 20 sharp peaks (intensity over 40 mAU) were observed (as shown in Fig. 3.4A). Similar chromatography was observed by using a couple of BSA (data not shown). It indicates a good BSA digestion and stability of the instrument. In Fig 3.4B, as expected, we can see that higher amount of samples used resulted in higher absorbance intensity (200 µg vs 20 µg). However, samples needed to be desalted by C_{18} columns before submission to the LC-MS/MS. To test the desalting efficiency, 200 µg of trypsin digested cell lysates was treated by different C_{18} columns

(smaller capacity from NEST, USA and big capacity from SUPELCO, UK) and results are shown in Fig 3.4C.

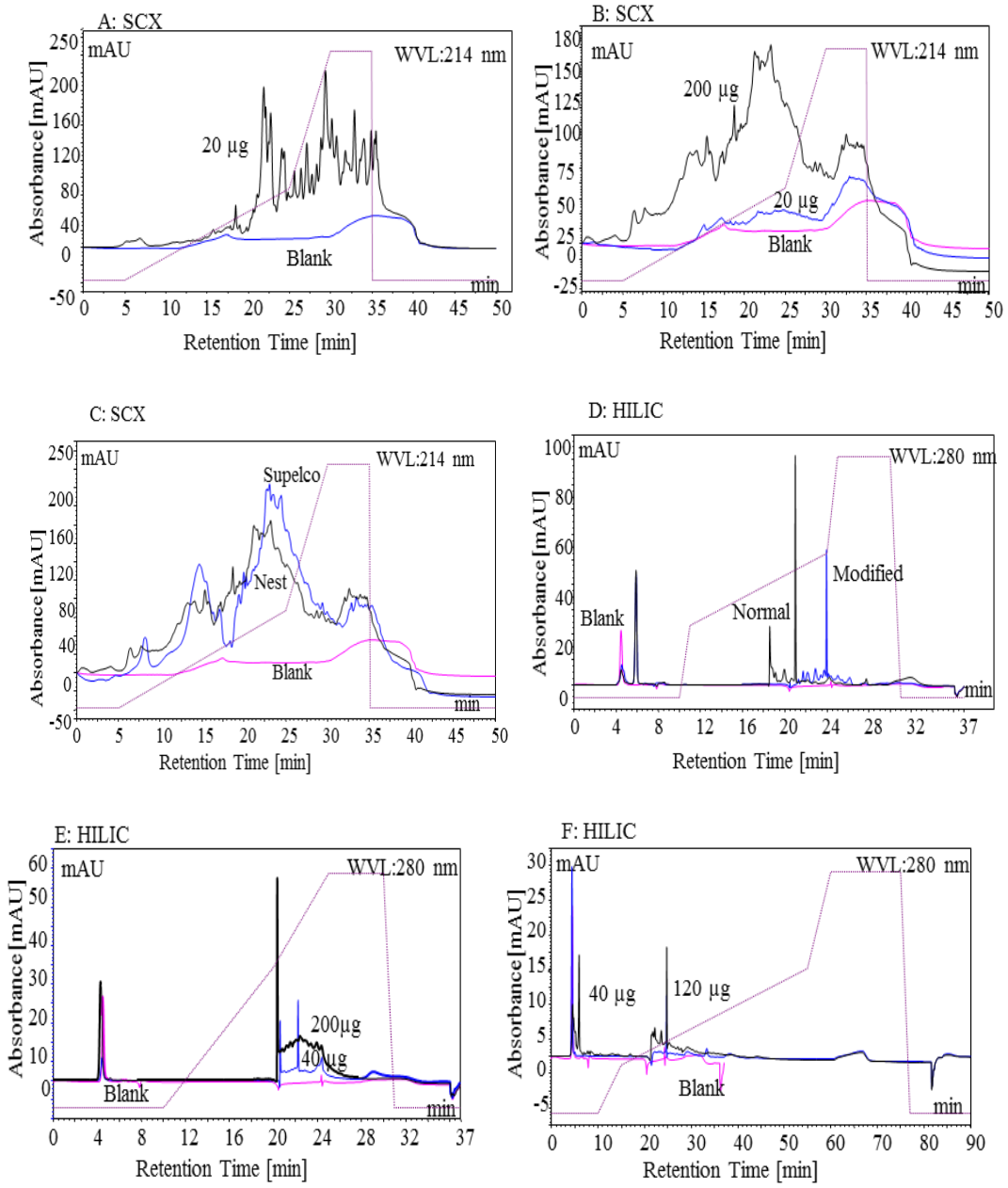

Fig 3.4 SCX chromatography of (A) 20 µg BSA and blank. (B) 200 µg and 20 µg of digested proteins from *S. solfataricus* grown on 0.4% glucose and blank. (C) 200 µg of digested proteins from *S. solfataricus* destalted by different C_{18} materials supplied by supelco, nest and blank. HILIC chromatography of (D) normal and modified gradient program for 200µg BSA and blank. (E) 40 µg and 200 µg digested protein and blank. (F) 90min program separation for 40 µg and 120 µg protein of *S. solfataricus* grown on 0.4% glucose and blank.

In Fig 3.4C, similar chromatography peaks were observed which indicate similar desalting capacity. The intensity between two desalting materials are almost the same except for the first two peaks (as indicated in 11-15min in Fig 3.4C) which might be caused by a difference in binding capacity of C_{18} materials supplied or insufficient washing out of glycerine surrounding the polypropylene (C_{18} material).

A total of 32 SCX fractions from 200 µg digested proteins (in Fig 3.4C) were collected by every minute and desalted using a SUPELCO C_{18} column (Sigma, UK) before submitting to a nano-HPLC coupled to an HCT Ultra MS. Biological replicates were run using SCX, but only one set of SCX fractions were run on LC-MS/MS (HCT Ultra) due to the limit of instrument. Furthermore, to test the difference between different MS, the 22th min fraction with highest SCX intensity was desalted and submitted to both HCT-Ultra and maXis UHR-TOF.

Tryptic digestion of proteins mainly occurred in lysine and arginine residues, and N-methylation of lysine residues in *S. solfataricus* [236, 237] proteins have been reported to be involved in thermal stability. In such a situation, the amino acid residues from peptides digested by trypsin are mostly arginine, and carboxyl terminal would carry a proton under an acidic environment (pH 2.85). Many studies have been carried out to improve tryptic digestion efficiency, and a detailed discussion was reviewed in Chapter 2. These include work from Taouatas and co-workers [238] who separated the same single charge phosphopeptide from N-acetylated peptides using metallo-endopeptidase (Lys-N cleavage specificity) and a modified gradient. Different protein/trypsin ratios were tested (data not shown), an optimum 35/1 ratio was recommend for *S. solfataricus* whole cell lysates.

3.4.3.1.2 HILIC chromatography

Basic peptides would be retained in HILIC, due to hydrophilic interaction with mobile phase (water in buffer B), followed by the phosphopeptides and highly hydrophilic peptides. A modified gradient was applied to improve peptide separation (Fig 3.4D). Fig 3.4E shows chromatography of different amounts of *S. solfataricus* samples (40 and 200 µg used). From Fig 3.4D, BSA peptides were eluted mainly from the 20th to 28th min, and numbers of peaks was less compared to which in SCX. One of the potential reasons was that most of the peptides were non-absorbable under an UV

wavelength of 280 nm. A modified 37 min gradient (c in Fig 3.4D) to get a better separation was used for 20 µg BSA. An extended gradient (90 min) was also investigated (Fig 3.4F) with the aim to improve peptide separation. However, improvements was not significant, as shown from the chromatograph in Fig 3.4F.

In summary, efficient LC separation chromatography is important for global quantitative proteomics/phosohoproteomics studies.

3.4.3.2 Protein identification based on shotgun technique

3.4.3.2.1 Integration of proteomic results from SDS-PAGE and SCX fractions

The identification number between SDS-PAGE and SCX fractions is not directly comparable, because the starting protein amount is different for these two methods, 20 and 200 µg, respectively. To assess about in-gel and in-solution tryptic digestion, selective gel slices from S1 to S10 and SCX fractions (even number start from 14 to 32 min fractions) were run on the HCT-Ultra MS. The number of proteins and peptides identified from each selected gel slice/SCX fractions are detailed in Tables TS3.4 and TS3.5 in the Appendix. As we can see, 412 unique peptides from 177 proteins were identified from selected in-gel digestion. 221 unique peptides from 137 proteins were detected from selective SCX fractionations using in-solution digestion.

Moreover, the identification of unique phospho- peptides/proteins from in-gel *vs* in-solution tryptic digestions are listed in Table 3.2. Also, detected numbers from selected runs are listed. It can be seen that 7 and 11 phosphopeptides were detected for in-gel and in-solution tryptic digestions respectively. A number of 6, 3 and 1 phosphosites were detected on Ser, Thr and Tyr for in-gel tryptic digestion, whilst the phosphorylation sites distribution on Ser/Thr/Tyr was 8, 10 and 2 for in-solution tryptic digestion. Furthermore, phosphoprotein coverage, which shows the localization of the identified phosphoproteins in the subset of protein samples analysed from SCX fractionations and SDS-PAGE, is shown in Fig 3.5. The results of course give an indication only as the whole proteome was not analysed. From our data (in Table 3.2), 7 phosphoproteins were identified from the 22^{th}, 24^{th}, 26^{th}, 28^{th} and 30^{th} min fractions among a total of 137 *S. solfataricus* proteins identified from selected SCX fractions.

Table 3.2 Identified phosphopeptides from selected SCX fractions and gel slides.

		AC	Peptide Score	MD Score	P-site	PTMs
	F22	13814256	39.33	35.47	S	D[pS]KKVLFLVGEEFEDIELLYPFYR
	F22	13813147	31.87	23.73	YS	ES[pY]ILV[pS]YPK
	F22	13815497	26.55	19.93	S	EVLEEYGF[pS]
	F24	13815813	25.81	21.7	TST	QI[pT]GR[pS]EFN[pT]VYFNNVK
	F26	13814125	27.79	21.65	T	ASELVDLMHKQGL[pT][pS]GK
	F28	13814112	26.19	18.75	S	NL[pS]EYKAANLMGLTPAAVSNYLKSR
SCX	F30	13815142	26.77	23.45	S	ELLGYL[pS]KLLK
	S2	13816095	26.65	20.51	YSS	PPTGA[pY]RGLGIPPAVLVLENLVK[pS]I[pS]K
	S4	13813729	25.89	17.3	T	GNEM[pT]DELR
	S6	13815002	28.19	14.91	Y	[pY]YEEIGINR
	S6	13813882	23.41	20.71	S	IMNAGALGAEIII[pS]GKL[pT][pT]ERAR
	S7	13813351	30.43	26.25	STT	FETKK[pY]FFTIIDAPGHR
	S8	13814861	30.95	24.61	TTS	MKI[pT]VV[pT][pS]GLRSNYSGGSVHVNNVVR
	S8	13813184	29.79	24.98	TS	ILL[pT]V[pS]ALVK
	S9	13813396	30.02	19.73	T	MKD[pT]KVK
	S9	13813584	26.59	21.76	STT	[pS]AYFMTAL[pT]F[pT]DGKIIK
	S9	13813724	26.1	22.9	T	[pT]REELAKK
Gel	S9	13813216	27.29	26.17	S	VGEKLIY[pS]K

Note: P is phosphorylation. MD score is Mascot delta score for P-site localization..

3 phosphoprotein from the 22th mins, and another 4 phosphoproteins were identified at the four fractions. It is in agreement with previous report that phosphopeptides can be eluted in early SCX gradient fractions due to their lower net charge state caused by negative charge of phosphate in low pH [228]. It also indicates that the gradient program needs to be improved to achieve a better separation performance. Moreover, large amounts of protein as starting material and phosphopeptides enrichment techniques need to be considered for global phosphorylation studies.

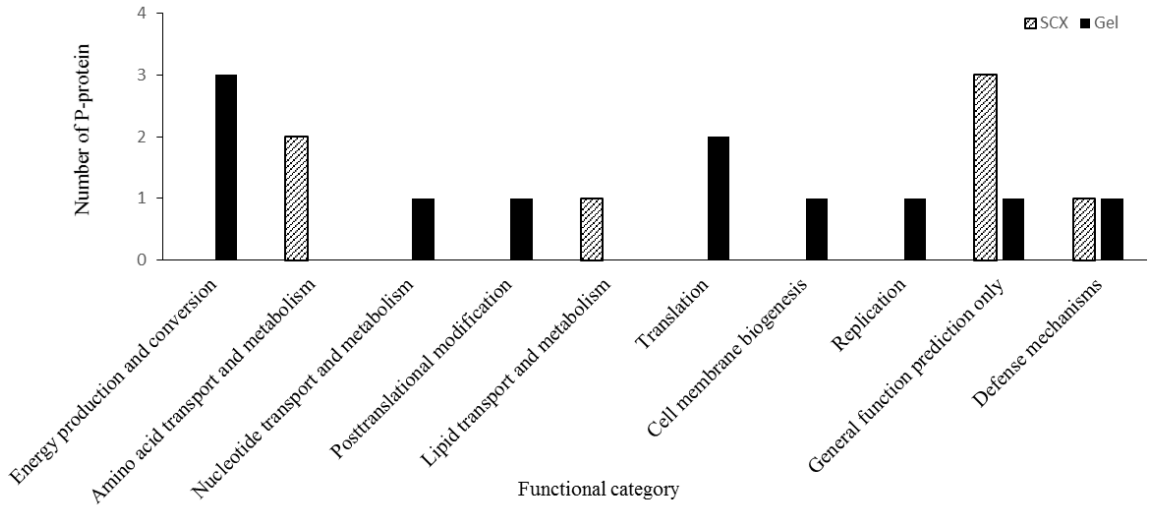

Fig 3.5 Localization of phosphoproteins identified from SDS-PAGE (black) and SCX (slash).

Compared to in-solution digestion, disadvantages of in-gel tryptic digestion include its time-consuming nature, handling in cutting gels (often is laborious) [160] when preparing samples for MS. Also, there is a high-risk of losing proteins (50-85%) from in-solution digestion [161] due to the multiple washing procedures (including stain, de-stain and in-gel digestion). In addition, the proteome coverage by employment of gel-based technique was better that from in-solution digestion, which is possibly due to better digestion efficiency. However, the limit detection of extremely acidic/basic proteins of SDS-PAGE gel need to be considered. Taken together, SDS-PAGE gel was used in our study to estimate protein extraction and trypsin digestion process, and in-solution digestion was carried out for further studies.

Trypsin digested cell lysates (about 4 mg proteins) were submitted to SCX chromatography, followed by TiO_2 enrichment. C_{18} desalted phosphopeptides were submitted to the HCT-Ultra ion trap MS. Unfortunately, not many phosphoproteins/proteins were determined (SCX chromatography were shown in Fig FS3.1 and Table TS3.6 in the Appendix B), which indicates the necessity for optimisation of phosphopeptide enrichment and MS detection techniques. Moreover, a method named ERLIC (Electrostatic repulsion hydrophilic interaction chromatography) has been reported for phosphoproteomics study by Alpert [186]. Furthermore, Hao and colleagues [229, 239] compared ERLIC and SCX techniques for identification of phosphopeptides/phosphoproteins, as a result, more phosphopeptides and phosphoprotiens were detected using ERLIC compared to SCX. It was suggested that the

loss of phosphopeptides might be caused by C_{18} desalting after SCX separation. The combination use of SCX, HILIC and ERLIC coupled with TiO_2 enrichment has been investigated [185], offering a powerful technique for phosphoproteomics studies.

In summary, LC separation chromatography is important, but efficient phosphopeptide enrichment seems to be necessity for global quantitative phosphoproteomic studies.

3.4.3.3 Different MS instruments

Besides HCT-Ultra ion trap, high resolution and sensitivity of MS instruments including Amazon and maXis-UHR-TOF were applied for phospho- peptide/protein detection. In total, 86 peptides from 70 proteins were identified from the 22thmin SCX fraction using HCT Ultra ion trap. In contrast, 195 unique peptides from 130 unique proteins were determined when the same fraction was submitted into the maXis-UHR-TOF MS. It is inferred that high resolution and sensitivity of the MS instrument shows a positive effect on peptides/proteins identification results. However, only 2 phosphopeptides corresponding to 2 phosphoproteins were identified using either MS instrument from the same fraction (as listed in Table 3.2). Moreover, only 2 phosphopeptides (from 148 peptides identified) corresponding to 2 phosphoproteins (from 77 proteins) were identified from a control experiment (without TiO_2 enrichment) using the Amazon MS. That is an obvious evidence of the ion suppression phenomenon. Therefore, conventionally used TiO_2 beads was applied to enrich phosphopeptides before submitting to the LC-MS/MS.

3.4.3.3.1 MS program for neutral loss of H_3PO_4

A MS program for assigning neutral loss of H_3PO_4 (97.97, 48.88 and 32.66 Da) on phosphopeptides was performed on the Amazon MS using β-casein as a standard. Using this program, the ion suppression phenomenon was mostly resolved. Phosphopeptides including ^{33}FQ[pS]EEQQQTEDELQDK^{48} (M_r 2061.83 Da) and ^{30}IEKFQ[pS]EEQQQTEDELQDK^{48} (M_r 2432.08 Da) were detected using with and without neutral loss scan. However, a peptide with multiple phosphosites ^{17}ELEELNVPGEIVE[pS]L[pS][pS][pS]EESITR^{40} (M_r2966.16) was uniquely detected by using the MS neutral loss program. Therefore, this MS neutral loss program was applied for identification of phosphopeptides of *S. solfataricus*, which were enriched by TiO_2 beads. As expected, the ion suppression phenomenon was eliminated for the complex samples. As a result, 5 phosphopeptides were observed from a total of 32 total

identified peptides (for neutral loss scan program). In contrast, only 2 unique phosphopeptides from 2 phosphoproteins were detected using the normal MS program. However, the total phosphopeptides and phosphoproteins number are not favourable for global phosphoproteome study, so further experiments will be applied on the high-resolution maXis Q-TOF.

Table 3.3 Phosphopeptide of β-casein detected using different MS programs

MS program	Detected phosphopeptide of β-casein
Normal MS	^{33}FQ[$_P$S]EEQQQTEDELQDK^{48}, ^{30}IEKFQ[$_P$S]EEQQQTEDELQDK^{48}
Neutral loss	^{33}FQ[$_P$S]EEQQQTEDELQDK^{48},^{30}IEKFQ[$_P$S]EEQQQTEDELQDK^{48} ^{17}ELEELNVPGEIVE[$_P$S]L[$_P$S] [$_P$S] [$_P$S]EESITR^{40}

Table 3.4 Phosphopeptide of *S. solfataricus* detected using neutral MS program

MS	P-rotein	P-peptide sequence	P-site	z	m/z
Neutral loss	15897630	EMAKILRDEA[$_P$S]WDYDEAK	S	3	2248.97
	15897546	CGVFAV[$_P$S]SPKEVNIQLVVEGIR	S	4	2480.25
	15897389	IGVV[$_P$S]GKGGVGKSFVSSNLAMAIAASGR	S	4	2955.28
	15897225	VALTSLG[$_P$S]K	S	2	954.48
	15897157	IF[$_P$T]AV[$_P$S]SSSLVEEYLKK	TS	3	2059.95

Note: P is phosphorylation. Z is charge state of peptide.

3.4.3.3.2 Limit of detection (LoD) for phosphopeptides

To determine the LoD for phosphopeptides, β-casein (Mr 24000 Da), with the commercialized and known phosphorylation on serine was chosen. Four of the five known phosphoserines can be found from trypsin digestion products and a phosphopeptide (residues 33-48, phosphorylation on Ser^{35}) from in-gel tryptic digestion of 250 fmol of β-casein was still detectable using nano-ESI LC-MS/MS [149]. To ensure detection of phosphopeptides, also, by considering the sample loss during desalting step, a large amount 5 pmol of in-solution trypsin digested β-casein was applied in our experiment. The LoD of phosphopeptides was determined by using a decreasing concentration of β-casein starting ranging from 5 pmol to 2.5, 0.5, 0.1, 0.02, 0.004 pmol and 8 fmol in amount. To investigate the interference of complex sample to the detection intensity of phosphopeptides, 500 µg of tryptic cell lysates of *S. solfataricus* as a simulation of the complicated biological samples was mixed with β-casein.

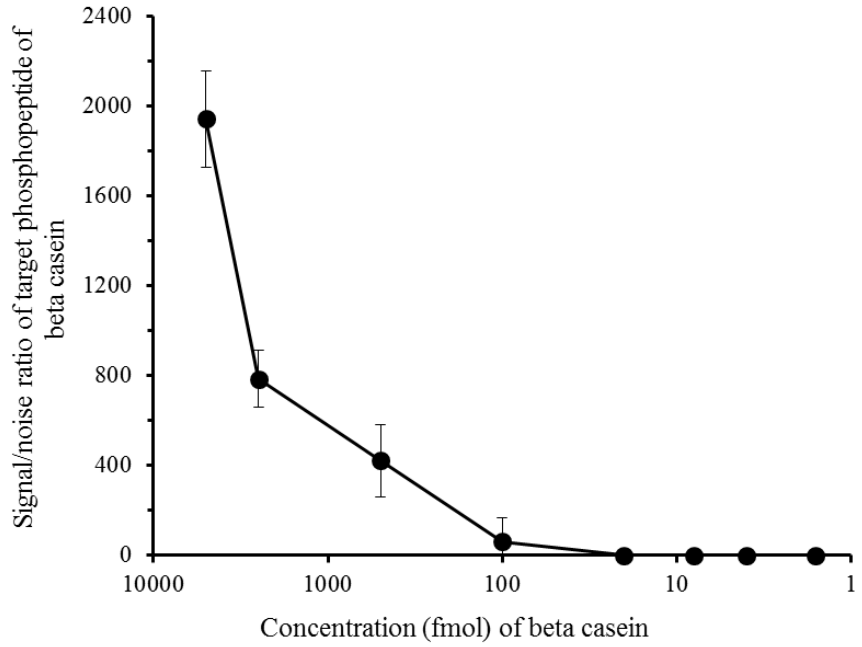

Fig 3.6 Realationship between different concentrations of of β-casein (from 5 x 10^3 fmol to 2500, 500, 100, 20, 4 and 1.6 fmol) with interference (generated by tryptic cell lysates of *S. solfataricus*).

Peptide ^{33}FQ[pS]EEQQQTEDELQDK^{48} (M_r 2061.83, +3 charged of 687.95) was mostly abundant peptide detected from the digested beta casein. It was used as a target peptide for investigating a LoD of phosphopeptide. The signal to noise ratio (S/N) of the target peptide *vs* β-casein concentrations with *S. solfataricus* is shown in Fig 3.6. There was an obvious decrease: approximately 2.5 times down for detection of target phosphopeptide when the concentration of β-casein decreased. When the amount of beta casein was less than 1 pmol (0.024 µg), the low abundance target phosphopeptide signal cannot be observed because of inadequate signal to noise ratio, which results from the complex background noise from digested cell lysates. The whole genome of *S. solfataricus* P2 is 3 Mb [3], and supposed the phosphorylation level was less than 0.01% [54]. Therefore, we could estimate the amount of sample (240 µg) for phosphorylation study by LC-MS/MS analysis.

The limit of detection of phosphopeptides for beta casein and *S. soltataricus* cell lysates were 0.024 µg and 240 µg respectively. However, higher amounts of *S. soltataricus* samples (500 µg) was recommended for phosphopeptide enrichment analysis due to the loss of sample during desalting, LC separation processes such as SCX.

3.4.3.4 Phosphopeptide enrichment

Based on the LoD curve determined above, peptides digested from 500 µg of *S. solfataricus* proteins were used for phophopeptide enrichment with a 50 µl TiO_2 beads slurry. Furthermore, a control sample where phosphopeptides were not enriched was also performed. Only 2 phosphopeptides (from 148 peptides identified) corresponding to 2 phosphoproteins were identified using the Amazon without enrichment, whilst 18 phosphopeptides (from 249 peptides identified) corresponding to 18 phosphoproteins were detected using TiO_2 beads enrichment. A total of 27 unique phosphopeptides corresponding to 22 phosphoproteins were detected without enrichment using the maXis Q-TOF, covering 9.2% of phosphorylated peptides compared to total peptides detected. A higher number of 54 phosphopeptides corresponding to 53 proteins were detected using TiO_2 beads enrichment. However, the total peptides and proteins identification number is also high using the maXis, which indicates the necessity of optimisation of the phosphopeptides enrichment workflow.

Furthermore, common issues were observed for phosphopeptide identification by analysing the sequence information of both phosphopeptides and non-phosphopeptides, which is in agreement with a previous report. For instance, compared to non-phosphopeptides, a higher percentage that were triply charged was observed: 12.5% of *vs* 5.1% from SCX fractions detected using HCT-Ultra ion trap and 50% *vs* 29.4% of neutral loss program detected by the Amazon and 61.1% *vs* 33.3% from TiO_2 enrichment results detected through maXis Q-TOF. In total, the phosphopeptide enrichment technique using whole cell lysates of *S. solfataricus* will be applied and detailed results can be found in Chapter 5.

3.5 Conclusions

Some studies have re-constructed the CCM pathways in *S. solfataricus* P2 as detailed in Chapter 2. However, these pathways have not been completely understood and there was no system report on amino acid metabolism [3, 6]. This study provides complementary information to the existing glucose glycolysis pathway in *S. solfataricus* and extends knowledge on how *S. solfataricus* can utilise various carbohydrates *via* a study of different *S. solfataricus* strains (P2, PBL2025 and PBL2073) in responding to different carbon sources.

The growth curves of different *S. solfataricus* strains grown on 0.4% glucose and tryptone as sole carbon source were determined. Cell pellets from P2, PBL2025 and PBL 2073 were collected in the late exponential phase further experiments including proteomic (Chapter 4), phosphoproteomic (Chapters 5 and 6) and metabolomic (Chapter 7) analyses. Moreover, pre-culture of PBL2025 and PBL2073 strains in ethanol, acetone and n-propanol shows that both strain cannot adapt and utilize these compounds as *S. solfataricus* strain P2. Further experiments will be applied when strains were grown on glucose and tryptone.

We speculate that pathways involved in degradation of glucose and tryptone in *S. solfataricus* PBL2025 might be different from that in *S. solfataricus* P2. The lag phase of *S. solfataricus* PBL2025 was different from that of P2 and PBL2073. The differences of specific growth rates among three strains (*S. solfataricus* P2, PBL2015 and PBL2073) might be due to lack of some important genes/proteins in PBL2025 and PBL2073, such as the ABC transporter coding genes. The uptake difference of glucose among these strains will be measured. Moreover, we speculate that the gene *SSO3117* disruption strain *S. solfataricus* PBL2073 is not adapted to utilize ethanol/n-propanol as the sole carbon source and might been converted to acetaldehyde/propionaldehyde while then accumulated in the cell as indicated in Fig 2.1 in Chapter 2. However, the conversion from acetaldehyde to acetate for ethanol metabolism is a unique direction manner, which perhaps causes its death in early growth phase.

The identification of peptides/proteins derived from SDS-PAGE coupled with in-gel tryptic digestion *vs* in-solution digestion coupled with SCX chromatography was performed, which lay a fundamental proteomics technique for further quantitative proteomic studies. Proteomic (based on isobaric tags for relative and absolute quantitation (iTRAQ)) and metabolomic analyses will be applied to elucidate pathways involving degradations of different carbon sources and results are presented in Chapter 4.

The LoDs of phosphopeptide for beta casein and *S. solfataricus* was 0.024µg and 240µg respectively. However, 500µg of *S. solfataricus* proteins were used for further phosphopeptides enrichment experiments because of the loss during sample preparation and other processes (e.g. SCX) as well as the application of technical replicates for MS

analysis. The optimisation on phosphopeptide enrichment methods for large scale phosphoproteome studies will be performed and results are discussed in Chapter 5.

Chapter 4 Quantitative proteomic analysis of *Sulfolobus solfataricus* utilising different carbon sources

4.1 Abstract

The extreme thermoacidophilic crenarchaeon *Sulfolobus solfataricus* has been widely used as a model organism in archaeal research. Here, three 8-plex iTRAQ experiments were performed to generate in-depth quantitative proteomic data sets for *S. solfataricus* strains P2, PBL2025 (a spontaneous mutant) and PBL2073 (a genetic mutant). This approach was used in a combination with bioinformatics methods (enrichment analysis) to identify the functional proteins present at different stoichiometry between P2 and PBL2025 when grown on the same (glucose) carbon source, as well as those affected by changing the carbon source from glucose to tryptone.

A total of 158 proteins (27 % of quantified proteins) were determined as being present in different ratios amounts between *S. solfataricus* P2 grown on glucose compared to *S. solfataricus* PBL2025. Among these differentially abundant proteins, 61 proteins involved in amino acid and carbohydrate metabolism were present in reduced amounts in the PBL2025 compared to P2 strains, whereas 97 proteins belonging to carbon fixation, butanoate metabolism and Val, Leu, and iso-Leu biosynthesis and so forth were greater in abundance.

In contrast, when changing carbon source from glucose to tryptone many changes occurred at proteome level for both these strains: 159 and 208 (27% and 34.2% of quantified proteins) proteins show abundance change for P2 and PBL2025. These changes involved central carbohydrate metabolism including glycolysis and pyruvate metabolism, nucleotide metabolism including purine and pyrimidine metabolism (for P2), or energy metabolism including oxidative phosphorylation (for PBL2025). A significant change in amino acid metabolic pathways was found in both strains. Furthermore, these results demonstrate that the absence of 46 genes (in PBL2025 strain) affected the regulation of many metabolic pathways, and the change of carbon sources from glucose to tryptone resulted in the inhibition of most amino acid biosynthesis.

4.2 Introduction

The extremely thermoacidophilic crenarchaeon *Sulfolobus solfataricus* grows at 80°C and pH 3 [1], and it utilises a variety of sugars and amino acids as sole carbon and energy sources [20]. *S. solfataricus* P2 was first isolated in Italy [2], whilst *S. solfataricus* PBL2025 strain (a spontaneous mutant of *S. solfataricus* 98/2 strain(s)) was isolated in the US [240]. The *S. solfataricus* P2 genome has been fully sequenced with about 3 Mb sequence consisting of a single chromosome and encoding 2977 proteins [3]. *S. solfataricus* P2 has been used as a model microorganism, while *S. solfataricus* PBL2025 has been widely used as a biological tool for construction of mutants (by deleting target genes) in biological research [241]. Compared to P2, PBL2025 lacks genes from *SSO3004* to *SSO3050* [242], in which 6 genes are supposed to be involved in central carbohydrate metabolism (CCM), 8 genes play role in energy metabolism, 8 genes are composed of transporters, and 2 genes function in the biofilm formation and extracellular polymeric substances (EPS) secretion [43].

Glucose or tryptone as a single source of carbon and energy for *S. solfataricus* induces different metabolic enzymes, transporters and uptake systems to support growth and cell survival. In 2006, Snijders *et al.*,[39] using ^{15}N metabolic labelling based proteomics (yeast extract and tryptone *vs* glucose), reported 3% of the identified genes and 14% of the identified proteins were differentially regulated more than 2 fold; these genes/proteins were involved in CCM. Further interest in quantitative proteome level of *S. solfataricus* (PBL2025 and P2) was inspired when the biofilm formation differences between these strains was observed [43] and the role of SSO3006 and SSO3019 in EPS formation was confirmed [42]. Recently, Esser, *et al.* [7] suggested that the blocking of sugar degradation pathways and activation of enzymes in gluconeogenesis of *S. solfataricus* was enhanced when tryptone was supplied. Esser, *et al*., [7] found the phosphoproteome of the P2 strain were dramatically affected when the carbon source was changed from tryptone to glucose. The immediate purpose of this study is to generate an in-depth quantitative proteome data set for both *S. solfataricus* strains (P2 and PBL2025). To achieve this, we apply an iTRAQ approach, since this technique has been routinely successfully applied to the proteomics studies of *S. solfataricus* [4].

Three sets of 8-plex iTRAQ experiments were performed to investigate the quantitative proteomics level of these two strains when tryptone or glucose was used as a sole carbon source. Furthermore, bioinformatics tool is applied to identify prominent pathways affected by changing carbon source from glucose to tryptone. Maps of comprehensive pathways for both *S. solfataricus* P2 and PBL2025 grown on either tryptone or glucose in standard medium are constructed based on the iTRAQ data.

4.3 Materials and methods

4.3.1 Growth conditions

S. solfataricus strains P2 and PBL2025 were grown in media containing either 0.2% tryptone (w/v) or 0.4% glucose (w/v) as a sole carbon and energy source in a horizontal shaking thermal incubator (Thermotron, Infors, UK) at 120 rpm at 80°C. Cells grown on glucose were used as a control. Stock cultures of these strains were stored at -80°C and activated using either 0.4% glucose grow up to $OD_{650} = 1.0 \pm 0.05$; or 0.2% tryptone grow up to $OD_{650} = 0.70 \pm 0.05$. All the experiments were performed at an initial OD_{650} of 0.2 ± 0.05 in 150 ml of standard media as described elsewhere [38]. Cell growth curves were obtained by measuring the optical density at a wavelength of 650 nm using a spectrophometer (Ultrospec-2100 Pro UV/Visible, Amersham Biosciences, US) against time. The specific growth rates (μ) were calculated (biological triplicates) graphically by plotting OD_{650} *versus* time. The OD_{650} values were also corrected against evaporation (negative control without *S. solfataricus*). The pH of growth medium was measured by (Mettler Toledo 320 pH meter, UK) against the growth at OD_{650nm} for all *S. solfataricus* strains when tryptone or glucose was supplied.

Based on the growth curves, cells were collected at late exponential growth phases for further experiments. A volume of 50 ml of cell culture was harvested and centrifuged at 5,000 x *g* for 10 min at 4°C and cell pellet was collected and then stored at -80°C until required. Numbers of cells were counted using a haemocytometer (Marienfield, Germany) under an optical microscope (Axiostar Plus, USA).All chemicals were purchased from Sigma-Aldrich (Gillingham, UK) unless otherwise stated.

4.3.2 Protein extraction and labelling

Crude protein extraction was performed as described elsewhere [38]. Briefly, cells were washed with extraction buffer (500 mM TEAB, triethylammonium bicarbonate pH8.5

containing 0.05% SDS (w/v)). Then re-suspended in the extraction buffer and one volume of acid washed glass beads (425-600 μm) (compared to cell pellet) was added. Protein was extracted using a disruptor (Genie Vortex, USA). Supernatants containing soluble crude extracts were centrifuged at 21000 × *g* for 30 min at 4 °C and then transferred into low binding tubes. To purify proteins, 5 time volumes of ice-cold acetone (compared to sample) were added and samples were left at -20 °C for overnight. Samples were centrifuged at 12000 × *g* for 20 min at 4 °C, these protein pellets were then re-suspended in the extraction buffer, and protein concentrations were determined using the Bio-Rad RC-DC Protein Quantitation Assay (Bio-Rad, UK).

A total of 100 μg proteins of each phenotype was firstly reduced with 50 mM tris 2-carboxyethyl phosphine hydrochloride at 60 °C for 1 h, then alkylated with 200 mM methyl methanethiosulfonate at room temperature for 10 min before digested by trypsin at a ratio of 1:40 (trypsin:protein) at 37 °C for overnight. Digested samples were then labelled using 8-plex iTRAQ reagents according to the manufacturer's protocol (AB Sciex, USA). After labelling, samples were mixed and dried in a vacuum concentrator. The detailed labelling of samples is shown in Table 4.1A and B.

Table 4.1 Labeling of iTRAQ experiments.

Table 4.1 (A) The 1st set of iTRAQ.

Strain	**P2**			**P2**			**PBL2025**	
Carbon source	Tryptone			Glucose				
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd
iTRAQ reagent	116	117	118	113	114	115	119	121

Table 4.1 (B) The 2nd set of iTRAQ.

Strain	**PBL2073**			**PBL2025**				
Carbon source	**Glucose**			**Glucose**			**Tryptone**	
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd
iTRAQ reagent	115	116	117	118	119	121	113	114

Note: 1st, 2nd, and 3rd indicates biological replicates.

Table 4.1 (C) The 3rd set of iTRAQ.

Strain	**PBL2073**			**PBL2025**		
Carbon source	**Tryptone**			**Tryptone**		
	1st	2nd	3rd	1st	2nd	3rd
iTRAQ reagent	113	114	115	116	117	118

Note: 1st, 2nd, and 3rd indicates biological replicates.

Common biological samples were applied in different iTRAQ sets. For instance, *S. solfataricus* PBL2025 grown on glucose were used for the1st and the 2nd iTRAQ sets. PBL2025 cells grown on tryptone were used for the 2nd and the 3rd iTRAQ sets.

4.3.3 Hydrophilic interaction chromatography (HILIC)

Labelled peptides were re-suspended in 90 µl of HILIC buffer A (80% ACN, 10 mM ammonium formate, pH 3) prior to loading onto a 4.6 x 200 mm Poly HYDROXYETHYL-A column (5 µm, 200Å, Hichrom Limited, UK) coupled with an Agilent 1100 Series HPLC system (Agilent, US) consisting of a G1311 Quat Pump, G1379 Degasser and G1314A UWD UV detector operated at a wavelength of 280 nm. Peptides were fractionated using a gradient as follows: 10 min of buffer A before ramping up to 20% of buffer B (5% ACN, 10 mM ammonium formate, pH 5) for 5 min then up to 60% of buffer B for 50 min, then ramped up to 100% of buffer B for 10 min and kept for 10 min and finally 0% of buffer B for 5 min at a flow rate of 0.5 ml/min. Fractions were collected every minute, then 35 high intensity fractions were chosen and dried in a vacuum concentrator (Eppendorf Concentrator 5301, Germany).

4.3.4 LC-MS/MS and data analyses

Each dried HILIC fraction was re-dissolved in 20 µl of MS loading buffer containing 3% acetonitrile and 0.1% formic acid, and then 10µl of sample was introduced to a nano-LC-ESI-qQ-TOF-MS/MS, QStarXL Hybrid ESI Quadrupole time of flight-tandem mass spectrometer, (Applied Biosystems, Framingham, MA; MDS-Sciex, Concord, Ontario, Canada). Details of the nano-LC system, MS/MS parameters and LC gradient were used as described in details elsewhere [38].

Protein identification and quantitation were mainly carried out as described elsewhere [224]. Briefly, MS/MS raw data were firstly converted into MGF format using a Mascot dll script in Analyst QS v. 1.1. Data were then submitted to a Phenyx searching engine (v.2.6, Geneva Bioinformatics, Switzerland) using *S. solfataricus* P2 protein database (2994 proteins) downloaded in April 2011 from NCBI (http://www.ncbi.nlm.nih.gov). Parameters for searching were set up as follows: MS tolerance was set at 0.4 Da and MS/MS tolerance were set as following: peptide tolerance 0.2 Da, charge +2 and +3, +4, minimum peptide length, *z*-score, maximum *p*-value and AC score were 5, 5, 10^{-6} and 5, respectively, and trypsin enzyme was used with two missed cleavages permitted for both cases. Modifications were set: 8-plex iTRAQ mass shifts (+304 Da, K and N-term) and methylthiol (+46 Da) as fixed modification and oxidation of methionine (+16 Da) as variable modification. The results were then exported to Excel (Microsoft 2010, USA) for further analyses. These data were also searched within the reversed *S. solfataricus* P2 database to estimate the false-discovery rate (FDR) as detailed elsewhere [214]. Data were then analysed using our in-house statistical approach uTRAQ server (v4.0) and SignifiQuant (v4.0, University of Sheffield) to determine regulated proteins [224]. The significantly up- or down- regulated proteins mentioned in the following text refers to the direction of change, because the fold change underestimation issue remains unsolved until now but direction of change remains unaffected [224].

The quantified proteins were categorized based on the arCOG functional code (http://archaea.ucsc.edu/arcogs/). Furthermore, to gain an understanding of how different *S. solfataricus* strains responded to different carbon sources at proteomic level and to determine which pathways were significantly regulated, an enrichment test analysis based on the hypergeometric distribution model was used as describe in the equation below [243]. A p-value was calculated for each pathway in every regulated group and pathways which achieved p-values < 0.05 were considered to be significantly affected.

$$y = f(k \mid N, M, n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

Where: *N*: total number of predicted proteins. *M*: number of proteins that is annotated to an arCOG category for gene function enrichment analysis or a specific KEGG pathway enrichment analysis, respectively. *n*: number of up or down regulated proteins. *k*: number

of proteins that are regulated and also annotated to specific KEGG pathways or gene functional categories.

4.4 Results and discussion

4.4.1 Cell growth profiles

4.4.1.1 Strain PBL2025 compared to P2 grown on glucose

From Fig 4.1A we can see that PBL2025 gained a similar growth trend with P2 (but PBL2025 shows a higher specific growth rate) in the standard glucose media, however, it showed a longer lag phase than P2 (18 hours compared to 10 hours). PBL2025 reached the stationary phase slower (with a slightly higher maximum OD_{650}) than P2. It is inferred that the growth difference might be a result from the lack of genes *SSO3004-SSO3050* in PBL2505 as discussed below, in which 6 genes are involved in CCM, 8 genes play role in energy metabolism and 8 genes are composed of transporters [242]. The protein phosphorylation involved in the CCM of *S. solfataricus* was noticed to be affected when changing carbon source from glucose to tryptone as reported previously by [7].

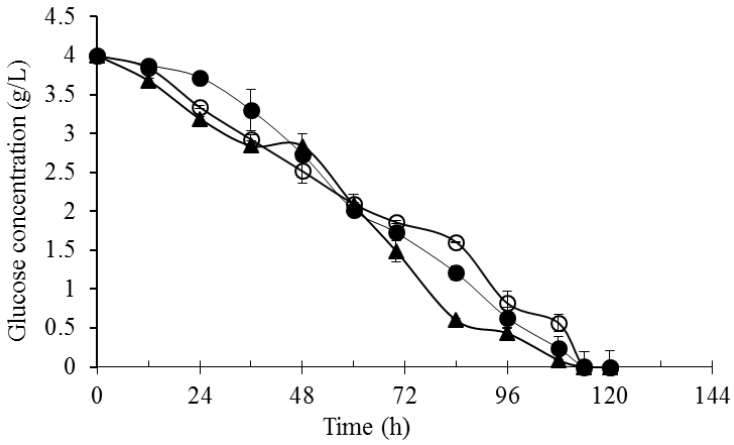

Fig 4.1. Glucose consumption profiles of *S. solfataricus* P2 (○) and PBL2025 (▲) grown on 0.4% (w/v) glucose. Data were performed from biological triplicate cultures.

The longer lag phase of PBL2025 could be explained by the lack of the ATP-binding cassette (ABC) transporters and secondary transporters in PBL2025 compared to P2 (SSO3012, SSO3045, SSO3046, SSO3043, SSO3047 and SSO3048) since the uptake of sugars is carried out *via* ABC and secondary transporters [244]. However, compared to strain P2, the average specific growth rate of PBL2025 was higher in 0.4% glucose

condition (as shown in Fig 3.2 in Chapter 3), 0.0106 ± 0.002 and 0.0214 ± 0.0001 h^{-1} for P2 and PBL2025 respectively. Correspondingly, an increase of glucose consumption (35.6mg/L compared to 40.2mg/L, calculation based on Fig 4.1) and a decrease of doubling time (32.5 ± 0.69 h compared to 65.4 ± 0.65 h) were also observed. The higher growth rate of PBL2025 might have resulted from the accelerating activities of some ABC transporters, since the uptake of glucose is mediated by ABC-transporters in *S. solfataricus* [245]. On the other hand, it might be due to the activation of some unknown transport systems relative to P2, as proposed in [70], where cellodextrin consumption in *S. solfataricus* 98/2 was investigated *via* a genetic mutants study. Furthermore, Lalithambika [70] found the inactivation of a putative ABC transporter involving SSO2847, SSO2848, SSO2849 and SSO2850 had no effect on the cell growth when grown on glucose. Accordingly, the author proposed the existence of additional transport systems in strain 98/2 relative to P2 [70]. The effect of growth media on cell biomass of *S. solfataricus* strains was also estimated. In glucose standard media, although PBL2025 grew faster, its biomass reduced up to 21% compared to P2 (at late exponential growth phase).

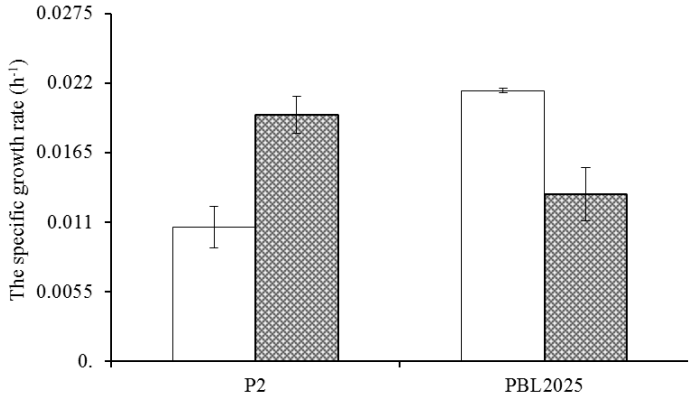

Fig 4.2 Specific growth rates of *S. solfataricus* P2 and PBL2025 strains grown on 0.2% (w/v) tryptone and 0.4% (w/v) glucose. A biological triplicate was performed for each condition.

4.4.1.2 *S. solfataricus* strains grown on tryptone compared to glucose

The specific growth rates of *S. solfataricus strains* grown on 0.2% tryptone and 0.4% glucose are shown in Fig 4.2. From Fig 4.2, we can see the average specific growth rates of P2 were 0.0106 ± 0.002 h^{-1} and 0.0195 ± 0.001 h^{-1} when grown on tryptone and glucose respectively, whilst the specific growth rates of PBL2025 were 0.0214 ± 0.0002 h^{-1} and 0.0132 ± 0.002 h^{-1} respectively. Correspondingly, there was a decrease of doubling time

of P2 grown on tryptone compared to glucose (35.5 h compared to 65.4 h). The faster growth of P2 on tryptone compared to glucose was reported in a previous study, where the oligopeptide binding proteins (encoded by *SSO1273* and *SSO2619*) of ABC transporters were strongly induced by the addition of tryptone in media [246]. This is also agreed with the observation of Elferink, *et al*., [244], who noticed the activity of glucose-binding proteins of ABC transporters in P2 showed the highest levels in tryptone compared to glucose. Transport proteins are supposed to catalyse the transfer of carbon sources across membrane into cell [247]. Above all, it might also contribute to the fast growth of *S. solfataricus* on tryptone.

Both *S. solfataricus* P2 and PBL2025 strains had a shorter lag phase and a faster growth rate in tryptone than in glucose media: 4 h and 12h differences for tryptone compared to glucose for P2 and PBL2025 respectively. Tryptone is a complex mixture of amino acids, and it is inferred that tryptone can be directly used as tricarbocitrate cycle (TCA) intermediates [248] for protein biosynthesis and also as intermediate metabolites for carbohydrate, lipid and nucleotide metabolism due to the close interaction of amino acid metabolism with glycolysis, gluconeogenesis and the citrate cycle [249]. Glucose is metabolized through the Entner-Doudoroff (ED) pathway with semi-phosphorylative and non-phosphorylative (np-ED) branches in *S. solfataricus* [221]. *S. solfataricus* can synthesise all 20 amino acids and utilize various amino acids as carbon sources [20]. TCA intermediates are used as precursors for cellular biosynthesis. The initial reaction step of TCA cycle is the adding of Acetyl-CoA to oxaloacetate to form citrate, which is catalyzed by the enzyme citrate synthase. Replenishment of oxaloacetate is necessary to keep TCA cycling [248]. The replenishment the oxaloacetate of the ED pathway is *via* carboxylation of pyruvate or phosphoenolpyruvate when *S. solfataricus* grown on glucose [39]. It may explain why the growths of both these two *S. solfataricus* strains on tryptone were faster than in glucose.

4.4.1.3 The effect of pH change of medium *S. solfataricus* growth

To test if the lower OD observed for all *S. solfataricus* strains (P2, PBL2025 and PBL2073) growth on tryptone compared to glucose might be a consequence of the observed alkalisation (pH change) of the growth medium, the pH of medium was measured against the growth at wavelength of OD_{650nm} and results are shown in Fig 4.3.

It can be seen that the change of pH for glucose and tryptone media is different, which may be caused by the metabolic pathway differences. All the strains were grown on an initial pH of 3.04, and it decreased down to 2.3 on 0.4% glucose media, but it increased significantly up to 6.9 for when 0.2% tryptone was provided as a sole carbon source.

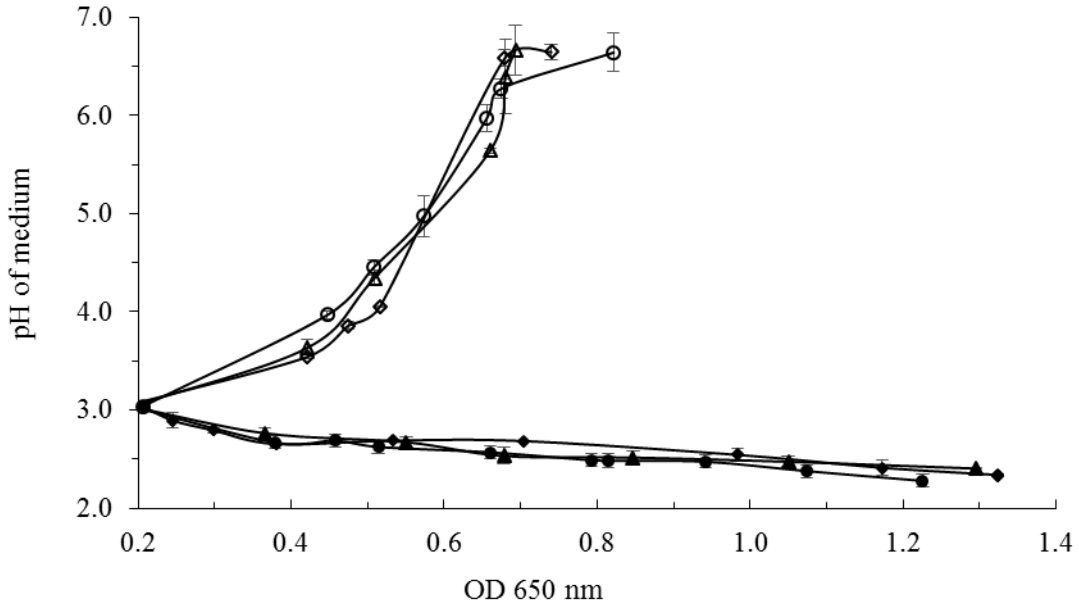

Fig 4.3 The pH change of medium against the cell growth.

The pH of decreases from 3.04 to 2.28 with cell growth from 0.20 to 1.23 for all strains grown on 0.4% glucose media (Fig 4.3). The minimum pH is similar: 2.35±0.03, 2.36±0.05, 2.41±0.03 for strain P2, PBL2025 and PBL2073, respectively. During the initial 20-24 h, the pH dropped from 3.04 to around 2.65 with the increase of cell density (OD_{650}) from 0.20 to 0.42, 0.32 and 0.65 for P2, PBL2025 and PBL2073 respectively. The maximum growth was achieved when no glucose was available in medium for consumption (it can be supported by the glucose consumption in Fig 4.1).

The pH of 0.2% tryptone growth medium increased with cell growth and the tolerant maximum pH for all strains is similar: 6.93±0.06, 6.89±0.02, 6.81±0.05 for P2, PBL2025 and PBL2073, respectively. The pH of tryptone medium increased from 3.04 to around 4.44 within 24 h with the increase of cell density (OD_{650}) from 0.20 to 0.44; and the maximum cell growth was achieved when the pH of medium reaches up to 6.5. The drop of OD_{650} was observed for all the strains when pH of growth medium reaches to the upper tolerance limit of cell membrane (pH 6.5-7.0). It is inferred that the low OD observed on

0.2% tryptone media can be contributed partly to the limit by up-taking of substrate since cell growth was observed again when the tryptone medium were acidified to 3.04 and it stops grow when pH of media reaches to around 6.5 (data not shown).

The effect of media on cell biomass of the *S. solfataricus* strains was also estimated. Both strains declined their biomass yields when grown on tryptone compared to glucose in stationary phases (0.41 ± 0.04 compared to 0.81 ± 0.01 g/L for P2, and 0.43 ± 0.05 compared to 0.64 ± 0.06 g/L for PBL2025 respectively). This was possibly caused by the pH change of medium that limit the uptake of carbon and energy source in tryptone compared to glucose.

4.4.2 Proteomic results

Many proteomics studies have been performed for *S. solfataricus* P2, however, comparable studies are still rare for *S. solfataricus* PBL2025. A previous study reported that changing of carbon source from tryptone to glucose significantly affected the protein phosphorylation level in *S. solfataricus* P2 [7]. However, this study focused on the phosphoproteome identification only, here we perform quantitative proteomics analysis for both two strains grown on either tryptone or glucose. Two sets of 8-plex iTRAQ experiments were carried out to investigate the responses of these *S. solfataricus* strains (PBL2025 *vs* P2) to different carbon sources (tryptone *vs* glucose) with details of each set iTRAQ shown in Tables 4.2A and B.

The numbers of identified and quantified proteins are shown in Table 4.2. A total of 2806 unique peptides corresponding to 740 proteins was identified (FDR < 1%) from the 1st iTRAQ experiment, whilst 2880 unique peptides and 702 proteins (FDR < 1%) were observed for the 2nd iTRAQ experiment, and 2194 unique peptides and 637 proteins for the 3rd iTRAQ experiment respectively. However, only proteins identified with more than 2 unique peptides were used for quantitation, resulting 609, 583 and 449 proteins quantified for the 1st, 2nd and 3rd iTRAQ experiments, respectively. Furthermore, t-tests were also performed to determine regulated proteins for each iTRAQ experiment [250]. The quantified protein number with abundance change from PBL2073 compared to PBL2025 grown on either glucose or tryptone were less than 100 (81 with the presence of glucose and 43 when tryptone was supplied), which are not enough for enrichment

test, so the following results are only discussed for comparisons between the 1st and 2nd set of iTRAQ experiments.

Table 4.2 Numbers of identified and quantified proteins from iTRAQ experiments.

Total predicted proteins[a)]	2994					
iTRAQ experiment	1st experiment				2nd experiment	
Proteins identified with ≥ 2 unique peptides[b)]	583				609	
Protein regulation	Number of proteins and their distribution (%)					
	PBL2025 *vs* P2 both grown on glucose		P2 grown on trypton *vs* glucose		PBL2025 grown on trypton *vs* glucose	
Up	97	16.6%	78	13.4%	110	18.1%
Unchanged	425	72.9%	424	72.7%	401	65.8%
Down	61	10.5%	81	13.9%	98	16.1%

a) She, *et al.*, 2001 and Esser, *et al*., 2011. b) More than 2 unique peptides and FDR < 0.01.

Moreover, to determine which cellular processes of *S. solfataricus* were significantly regulated under tryptone compared to glucose conditions, proteomics data were analysed based on a gene functional analysis (with p-value < 0.05, protein function categories and localization are download from arCOG (http://archaea.ucsc.edu/arcogs/)) [251]. Furthermore, the response of the CCM and its relationship with other significantly affected pathways was also determined based on the pathway enrichment analysis [243]. Details of these analyses are discussed below.

4.4.2.1 Characterisation, functional classification and localization of regulated proteins

Only proteins quantified with more than 2 unique peptides were used for classification and detailed results are shown in Supporting Information Table SP4.1A-C. Among these significantly regulated proteins, cytoplasmic proteins were predominantly observed (approximate of 50% for each phenotype comparison), followed by unknown proteins.

Furthermore, a numerical distribution of regulated proteins according to localization and function of cells is also shown in Supporting Information Table SP4.1D and E. Archaea generally possess a single cytoplasmic membrane and S-layer, but lack an outer periplasm, however, they contain proteins attached to the outside the cytoplasmic membrane, a region in between S-layer and a cytoplasmic membrane called pseudo-periplasm [152]. In our investigation, some pseudo-periplasm proteins were detected under different phenotype comparisons; however, these pseudo-periplasm proteins were much less than that of cytoplasmic membrane proteins. According to enrichment analysis, many processes such as amino acid biosynthesis and carbohydrate metabolism were inhibited under tryptone compared to glucose conditions for both strains. Furthermore, nucleotide metabolism and energy metabolism were also inhibited for tryptone compared to glucose (for P2) and tryptone compared to glucose (for PBL2025), respectively.

The differentially regulated proteins were classified according to their functional roles and implicated in metabolic pathways such as carbohydrate, energise, amino acid metabolic pathways, main cellular genetic information processing such as transcription and translation; results are shown in Supporting Information Table SP4.1D-E. In brief, differentially abundant proteins were classified as follows: comparison C1: 158 proteins were classified into 14 cellular processes for PBL2025 compared to P2 (both grown on glucose); comparison C2: 159 proteins involved in 15 processes for P2 grown on tryptone compared to glucose, and comparison C3: 208 proteins involved in 15 processes for PBL2025 grown on tryptone compared to glucose.

As expected, the largest group of detected proteins was hypothetical proteins for all comparisons, since hypothetical proteins are predicted at up to 40.3% of the whole *S. solfataricus* proteome [3]. The next largest groups were proteins involved in translation and energy metabolism. Furthermore, most of the regulated proteins observed in the translation group were ribosomal proteins and tRNA synthetases. Following this were a large number of proteins involved in amino acids biosynthesis (as showed in Fig 4.4 and Table SP4.3A-C in supplymentary information). These processes are linked together, for instance: metabolism of amino acids provides various metabolites and precursors for other cellular processes such as glycolysis, citrate cycle, pyruvate metabolism etc. and eventually for the biosynthetic requirements of new cell construction [249].

4.4.2.2 Gene function enrichment analysis

The differentially regulated proteins were assigned to different functional ontologies based on their gene functions (from arCOG). To provide different viewpoints of functional representation of responses of *S. solfataricus* to different carbon sources, proteomics data were analysed in terms of gene functional analysis with a p-value < 0.05 [251]. Regulated proteins were classified into 18 or 19 out of 26 arCOG functional categories in *S. solfataricus.* The results are depicted in Fig 4.4 and Fig 4.5. All detailed numbers can be seen in supplementary data Table SP4.3.

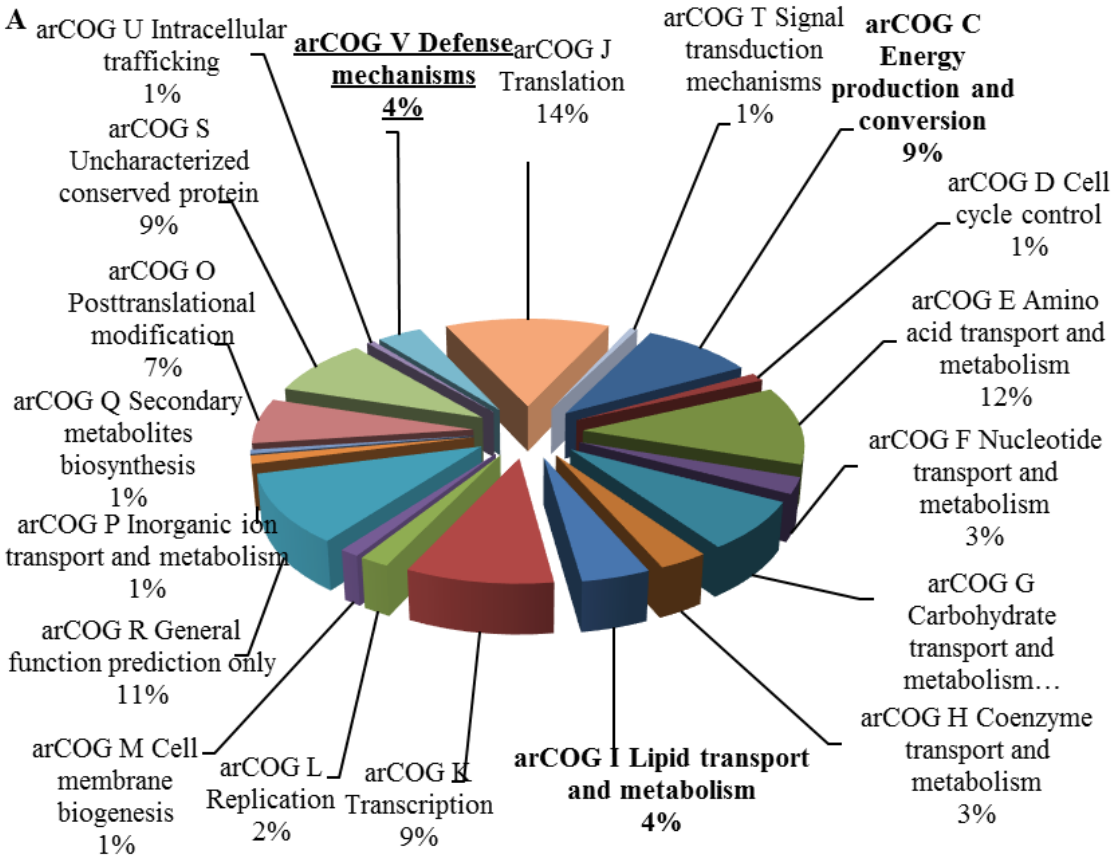

Fig 4.4 Classification of regulated proteins based on arCOG gene functions. (A) PBL2025 *vs* P2 both grown on glucose. Categoreis with font bond indicates changes are significantly based on enrichment test, and underline means down-regulation.

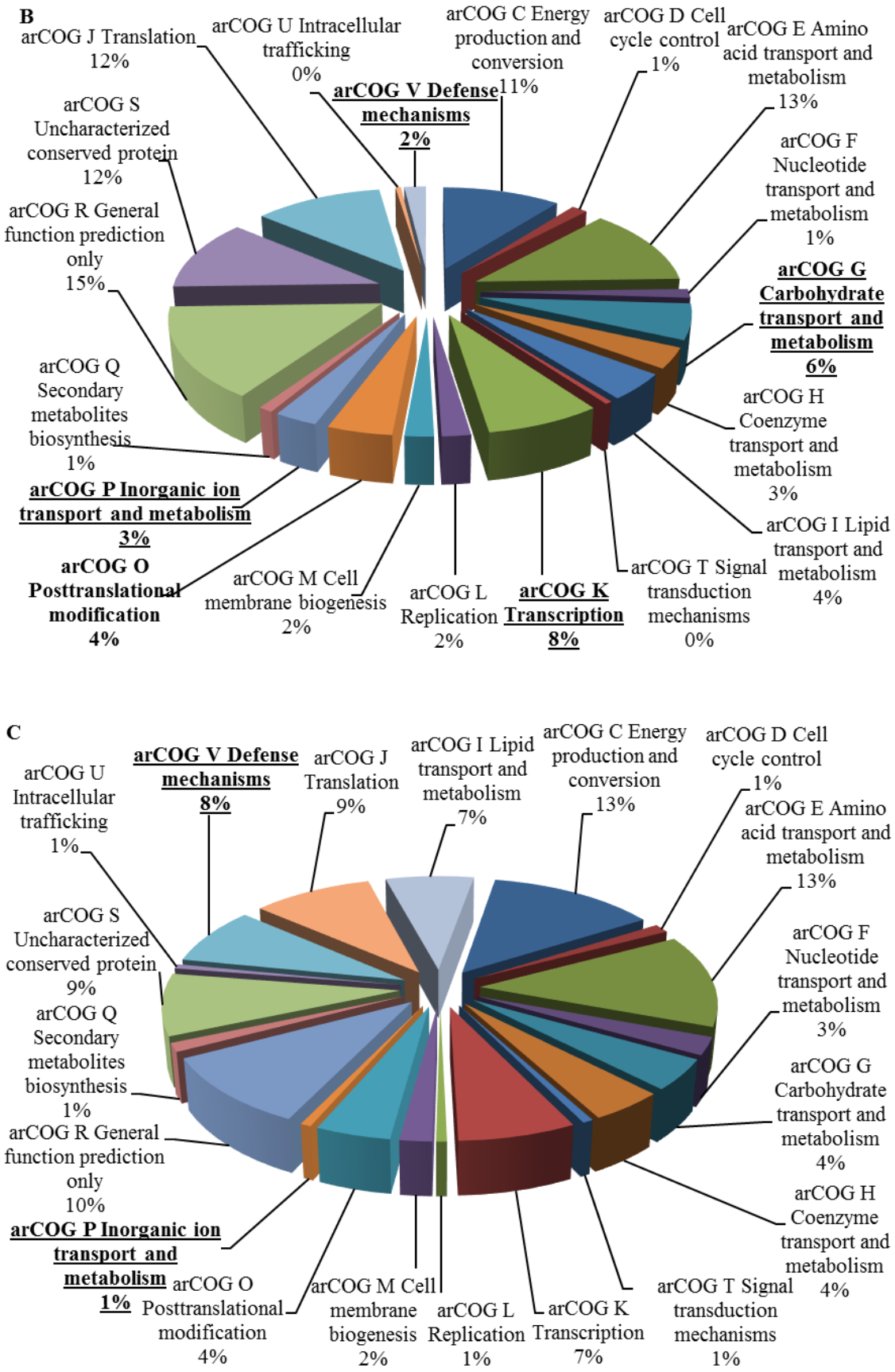

Fig 4.4 Classification of regulated proteins based on arCOG gene functions. (B) P2 grown on tryptone *vs* glucose, and (C) PBL2025 grown on tryptone *vs* glucose. Categoreis with font bond

indicates changes are significantly based on enrichment test, and underline means down-regulation.

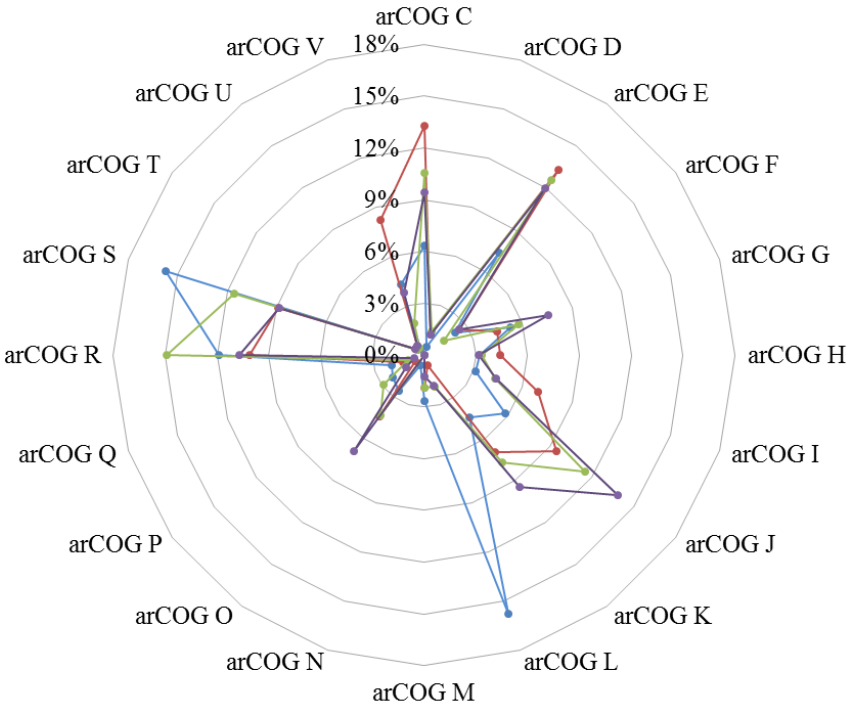

Fig 4.5. Spider chart of classification of regulated proteins based on arCOG gene functions. Blue, red, purple and green line are for predicted, PBL2025 *vs* P2 both grown on glucose, P2 grown on tryptone *vs* glucose, and PBL2025 grown on tryptone *vs* glucose.

Gene function categories identified with less than 5 quantified proteins and those with only one regulated gene are not considered for classification. For *S. solfataricus* PBL2025 compared to P2 grown on glucose, a lower relative abundance of energy production and conversion (arCOG C) and defence mechanism (arCOG V) categories were observed (red line in Fig 4.5), as well as the higher abundance of members of lipid transport and metabolism (arCOG I) was also noticed. Comparisons of the arCOG functional code profiles in the quantitative proteome under tryptone *vs* glucose conditions revealed obvious variations between strain P2 and PBL2025 (purple and green lines in Fig 4.5). By focusing on proteins classified into cellular metabolism (arCOG C, E-I, P-Q), enrichment analysis revealed that proteins of arCOG V (Defense mechanisms) and arCOG P (Inorganicion transport and metabolism) were down regulated in both comparisons (Fig 4.5).

Additionally, arCOGs carbohydrate transport and metabolism (arCOG G), and transcription (arCOG K) were down regulated in tryptone compared to glucose of P2 as depicted in Fig 3 (purple line). In contrast, only the down regulation of arCOG P and arCOG V were found in tryptone compared to glucose for PBL2025 (green line in Fig 4.5). Moreover, up regulated proteins belonging to arCOG O (posttranslational modification) was significantly observed in comparison of PBL2025 grown on tryptone *vs* glucose (green line in Fig 4.5). This agrees with observations reported by [7], in which a significant change of protein phosphorylation when provided different carbon sources in P2 was found [7]. Transport systems (carbohydrate, lipid and inorganicion transport) were significantly affected in tryptone media compared to glucose, as well as the defence mechanisms.

4.4.3 Carbohydrate transporter proteins

The uptake of glucose is mediated by ABC-transporters in *S. solfataricus* [245]. It has been discussed in two super-families according to its substrate binding activities: the carbohydrate and the di-/oligopeptide uptake transporter classes [252]. Different transporters show various substrate binding activities [252]. Transporting of various substrates impacts on different metabolism and eventually growth patterns. Most of the quantified transporters in this investigation belonged to the ABC transporters and they are listed in Supporting Information Table SP4.2.

A typical ABC transporter consists of 2 ATPases, 2 permeases and 1 membrane anchored substrate-binding protein [252]. In *S. solfataricus*, there are 37 putative ABC transporters [246] based on the genome sequence information and 15 binding proteins: whereas 6 belong to sugar class, 8 belong to oligopeptide class and 1 ungrouped [252]. The up/down-regulated ABC transporters detected here are listed in Table SP4.2. Furthermore, the affected transporters from previous studies where *S. solfataricus* grown on other carbohydrates are also summarized in Table SP4.2. The up regulation of oligo/dipeptide superfamily transporters were detected in both strains (in tryptone compared to glucose). When P2 was grown on tryptone compared to glucose, 10 proteins belonging to 7 ABC transporters were quantified, but only one dipeptide ABC transporter, periplasmic dipeptide binding protein (dppA, encoded by *SSO2619*) was up regulated and another ATP-binding protein of maltose ABC transporter (encoded by *SSO3055*) was down-regulated. However, 4 of 5 quantified ABC transporters including

SSO1003, SSO1275, SSO2619 and SSO3055 were up-regulated when PBL2025 was grown on tryptone compared to glucose. It can be hypothesised that the up regulation of SSO3055, SSO2619, permease protein (SSO1275) and sugar ABC transporter (SSO1003) could lead to the activation of these transporters, possibly, contributed to a shorter lag phase and a faster growth on tryptone compared to glucose mediated growth of PBL2025. These observations support the work carried out by Elferink, *et al*., [244], who found the expression of glucose-binding activity of the membrane-bound sugar-binding proteins reached highest level in cells grown on tryptone. It is also consistent with a previous study [246], where a strong expression of SSO2619 and SSO1273 (oligopeptide binding protein) under the addition of a peptide mixture (tryptone) condition was reported. Interestingly, the expression of SSO0999 (ABC-type trehalose transporter system) was not affected under tryptone conditions in both strains, which is consistent with the study reported by [246]. The abundance change of SSO0999 seemed to be affected by the supplement of various carbon sources. This protein showed its up-regulatetion under n-propanol [40] and ethanol [35, 41], and down-regulation under iso-propanol and phenol and no change under acetone [35] and tryptone conditions. Taken together, these may explain why the lag phase of PBL2025 and P2 were similar when both were grown on tryptone, despite the absence of some transporter encoding genes in PBL2025.

4.4.4 Effect of carbon sources on metabolism pathways

Regulated proteins were distributed into 53, 50 and 55 different pathways (a total of 77 pathways from KEGG: http://www.genome.jp/keeg/) for three comparisons (see Table 4.3 for details of comparisons), respectively. To gain a better understanding of how different *S. solfataricus* strains responded to different carbon sources, a relative frequency of the regulated enzymes in each pathway was calculated using the enrichment test (hypergeometric model) with a p-value < 0.05 [253] and the results are shown in Table 4.3. As listed in Table 4.3, the significantly affected pathways were involved in amino acids biosynthesis, carbohydrate metabolism, energy metabolism and nucleotide metabolism. Due to the specificity of archaeal metabolism and limit aviliability of characterized proteins in different KEGG pathways, only those pathways with proteins mostly characterized such as glycolysis pathway will be discussed.

Table 4.3 Summary of pathway enrichment analysis (hypergeometric model).

	Pathways	C1	C2	C3
Carbohydrate metabolism	Butanoate metabolism	+	None	None
	Glycolysis	None	_	+
	Pyruvate metabolism	None	_	None
Amino acids metabolism	Ala, asp and glu metabolism	_	+	_
	Arg and pro metabolism	_	_	None
	Val, leu and iso-leu biosynthesis	+	_	_
	Val, leu and iso-leu biodegradation	None	_	None
Energy metabolism	Carbon fixation pathways	+	None	None
	Oxidative phosphorylation	None	None	_
Nucleotide metabolism	Purine metabolism	None	_	None
	Pyrimidine metabolism	None	_	None

Note: C1: PBL2025 *vs* P2 both grown on glucose; C2: P2 grown on tryptone *vs* glucose; C3: PBL2025 grown on trypton *vs* glucose. +: Up regulation. -: Down regulation. None: change was not significant in a given pathway. A pathway was considered if more than 5 quantified proteins were observed for that pathway.

4.4.4.1 Responses of *S. solfataricus* PBL2025 vs P2 grown on glucose conditions

For PBL2025 *vs* P2, both grown on glucose, a proteomic analysis showed that the missing of genes *SSO3004-SSO3050* in PBL2025 likely resulted in different regulations of pathway metabolisms. According to the pathway enrichment test (as detailed in Table 4.3), proteins with an increased abundance were involved in the butanoate metabolism, Val, Leu and iso-Leu biosynthesis and carbon fixation pathway [254].

4.4.4.1.1 Regulation of central carbohydrate metabolism

52 proteins involving in CCM were quantified for comparison C1. The increased abundance of some enzymes involving in this pathway was reasonable. For instance, glucose is metabolised through Entner-Doudoroff pathway: semi-phosphorylative and non-phosphorylative ED (np-ED) branch [221]. The first few steps of np-ED pathway is conversion of glucose to glucono-1,5-lactone then to gluconate, catalysed by glucose 1-

dehydrogenase (gdh-1, -2, -3, encoded by *SSO3003*, *SSO3042* and *SSO3204* separately) and gluconolactonase (encoded by *SSO2705* and *SSO3041*separately) [39]. It is obvious to assume that the absence of *SSO3041* and *SSO3042* genes resulted in the accumulation and thus the requirement for higher abundance of gdh-1 (encoded by *SSO3003*), as indicated in Fig 4.6. However, although the substrate specificity of gdh-3 (SSO3204) in P2 has been shown to specific to glucose [255], its expression was the same level in PBL2025 (as shown in Fig 4.6). The pentose phosphate pathway enzymes also showed similar abundances, although PBL2025 lacks the *SSO3032* and *SSO3036* genes.

EPS metabolism was involved in in propanoate metabolic pathway based on KEGG pathway map. The increased relative abundances of proteins involving in propanoate metabolism included: Acetyl-CoA C-acetyltransferases (encoded by *SSO2061, SSO2062* and *SSO2625*), methylmalonate-semialdehyde dehydrogenase (SSO1218) and 4-aminobutyrate aminotransferase (SSO3211) for PBL2025 compared to P2 (both grown on glucose). The formation and amount of EPS differences between PBL2025 and P2 has been showed in a previous study [42]. Moreover, the complementation of *SSO3006* (α-mannosidase) and *SSO3019* genes confirmed their roles in EPS formation [43]. No expression of *SSO3006* and strong induced expression of *SSO3019* at transcriptional level were reported by using q-PCR upon surface attachment in P2 [42]. At the proteomic level, SSO3006 was quantified but its abundance was not different to that observed in P2, and SSO3019 was not detected in PBL2025 compared to P2 (both grown on glucose standard media).

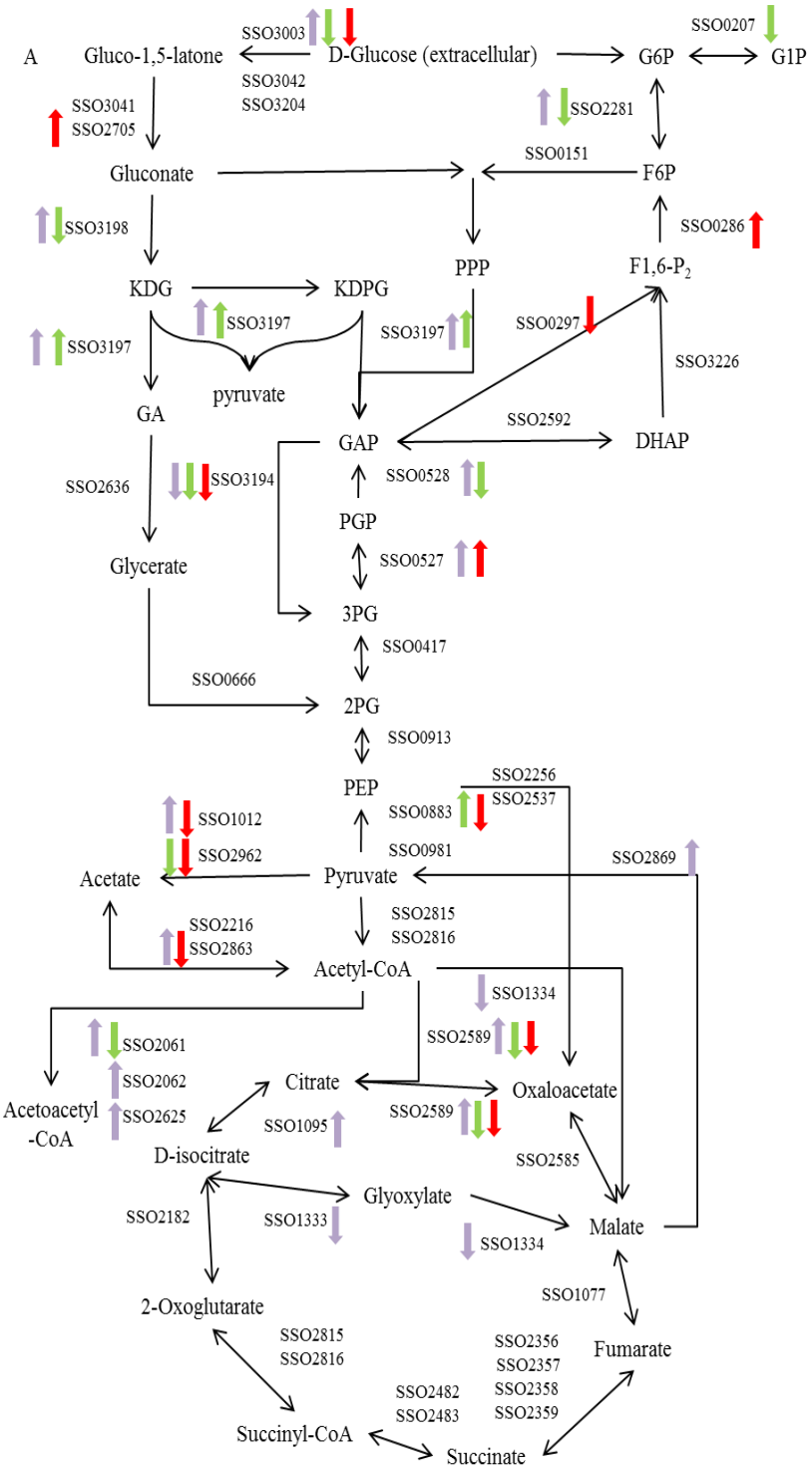

Fig 4.6 Regulated pathways for carbohydrate metabolism (Glycolysis, TCA and pyruvate metabolism). Comparison 1: PBL2025 *vs* P2 both grown on glucose (purple). Comparison 2: tryptone *vs* glucose of P2 (green); Comparison 3: tryptone *vs* glucose of PBL2025 (red). Direction arrows are for up-regulated and down-regulated proteins respectively. G1P: glucose 1-phosphate; G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; F1,6P: fructose 1,6-bisphosphate; DHAP: dihydroxyacetone 3-phosphate; KD(P)G: 2-keto-3-deoxy-(6-phospho)gluconate; GAP: glyceraldehyde 3-phosphate; PGP: 1,3-bisphosphoglycerate; 3-PG: 3-phosphoglycerate; 2-PG: 2-phosphoglycerate; GA: glyceraldehyde; PEP: posphoenolpyruvate; PPP: pentose phosphate pathway.

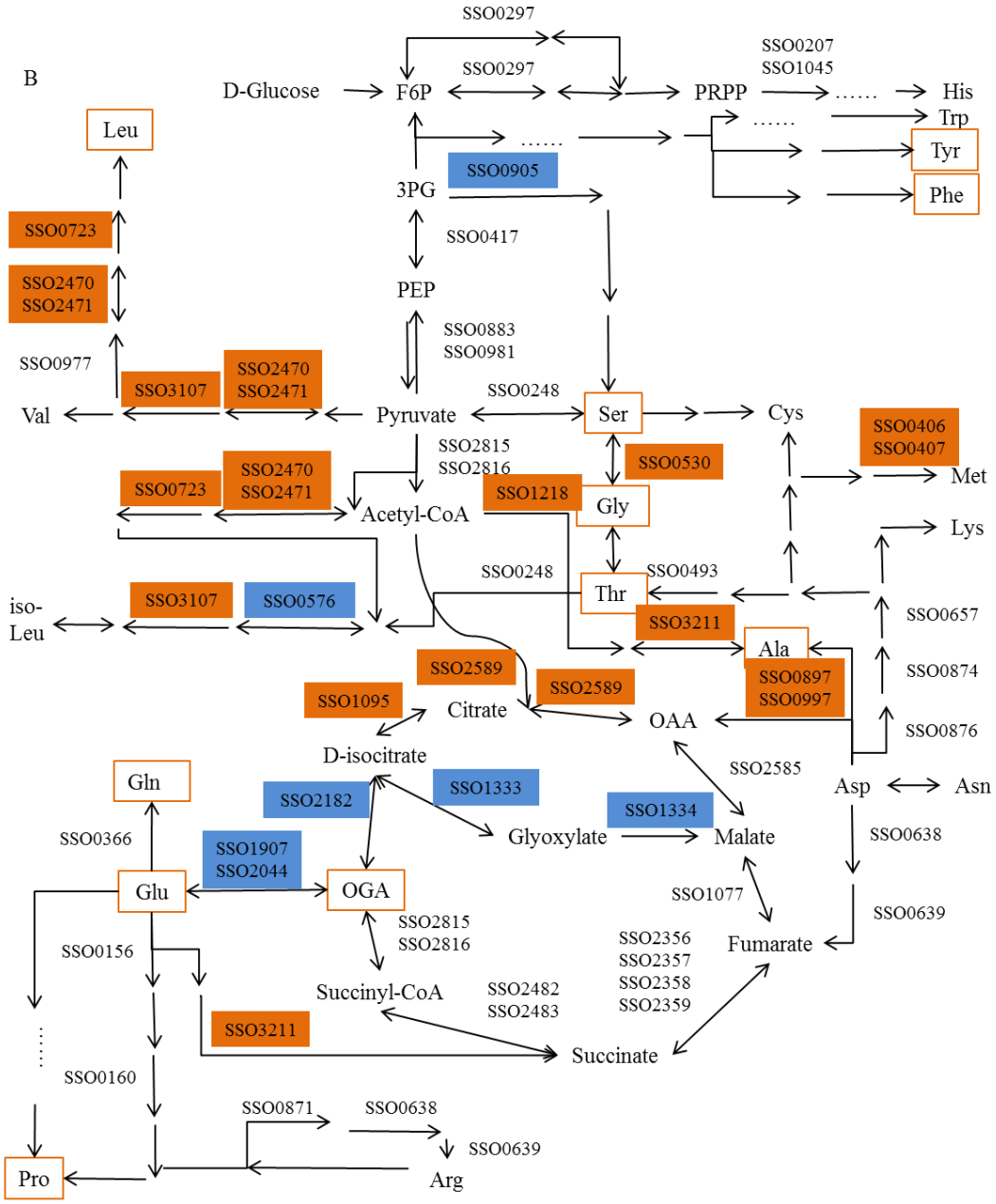

Fig 4.6 Regulated pathways for PBL2025 *vs* P2 both grown on glucose. (B) Biosynthesis of amino acids. Red and blue backgrounds are for up-regulated and down-regulated proteins respectively. Metaboloites with red cicle are up-regulated (data from Chapter 7). F6P: fructose 6-phosphate; 3-PG: 3-phosphoglycerate; PEP: posphoenolpyruvate; OAA: oxaloacetic acid; OGA: 2-oxoglutarate; PPP: pentose phosphate pathway.

In addition to changes involving in the CCM, the increased abundances of malate oxidoreductase (SSO3197) and 2-keto-3-deoxy gluconatealdolase (SSO2869) were found in PBL2025 compared to P2 (as shown in Fig 4.6A). *S. solfataricus* can synthesise

all 20 amino acids [20] and these amino acids can be used for protein biosynthesis. Biosynthesis of some amino acids was significantly different between PBL2025 and P2, as shown in Fig 4.6B. The up-regulation of these proteins might accelerate the utilization of pyruvate. The pyruvate production from malatate and KD(P)G catalysed separately by these enzymes seems to prefer to be utilized by the biosynthesis of Leu, Val and iso-Leu compared to other amino acid biosynthesis (as shown in Fig 4.6B). It can be pursued from the higher abundance of proteins involving in Leu, Val and iso-Leu metabolism, especially 3-isopropylmalate dehydratase subunits (encoded by *SSO2470* and *SSO2471*), dihydroxy-acid dehydratase (SSO3107), 3-isopropylmalate dehydrogenase (SSO0723) in PBL2015 compared to P2. Above all, more active amino acid pathways might contribute to the higher specific growth rate of PBL2025 compared to P2 when grown on glucose.

4.4.4.1.2 Energy metabolism

Energy produced from various metabolic pathways is used for cell growth [248]. Among the 21 quantified proteins involved in the CO_2 fixation pathway, 5 proteins were up-regulated in PBL2025 compared to P2 in glucose condition, leading this pathway could be considered as significantly affected. CO_2 fixation refers to the assimilation of CO_2 into cellular organic materials [256]. It sustains autotrophic growth and can be also used for energy conservation and as a sink for cycling of reduced electron carriers, as discussed by Bar-Even, *et al.*,[254]. Autotrophic CO_2 fixation has been attracting attention since the characterisation of three new CO_2 fixation pathways was made in the last decade [257]. As of today, six mechanisms of CO_2 assimilation have been identified, as reviewed in detail elsewhere [257]. Among them, the 3-hydroxypropionate-4-hydroxybutyrate cycle (3HP/4HB) was found in *Sulfolobales* [258], leading to the assimilation of bicarbonates. This cycle was initially reported in Crenarchaeota *Metallosphaera sedula* [258]. Although not all of the enzymes have been characterised, key genes in this pathway have been already found in *Sulfolobus* and other Crenarchaeota species genomes [258]. Usually, the 3HP/4HB cycle is divided into two parts: (1) the transformation of acetyl-CoA and two bicarbonate molecules into succinyl-CoA, and (2) the conversion of succinyl-CoA into two acetyl-CoA molecules [258].

The detection of up-regulated proteins involving in this pathway included acetyl-CoA C-acetyltransferase (acaB), acetyl-CoA synthetase (acsA) and aconitatehydratase. AcaB-2, acaB-3 and acaB-7 (encoded by *SSO2061, SSO2062* and *SSO2625* respectively) catalyse

the conversion of acetoacetyl-CoA to acetyl-CoA, while protein acsA (encoded by *SSO2863*) catalyses the formation of acetyl-CoA from acetate. In addition, aconitatehydratase (encoded by *SSO1095*) catalyses the conversion of isocitrate to cis-aconitate, then to citrate. Acetyl-CoA goes directly to pyruvate, propanoate metabolism as well as Val, Leu and Iso-Leu biosynthesis pathways (source: KEGG: http://www.genome.jp/keeg/). It can be inferred that the accumulation of acetyl-CoA involved in the regulation of intracellular biosynthesis metabolism, and further resulted in an activation of carbon fixation in PBL2025 and eventually a faster growth of this strain than P2.

4.4.4.2 Responses of *S. solfataricus* under tryptone *vs* glucose conditions

4.4.4.2.1 Regulation of central carbohydrate metabolism

52 and 51 proteins involved in CCM were quantified for comparisons C2 (P2 grown on glucose *vs* tryptone) and C3 (PBL2025 grown on glucose *vs* tryptone), respectively. The differential abundances of these proteins are shown in Figs 4.6A (based on an enrichment test). From Fig 4.6A, we can see that the down regulation of proteins in glycolysis and pyruvate metabolism was significantly in comparison C2 (green color). However, proteins involved in glycolysis showed an increased abundance but no abundance change in pyruvate metabolism in comparison C3 (red color). These differences might be explained by the 'missing' genes in PBL2025 compared to P2. It seemed that when cells were grown on tryptone, a high abundance of amino acids supplemented in the media resulted in an increase of pyruvate/TCA intermediates [248]. Accumulation of DL-2-methylglutamic acid and pyroglutamic acid were detected based on GC-MS (data from Chapter 6). Unsurprisingly, the TCA pathway remained unaffected for both comparisons, which agreed well with a previous study in *S. solfataricus* [39]. Also, it might explain why the growth of these two *S. solfataricus* strains on tryptone were faster than on glucose.

4.4.4.2.2 Regulation of amino acids metabolism

Catabolism of amino acids is different from their biosynthesis, this catabolism process involves breakdown to remove the α-amino groups and then the carbon skeletons including pyruvate, Acetyl-CoA, Acetoacetyl-CoA and TCA cycle intermediates (oxaloacetate, fumarate, succinyl CoA and α-ketoglutarate). These metabolites are directly used for carbohydrate, nucleotide metabolism and other cellular processes [248].

However, for both strains, of cells ceased growth at ca. 40 h when tryptone was supplied, and their biomasses were less than that obtained in the glucose condition. This was possibly due to the lack of available carbon sources for cell construction.

At the proteome level, the down regulation of amino acid biosynthesis pathways under the tryptone condition in both strains was reasonable since cells did not require the production of more amino acids to use as storage sources [259]. Conversely, it also meant that the protein synthesis process happened based on its functional properties rather than a need to store extra amino acids as mentioned elsewhere [259]. Furthermore, bacterial amino acid biosynthesis processes are regulated at different levels: through enzymatic activity or through metabolite flow regulation or, even, at the DNA level (enzyme formation or gene expression) regulation [260]. A negative-feed-back regulation mechanism has been proposed in those cases, which means that the end-product inhibits its own biosynthesis through affecting the biosynthetic enzyme activity (for a review see elsewhere [260]). In other words, the supplement of external amino acids interferes with their intracellular formation. Subsequently, when microorganisms resort to an exogenous supplement of amino acids, their intracellular biosynthetic pathways tend to be blocked or shut down. This was observed, for instance, for Val [261] and iso-Leu biosynthesis in *E. coli* [262]. A similar phenomenon has been found for anthranilate synthase in *S. solfataricus*, which catalyses the synthesis of anthranilate from chorismate and glutamine and is feedback-inhibited by tryptophan [263]. In the light of this, we can reasonably speculate that the down regulation of the detected amino acids metabolic pathways followed a similar pattern.

4.4.4.2.2.1 Regulation of Val, Leu and iso-Leu metabolism

Val, Leu and iso-Leu are hydrophobic branched-chain amino acids (BCAAs). Significant down regulation of BCAAs metabolism were found in comparisons tryptone *vs* glucose comparisons in both P2 and PBL2025 strains. BCAAs have been investigated substantially in humans due to their important biological roles, such as the regulation of protein synthesis in a variety of tissues, the role in secondary structure of proteins, nutrient signaling as reviewed elsewhere [264]. The first common step of their catabolism involves the BCCA aminotransferase and keto acid dehydrogenase. Eventually, they are converted to succinyl-CoA or acetyl-CoA, which enters into pyruvate metabolism and TCA cycle respectively. From a biosynthetic aspect, the last four reactions are catalysed

by the same four enzymes in the biosynthesis of Val, Leu and iso-Leu (Fig 4.7), therefore, they are mostly discussed synchronically [260]. Enzymes with multiple substrates such as enzymes encoded by *SSO2470*, *SSO2471* and *SSO0576* are easily affected by the existence of any of BCCAs substrates (as shown in Figure 4.7). The substrates of 3-isopropylmalate dehydratase subunits (encoded by *SSO2470* and *SSO2471*) are 2-methylamate, D-erythro-3-methylamate, (2S)-2-Isopropylmalate, 3-Isopropylmalate and (2R, 3S)-3-Isopropylmalate, where (S)-2-Aceto-2-hydroxybutanoate, (R)-3-Hydroxy-3-methyl-2-oxopentanoate, (R)-2,3-Dihydroxy-3-methylpentanoate can be catalyzed by ketol-acid reductoisomerase (SSO0576). As speculated above, the down regulation of Val, Leu and iso-Leu biosynthesis metabolism may follow a feedback-inhibited pattern in *S. solfataricus* [263].

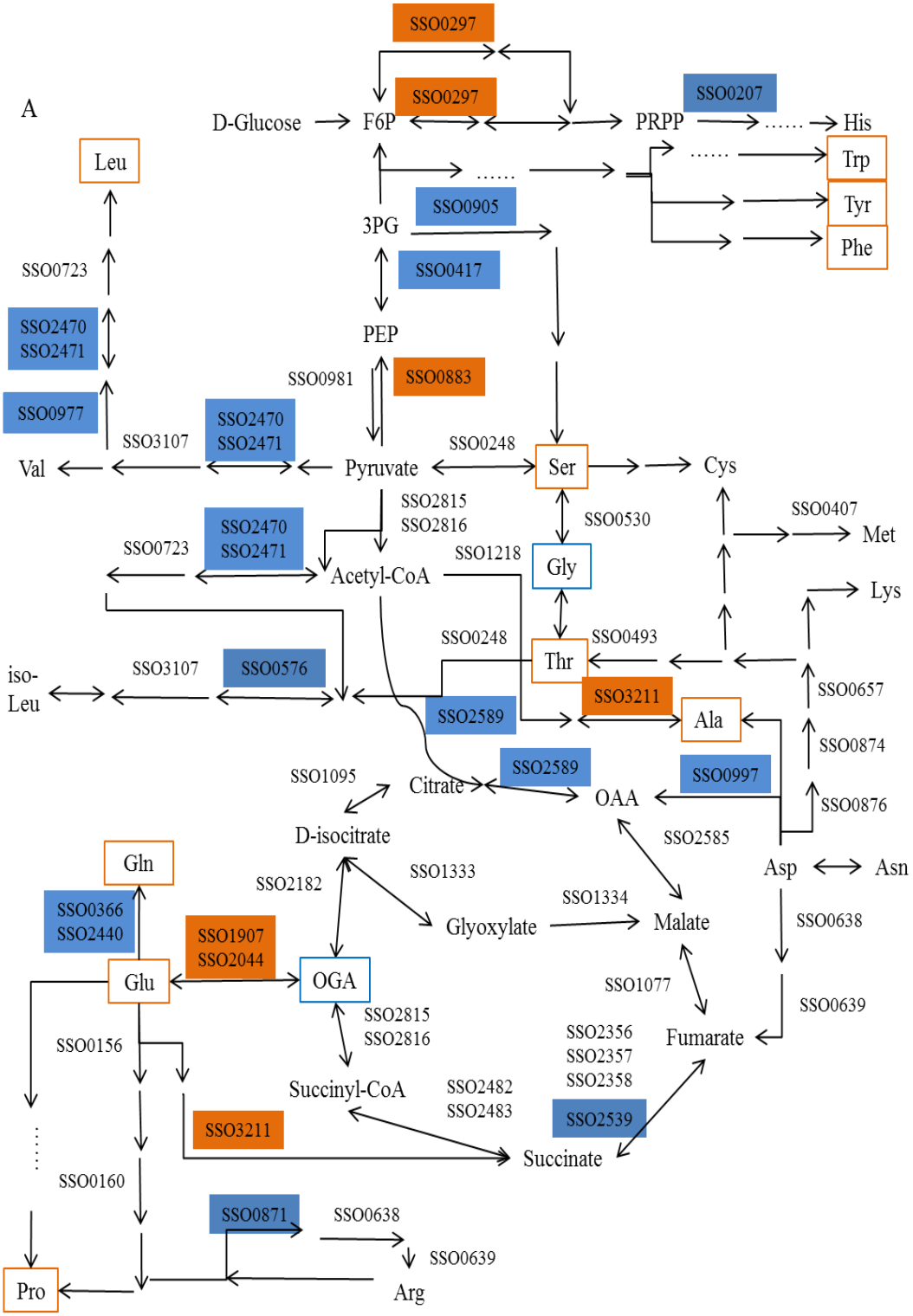

Fig 4.7 The amino acids metabolism regulation of *S. soltataricus* in P2 for tryptone *vs* glucose. Red, blue, and white background are for up-regulated, down-regulated, un-changed proteins respectively. Orange and blue cycle, and underline is for up-regulated, down-regulated, un-changed metabolites. F6P: fructose 6-phosphate; 3-PG: 3-phosphoglycerate; PEP: posphoenolpyruvate; OAA: oxaloacetic acid; OGA: 2-oxoglutarate.

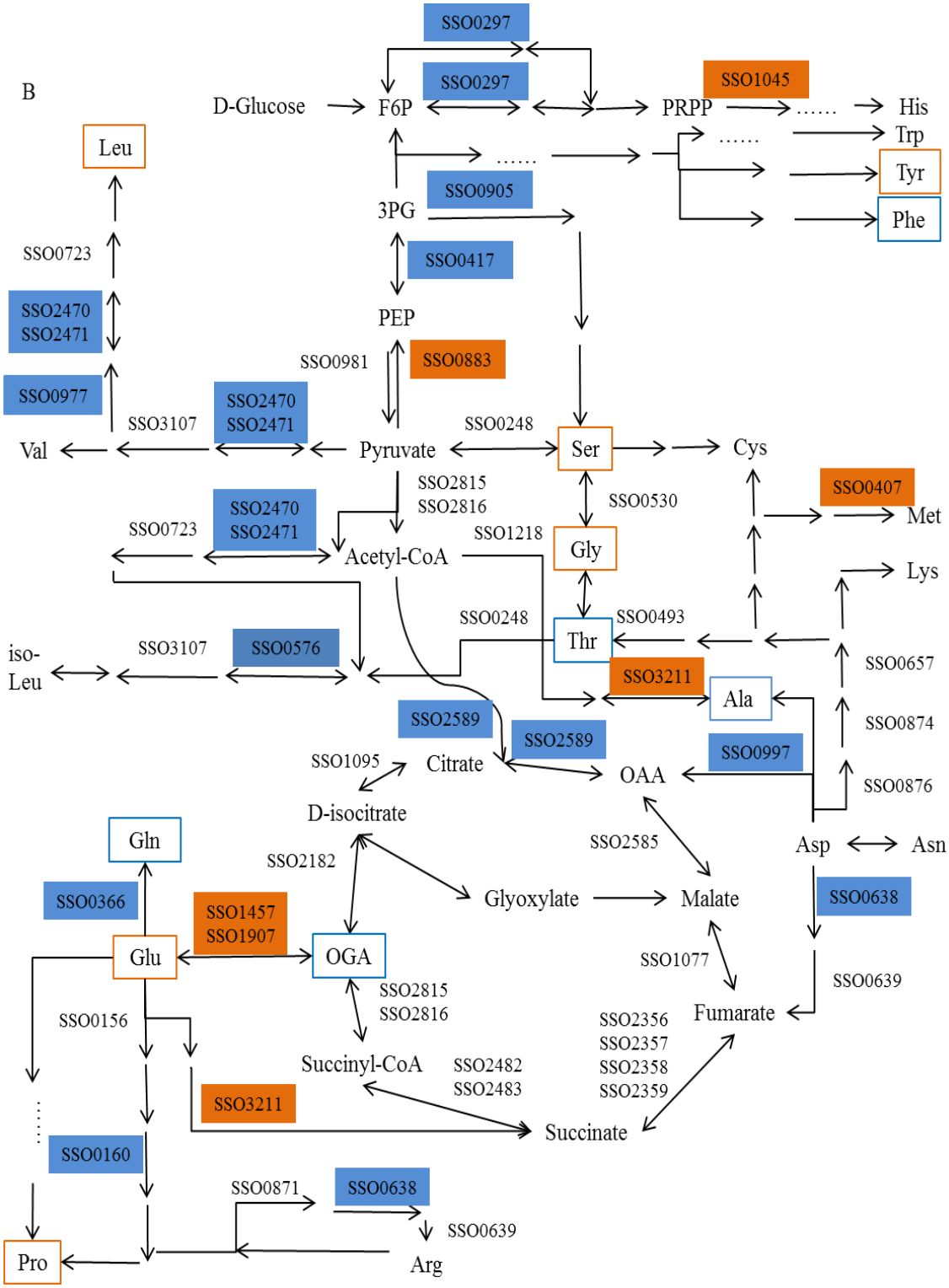

Fig 4.7 The amino acids metabolism regulation of *S. soltataricus* in PBL2025 (B) for tryptone *vs* glucose. Red, blue, and white background are for up-regulated, down-regulated, un-changed proteins respectively. Orange and blue cycle, and underline is for up-regulated, down-regulated, un-changed metabolites. F6P: fructose 6-phosphate; 3-PG: 3-phosphoglycerate; PEP: posphoenolpyruvate; OAA: oxaloacetic acid; OGA: 2-oxoglutarate; PPP: pentose phosphate pathway.

4.4.4.2.2.1 Regulation of Ala, Asp and Glu metabolism

For the tryptone *vs* glucose comparisons, most of the detected proteins were down regulated; only 0.5% and 0.8% of the quantified proteins were up regulated for P2 and PBL2025 respectively; these proteins are involved in Ala, Asp and Glu metabolism in P2, and Arg and Pro metabolism in PBL2025, as shown in Fig 4.8. These observations agree with a previous study, where certain concentrations of L-Glu and L-Asp enhanced cell growth by the presence of glucose, whereas Gly, L-Leu, L-Val showed no effect on cell growth [265].

The growth-stimulation effect of L-Glu in the presence of glucose was significant compared to other amino acids [265]. Glu plays a central role in various metabolic processes, as well as stress responses in bacteria as reviewed elsewhere [266]. Also, it has been reported that Glu-, Arg- and Lys-dependent acid resistance systems play an important role in protecting against acidic [267] and oxidative stress [268] in *E. coli*. Similar functions of Arg and Lys have also been found in bacteria *Salmonella typhimurium* CECT 443, but not for Glu [269]. Although the enzymetic properties may be different between *E. coli* and *S. solfatarics* due to typical living conditions for the latter organism, it could provide some hints for functional elucidiation of archaeal enzyme.

Glu catabolism is mainly carried out *via* the glutamate dehydrogenase (GDH) or glutamate decarboxylase (GAD) in bacteria [266]. Extracellular Glu also enters directly into the TCA cycle through 4-aminobutyrate aminotransferase (SSO3211). The up regulation of NAD specific GDH (gdhA-1, gdhA-2 and gdhA-4, encoded by *SSO1457*, *SSO1907* and *SSO2044*), which catalyses the revisable oxidative deamination of Glu to produce 2-oxoglutarate and ammonia with reduction of NAD^+, and SSO3211 were found, as shown in Figs 4.8 (A and B). This aspect agrees well with a previous study, which showed a large abundance of GDH in *S. solfataricus* [270].

GAD catalyses the decarboxylation of Glu to 4-aminobutanoate that leads to the further production of succinate semi-aldehyde in bacteria [266]. Furthermore, GAD systems were found to be especially related with resistance against acidic conditions [266]. Purification and characterisation of archaeal GAD were found in *Pyrococcus horikoshii* [271], *Thermococcus kodakaraensis* KOD1 [272] and *Pyrococcus furiosus* [273].

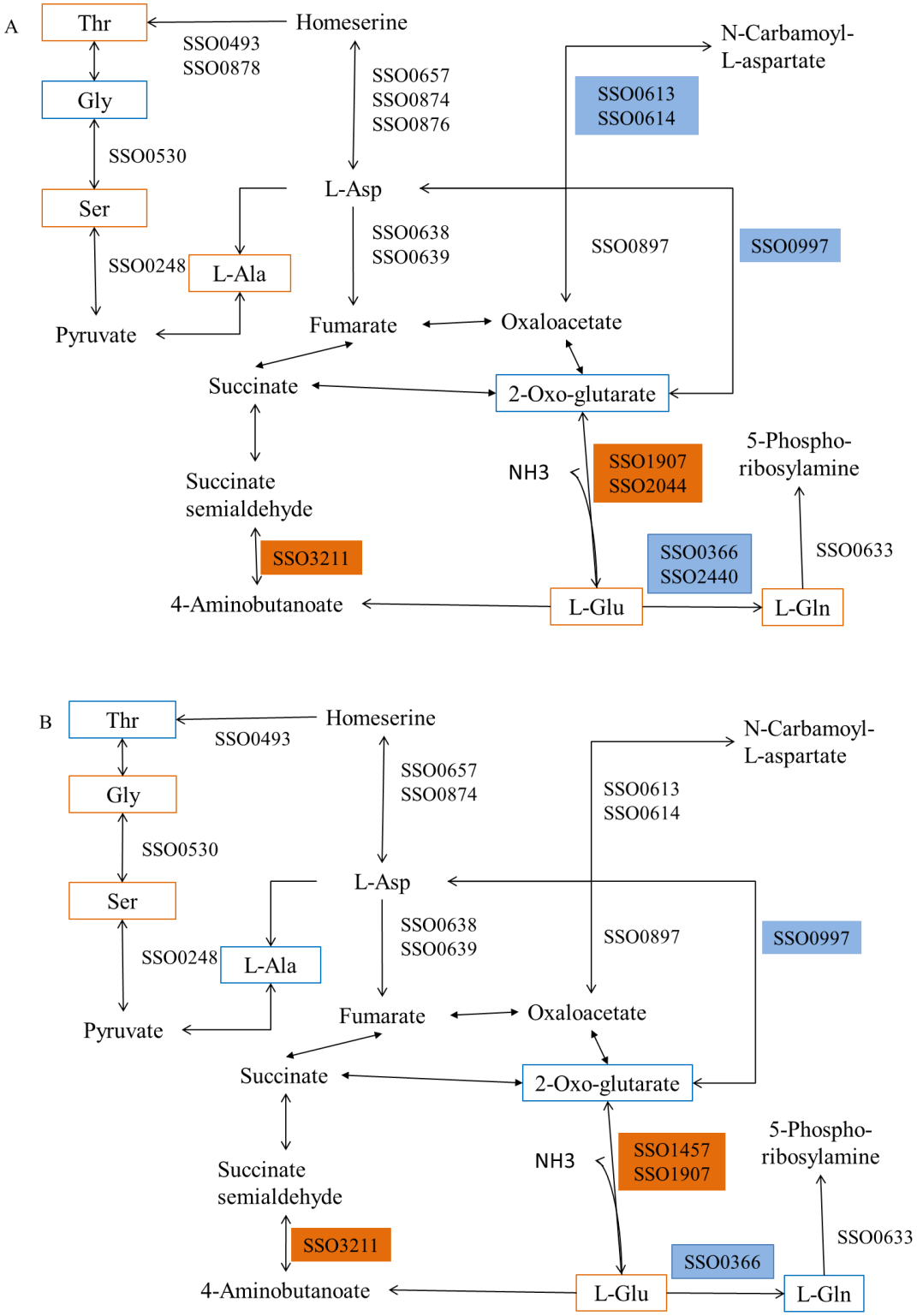

Fig 4.8 Regulations of Ala, Asp and Glu metabolism in *S. soltataricus* grown on tryptone *vs* glucose for strain P2 (A) and PBL2025 (B). Red, blue, and white background are for up-regulated, down-regulated, un-changed proteins respectively. Orange and blue cycle, and underline is for up-regulated, down-regulated, un-changed metabolites.

Moreover, down regulation of Gln synthetase (encoded by *SSO0366* and *SSO2440*) was observed in this investigation. Gln synthetase is supposed to catalyse the synthesis of Gln from Glu and NH_3 [274]. Its catalytic activity is solely regulated by end-product inhibition mechanism through Gly and Ala in *S. acidocaldarius* [275]. High abundance of Ala and low abundance of Gly has been detected for P2 grown under tryptone compared to glucose conditions (data in Chapter 7). The regulation of Gln synthetase in *S. solfataricus* may indicate follow the end-product inhibition mechanism (Fig 4.8).

In summary, the down regulation of most of the detected amino acid metabolic pathways may explain the reduced biomass (P2 and PBL2025) in tryptone as compared to compared to glucose-grown cells.

4.4.4.2.3 Energy metabolism

Under tryptone *vs* glucose conditions, no differentially-regulated proteins involved in energy-related pathways were found in P2; however, statistically significant down regulation of oxidative phosphorylation proteins (based pathway enrichment analysis) were observed in PBL2025 as a result of energetic purposes.

For tryptone growth compared to glucose for PBL2025, 30 proteins involving in energy metabolism were found to be significantly regulated (13 up and 17 down regulated). Usually, NADH and succinate generated from TCA cycle are oxidized and release energy to power the ATP synthase, subsequently ATP provides energy to other cellular process [276]. In our investigation, inorganic pyrophosphatase (SSO2390) and three NADH dehydrogenases subunits C, D and I (SSO0323, SSO0324, and SSO0326) involving in the oxidative phosphorylation pathway were down regulated, but the ATP synthase (SSO0559, SSO0561, SSO0563, SSO0564 and SSO0566) were unchanged. It can be speculated that up-regulation of proteins involved in glycolysis was activated for energy requirements. Ultimately, these regulations supported the shorter lag phase of PBL2025 grown in tryptone media compared to glucose.

4.5 Conclusions

S. solfataricus PBL2025 exhibited a longer lag phase when grown on glucose standard media, possibly due to the missing genes (absence of genes *SSO304-SSO3050* compared to P2). Both *S. solfataricus* P2 and PBL2025 strains could utilise tryptone as a sole carbon

source and demonstrated fast growth rates on tryptone compared to glucose conditions (1.25 and 1.62 folds for P2 and PBL2025 respectively). However, maximum cell biomasses for both *S. solfataricus* P2 and PBL2025 strains grown on tryptone were less than those grown on glucose.

At the proteome level, almost a third of the quantified proteins (detected with ≥ 2 MS/MS spectra) were regulated in tryptone compared to the glucose conditions (27.3% of 583 proteins for *S. solfataricus* P2 and 34.2% of 608 proteins for *S. solfataricus* PBL2025). Among these proteins, 83 quantified proteins overlapped between tryptone compared to glucose conditions (for both strains). Categorization of regulated proteins based on the new arCOGs genome annotation showed that 18 and 19 of 26 functional categories were affected when both *S. solfataricus* P2 and PBL2025 strains were grown on tryptone compared to glucose conditions. A pathway enrichment test was applied to determine regulated pathways based on the iTRAQ data. It has been shown that most cellular processes, especially amino acid metabolism pathways in *S. solfataricus* grown on tryptone, decreased in abundance compared to glucose conditions (for both *S. solfataricus* P2 and PBL2025), and down regulation of glycolysis and pyruvate metabolism was also found for *S. solfataricus* P2. This observation supports a previous study reporting sugar degradation was blocked in the presence of tryptone in P2 [7] However, the glycolysis pathway seemed to be accelerated in the *S. solfataricus* PBL2025 strain grown on tryptone compared to glucose. The reasons for these differences are not clear yet, further quantitative phosphoproteomic studies are in progress. Up and down regulation of proteins involved in sp-ED and gluconeogenesis were also observed, but these proteins did not show significant differential regulation. This study provides useful quantitative proteomics data for metabolic pathways manipulation of the model archaea *S. solfataricus* P2 and PBL2025.

Chapter 5 Evaluation of phosphopeptide enrichment strategies for global phosphoproteomic study of *Sulfolobus solfataricus* P2

5.1 Abstract

In terms of Phosphopeptide (P-peptide) characterisation, enrichment is a bottleneck for global phosphoproteomic studies. P-peptide enrichment is a critical step for proteomic analysis of *Sulfolobus solfataricus* due to low stoichiometric phosphorylation as well as low abundance of phosphorylation (P-proteins) in biological samples. Optimisation of the P-peptide enrichment process was performed to minimise the non-P-peptides binding to achieve high enrichment efficiency and sensitivity. Different affinity resins including three different TiO_2 formats (spin tips, magnetic beads and pipette tips) and PHOS-Select Iron Affinity Gel (IMAC-Fe) were tested alone and in combination of both. Different acids were included in the sample loading buffer to evaluate their ability to act as non-P-peptide excluders during P-peptide enrichment using TiO_2 beads. Optimisation was also performed for peptide-to-beads ratio, consecutive incubations with TiO_2 beads. Different elution buffers were tested for IMAC. In addition, 500 µg of tryptic digest cell lysates of *S. solfataricus* grown on standard glucose media was used for each experiment.

The key finding was that TiO_2 beads showed the best performance compared to pipette and spin tips for P-peptides enrichment. The use of non-P-peptide excluders (glycolic and lactic acid) in loading buffer improved P-peptides enrichment efficiency using TiO_2 beads. Furthermore, a 20/1 (w/v) peptide-to-TiO_2 beads ratio and 4 cycles of incubation of beads were found to be the optimum conditions for P-peptide enrichment for this system.

The use of elution buffer (ammonium in ACN) together with MOAC-TiO_2 resulted in a high number of detected P-peptides compared to the sole use of ammonium. The number of P-proteins identified from modified SIMAC and TiO_2 techniques were complementary and the overlap was only 13.2 %. Therefore, a combined use of modified SIMAC and 4

cycles of continuous incubation of TiO_2 beads were the optimal condition for P-peptide enrichment here.

5.2 Introduction

The acidic thermophilic archaeon *Sulfolobus solfataricus* P2 has been used as a model organism in archaeal research. It grows typically at 80 ^{o}C and low pH around 2-3. *S. solfataricus* can use different compounds as carbon source such as D-glucose and tryptone. Recently, a high number of P-proteins (540 proteins) was detected for this archaeon in a report studying the effect of different carbon sources [7]. However, a global quantitative phosphoproteomic study has not been performed for archaea yet.

P-peptide enrichment is a crucial step for phosphoproteomic analysis. In addition to issues of abundance and stoichiometry, there are further technical considerations for phosphoproteomic analysis using MS-based techniques: such as ionization suppression phenomenon, dynamic range as well as poor fragmentation using collision induced dissociation techniques. Low sensitivity of P-peptide detection in positive ionization conditions results from the negative charge carried on the phosphate group.

The development of affinity-based P-peptide enrichment enables an efficient quantitative phosphoproteomics study, especially the widely use of immobilized metal ion affinity chromatography (IMAC), TiO_2 and a combined use of both as reviewed elsewhere [144, 145, 164]. However, both of these workflows have advantages and disadvantages. The application of IMAC is hampered by the non-specific binding of acidic peptides, which is caused by the affinity between the carboxylate group of amino acid residues and IMAC materials.

IMAC results in selective binding of multiple P-peptides than mono-P-peptides [157]. Improvements of IMAC applications for p-protein/P-peptide enrichment are still ongoing, as reviewed in Chapter 2. For instance, the test of different materials for efficient binding and various support resins, the optimisation of enrichment protocols to reduce non-specific binding especially acidic peptides, as well as efficient P-peptides elution [164]. Furthermore, the application of tandem IMAC-IMAC and a combined use of different ions IMAC have proved to be superior to the one step IMAC enrichment [164].

Commercially available magnetic TiO_2 beads have been widely used by the community because of their high selectivity and specificity. Various acids in loading buffer used as non-P-peptide excluders such as 2,5-DHB (2,5-dihydroxybenzoic acid), phthalic acid [174], glycolic acid [157], lactic acid [277], and citric acid [176] have been found to significantly reduce the unspecific bindings of non-P-peptides. Glycolic acid or lactic acid was recommended due to its high adsorption efficiency for TiO_2 [157]. However, Sugiyama, *et al*., [277] and Aryal and Ross [278] observed a high non-specific binding by using glycolic acid. Phthalic acid and 2.5-DHB were not applied here to avoid a potential contamination of a LC system and the inlet of MS as proposed by [157] and reviewed by [279].

Other factors affecting P-peptide enrichment by TiO_2 beads are the ratio of peptide-to-beads, incubation time and consecutive incubation were investigated by [179] and [180]. The optimum peptide to beads ratio ranged from 1:2 to 1:8 (mass to mass) was proposed for P-peptide enrichment from HeLa cells. The author recommends to test pre-experiments on the optimal peptide/beads ratio and incubation cycles, since this will potentially vary between cell types [179].

In contrast to IMAC, the advantage of TiO_2 lies in its high selectivity for P-peptides as well as its robustness and tolerance towards commonly used detergents, salts, and reagents for biological analysis [157]. Therefore, to improve phosphoproteome coverage, especially to increase the identification of multiple P-peptides, a combination of gallium-IMAC and TiO_2-MOAC enrichment, termed SIMAC (sequential elution from IMAC) was performed by [166]. However, the enrichment of P-peptides is still a challenge, and the combination of different techniques seems necessity at this stage [184].

The purpose of this study is to obtain an optimised strategy for P-peptides/proteins detection (including sample preparation, peptide enrichment, MS operation, and bioinformatics approaches), for application to quantitative phosphoproteomic analysis of *S. solfataricus* P2 (Chapter 7) to compare samples grown on either glucose or tryptone. To achieve this aim, we tested various TiO_2 formats (magnetic beads, spin tips and off-column tips) and iron gel based IMAC, as well as the combination of both techniques (SIMAC) to evaluate the efficiency of P-peptides enrichment procedures. This evaluation

will be based on the comparing different forms of materials, the robustness of technique (estimated by detection number of P-peptides), selectivity of P-peptides (estimate by phosphorylation ratio of P-peptides: total peptides) as described [179], and easy-to-operate for sample handling.

5.3 Materials and methods

5.3.1 Cell growth, protein extraction and trypsin digestion

Cell collection of *S. solfataricus* strain P2, cell lysis, protein extraction and trypsin digestion were the same as described in Chapter 3. Briefly, cells were collected in the late-exponential growth phase by centrifugation at 5,000 *g* for 10 min at 4°C. Cell pellets were then stored at -80°C until required and protein extraction were carried out as detailed in Chapter 3.

To improve enrichment efficiency, the resultant peptides were desalted by Discovery DSC-18 columns (according to the manufacturer's instructions) before being dried by vacuum concentration (Eppendorf Concentrator 5301, Germany). Dried peptides were stored at -80°C for further analysis.

5.3.2 P-peptides enrichment

Limit of detection (LoD) of P-peptides was investigated using β-casein (data shown in Chap 3). Based on the LoD curve, 500 µg proteins were used as initial materials, and biological duplicates or triplicates were carried out for each experiment. For PHOS-Select Iron Affinity Gel experiments. Procedures were performed as manufacturer's instructions (Sigma Aldrich, UK). Optimal elution conditions were tested for 400 mM ammonia with or without the addition of 30% ACN, as indicated in Fig 5.1 A.

Different formats of TiO_2 including magnetic TiO_2 beads (GE Healthcare TiO_2 Mag Sepharose, UK), Mono Tip TiO (GL Sciences) and the titan spherephos-TiO kit spin tip (Hichrom Limited, UK) were tested. For magnetic beads, besides the recommend glycolic acid, different acids including lactic acid, citric acid and glutamic acid were tested as non-P-peptide excluders. In addition, 125 mM asparagine (NQ-Asp) and glutamine (NQ-Glu) was added before a washing step when lactic acid was used [178]. Loading, washing and elution steps were performed as recommended from the

manufacturers' documentation and solvents were prepared as listed Table TS 5.1 in the Appendix.

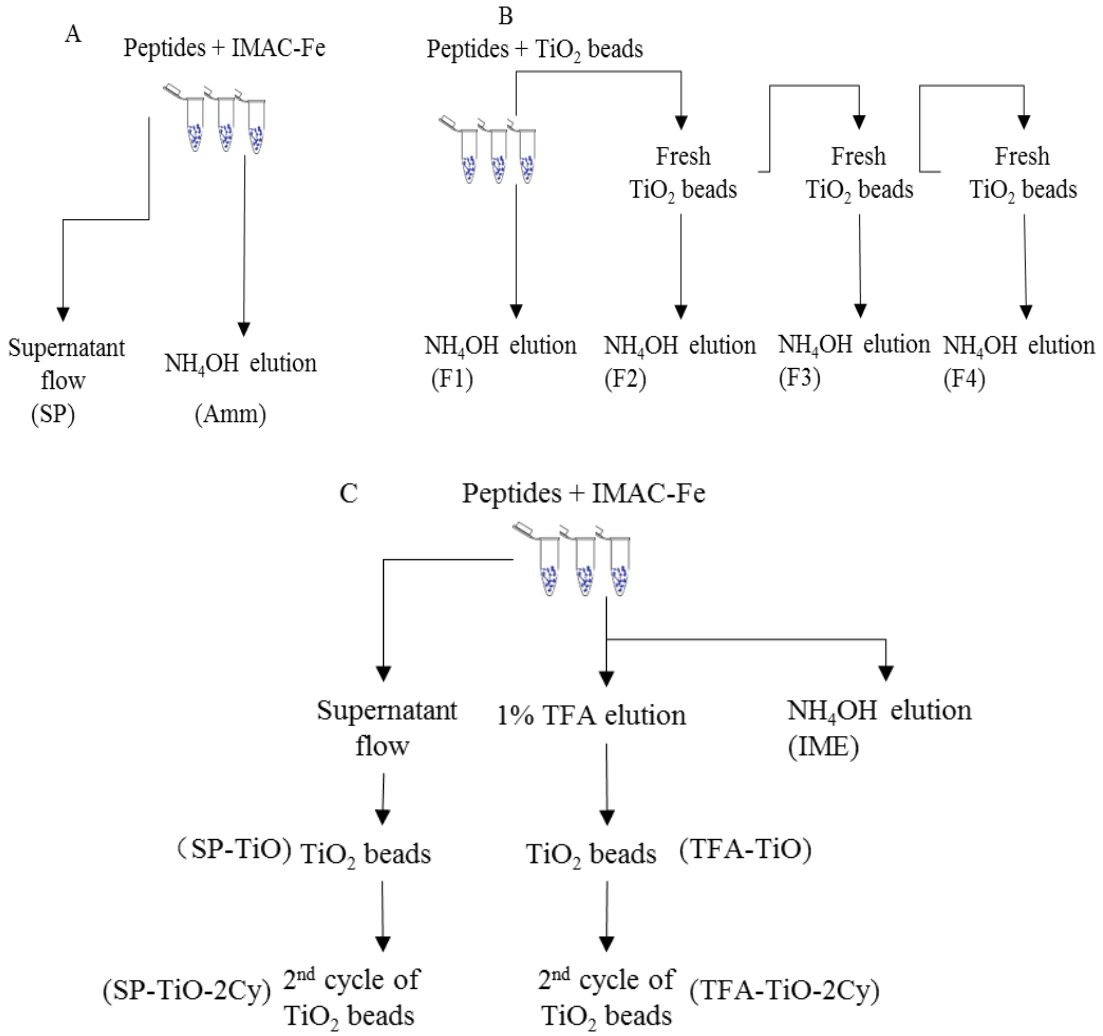

Fig 5.1 Flow chart of P-peptides enrichment of IMAC (A), consecutive TiO_2 beads (B) and modified SIMAC (C). F1 to F4 is the set of elutions using consecutive TiO_2 enrichment.

Different ratios of peptides/ TiO_2 (w/v of 80/1, 40/1, 20/1 and 10/1) and 4 cycles consecutive incubations (F1-F4 as shown in Fig 5.1B) were tested for magnetic TiO_2 beads. Briefly, digested cell lysates (500 µg) were resuspended in loading buffer: 1 M glycolic acid to prevent binding of non-P-peptides with 80% acetonitrile (ACN) and 5% TFA mixed with magnetic TiO_2 beads (12.5, 25, 50 and 100 µl of TiO_2 beads slurry) respectively and vortex for 30 min. After that, samples were washed with loading buffer once and washing buffer (80% ACN, with 1% TFA) twice. Finally, P-peptides were incubated with 50 µl of 5% ammonium hydroxide pH 12 for 5 min, and eluted. Vacuum dried samples were kept at -80°C.

Peptides were also subjected to SIMAC enrichment as described in details by [183] with some modifications, as shown in Fig 5.1C. P-peptides binding with IMAC-Fe gel were firstly eluted by 1% TFA buffer and then 40 mM ammonium hydroxide in 30% ACN. To achieve a good enrichment efficiency and phosphproteome recoverage, TFA elution and supernatant flow were dried separately and submitted to another two cycles of TiO_2 incubations, respectively.

5.3.3 LC-MS/MS analysis

C_{18} desalting following supplier's instructions (Nest group, US) was applied after P-peptides enrichment. Dried samples were then dissolved in 10 µl of MS loading buffer containing 3% ACN and 0.1% formic acid buffer and then 5 µl of sample was introduced to an LC-MS/MS. All experiments were run in biological triplicates and technical duplicates on a maXis UHR-TOF MS (Bruker, Germany) except the test of different acids that were run on Amazon MS (Bruker, Germany) and only biological duplicates were carried out due to instrument time limitation. LC gradient runs were from 3% to 40% B in 75 min followed by a 9 min wash at 90% B and a 14 min equilibration step at 3% B.

5.3.4 MS data analysis

MS raw data were converted into .mgf files using DataAnalysis (version 4.1, Bruker Daltonics) and submitted to an in-house Mascot Daemon (Version 2.4; Matrix Science) using a concatenated target/decoy database. The protein sequences of *S. solfataricus* P2 were downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Bacteria/Sulfolobus_solfataricus_uid108/) in 2013 Oct and its decoy format, the reversed sequence was created using Peptide Shaker (v 0.22.6, https://code.google.com/p/peptide-shaker/). The Mascot Daemon search parameters were as used as follows: trypsin as digestion enzyme with up two missed cleavages permitted, carbamidomethyl (C) as a fixed modification and oxidation (M), phospho (STY) as variable modifications. Peptide tolerance was set up with $\pm$0.02 Da and MS/MS tolerance with ± 0.1 Da [280]. For individual peptides search, a false discovery rates (FDR) calculation was determined as described in [214], and a FDR of 5% was applied with peptide score cut-off >25. Furthermore, for determination of P-sites, a threshold mascot delta score (MD-score) of 6 was used, corresponding to 5% of FLR (false localization rate) [216]. MD scores were calculated based on Mascot ion scores that were extracted from the corresponding Mascot dat-files. Proteins and peptides

extracted from Mascot were further analysed using Microsoft Excel 2013 (Microsoft, USA).

Mascot Daemon worked as a search engine, it can also process MS raw files into .mgf format. Phenyx only accept .mgf format files. To compare the results by using different search engines, the same .mgf files were submitted to either Phenyx or Mascot Daemon. One of advantages of Phenyx over Mascot is the superiority in a high-throughput protein identification project [208]. The similar searching parameters (compared to Mascot) were applied to our in-house search engine Phenyx (v.2.6, Geneva Bioinformatics, Switzerland). Briefly, MS tolerance of 0.02 Da and MS/MS tolerance of 0.1Da were used together with a charge stage of +1, +2, +3 and +4, minimums peptide length of 6, with peptides z-score of 4, p-value of 10^{-4} and AC score of 4. For individual peptides search, 5% of FDR was calculated as described in [214] and applied.

5.4 Results and discussions

5.4.1 Search parameters and search engine optimization

MS tolerance of 0.02Da was applied to both Mascot Daemon and Phenyx search engine for P-protein identification. Moreover, to compare the results derived from different searching engines, the same .mgf files were submitted to either Phenyx or Mascot Daemon. And results were shown in Table 5.1.

The results by using Mascot were divided into two parts: with or without P-site localization filter while P-site localization was not applied for Phentx results at the present because there is no available bioinformatics tool that can be used in conjunction with Phenyx data for P-site localization. The overlapped number between different searching parameters and searching engines is shown in Fig. 5.2.

Table 5.1 Number of peptide/proteins identified using different search engine and parameters.

Search engine	FLR	MS tolerance	P-pep	P-pro
Mascot	0.05	0.02	7	6
	1	0.02	9	7
Phenyx	-	0.02	205	178

FLR: false localisation rate. -: not determined. P: phosphorylation.

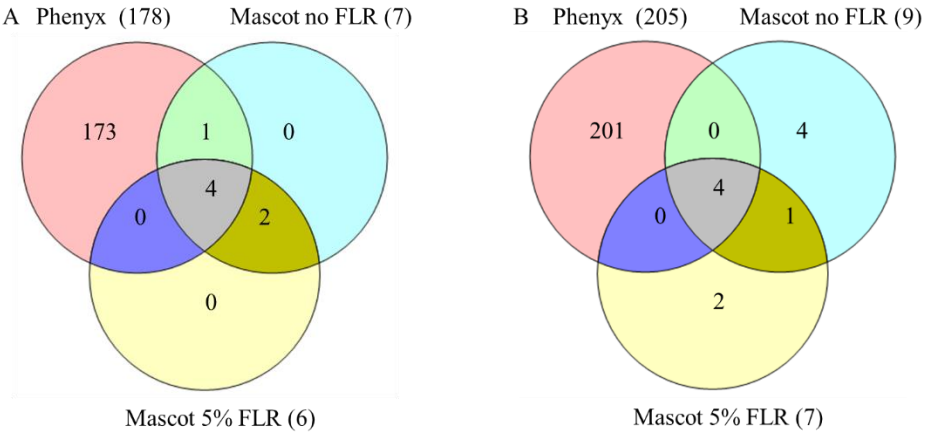

Fig 5.2 Overlapped number identified using 0.02Da as MS tolerance for Mascot and Phenyx: P-proteins (A) and P-peptides (B).

It can be seen from Fig 5.2 that only 5 P-proteins and 4 P-peptides were identified by using Phenyx and Mascot. The percentage of overlapped number of P-proteins from two search engine is not unexpected. One of the previous studies reported 247 overlapped proteins among the totally 801 proteins identified by using Phenyx (786) and Peptideshaker (372) [281]. The latter is an amalgamation of Mascot, Omssa and X! Tandem. It is a clear message that number of P-proteins identified by Phenyx was higher than Mascot. It agrees well with that Phenyx is superior to Mascot for protein identification [208].

Moreover, other advantages by using Phenyx search engine are: firstly, the in-house software tool kit can be easied used for iTRAQ quantifiecaiton analysis [224, 282]. Secondly, the identified P-protein/P-peptides can be compared with previous published data using PAcIFIC approach [7]. Thirdly, the P-sites annotation can be easied performed using the in-house script. However a major disadvantage (at this stage) is the accurancy of P-site localization information cannot be estimated (since there is no available bioinformatics tool that can be used for Phenyx data output). Therefore, taking into account of number of P-proteins identified, the possibility for comparision with published data and compatibility for using in-house software for quantitative phosphoproteomics analysis (iTRAQ data in Chapter 6), Phenyx data were used for further discussions as well as for analysing other data (Chapter 6).

The same amount of proteins (500 µg) were used for comparison between the different methods. The P-peptide enrichment efficiency of three different commercial materials were evaluated, including pre-packed TiO_2, IMAC-Fe and SIMAC. The number is the total identification summarised from 6 runs (biological triplicates and technical duplicates) for each sample run on the maXis UHR Q-TOF. A brief summary of the detection number from different experiments is shown in Table 5.1 The detailed effects of different materials are shown in Figs 5.2-5.10, and Tables ST 5.2-ST 5.9 in the Appendix, whilst the distribution of P-sites (single, double, triple and more than four P-sites) of different experiments are shown in Tables 5.2A-G. Optimisation of various acids as non-phosphopeptides excluders were run on the Amazon and only biological duplicates were tested due to instrument time. Only 5 unique P-peptides corresponding to 5 P-proteins were detected in control samples, where no P-peptides enrichment was applied. A total number of 205 unique P-peptides corresponding to 178 P-proteins were detected from all P-peptides enrichment experiments: merged data from all experiments detected by LC-MS (maXis UHR Q-TOF, Bruker), and a ratio of 10.6% for P-peptides /total peptides was also calculated. Furthermore, 151 and 54 P-peptides were observed for monophosphorylated (less than one P-sites) and multiphosphorylated (more than two P-sites) sites respectively. Among 205 P-sites observed, 83 (31.4%), 66 (25.0%) and 115 (43.6%) were found on Ser, Thr and Tyr, respectively. The distribution of P-sites, ratio of Ser/Thr/Tyr, in this study was 31.4%/25.0%/43.6%, and this differed slightly from a recent phosphoproteome study where the ratio of 25.8%/20.6%/53.6% was reported [7]. There is a note that Esser, *et al.* applied PAcIFIC (precursor acquisition independent from ion count) technique, where digested whole cell lysates were directly analysed by LC-MS/MS with an MRM program without P-peptides enrichment [7, 8]. In that study, a precursor range of 400-1200 m/z with 10 m/z intervals were applied. As a result, they also detected 1318 P-sites from a total of 540 P-proteins [7]. Therefore, it is a clear message that the PAcIFIC technique gave a higher identication number than the combination of different P-peptides enrichment approaches. However, the advantage of enrichment strategies over PAcIFIC approach is its potential application for further quantitative study. Also, LC-MS/MS running time reduced significantly from at least 8 weeks to 2 weeks. Additional cost including bulk batch of trypsin, IMAC or MOAC enrichment materials and C_{18} desalting columns need to be considered for choosing enrichment methods.

Table 5.1 Brief summarised detection number from different experiments.

Treatments	Protein amounts	Elute conditions	P-pep	Tot number of P-pep
IMAC	500 µg x 3	Amm + ACN	13	13
		SP	12	12
	500 µg x 3	Amm	8	8
SIMAC	500 µg x 3	IME	10	31
		TFA-TiO	8	
		TFA-TiO-2Cy	7	
		SP-TiO	12	
		SP-TiO-2Cy	8	
TiO_2 (Ratio 20/1)	500 µg x 3	F1	23	59
		F2	28	
		F3	18	
		F4	17	

Note: 500 µg of peptides was used for each experiments; n=3, biological triplicates with technical duplicates were run.

5.4.1 IMAC

Commercially-available IMAC-Fe was used to test its capability for P-peptide enrichment and results are shown in Fig 5.2; detailed information of P-sites is listed in Table TS5.1 in the Appendix. We noticed that there was a slight improvement of numbers P-peptides/proteins detected when ACN was added to the elution buffer (Fig 5.2): 13 and 13 unique P-peptides corresponding to 13 and 8 P-proteins with and without the presence of ACN. This finding agrees with a previous study reporting that the binding between P-peptides of α- and β-casein to Fe-IMAC can be enhanced by the addition of ACN [169]. Ye, *et al*, [169] suggested that the ionization of P-peptides can be relatively enhanced through the negative affect of the ionisation of acidic residues such as aspartate and glutamate by using a high concentration of up to 30% ACN.

Interestingly, the peptide sample after incubation with IMAC-Fe was submitted to LC-MS/MS to assess the enrichment efficiency. There is a higher number of P-peptides (12) corresponding to 11 P-proteins were detected from the supernatant flow. It indicates that multiple incubation cycles: tandem IMAC [166] or a subsequent combination with MOAC enrichment was necessary as proposed in previous reports [183] .

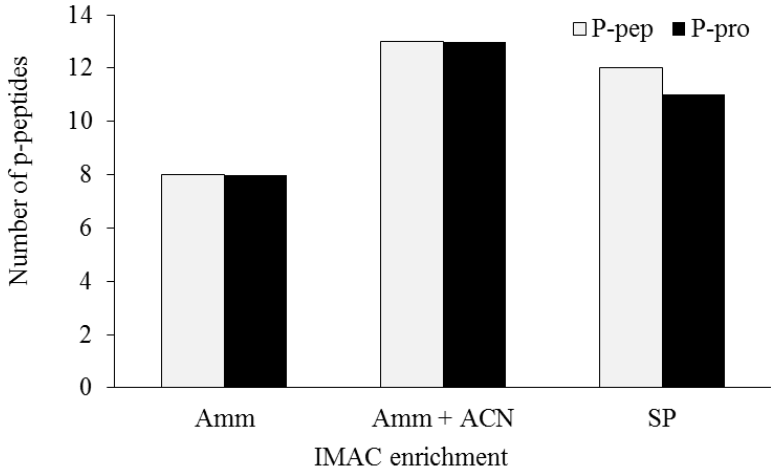

Fig 5.2 Total numbers of unique P-peptides and P-proteins identified by using IMAC. Amm is eluted by 40mM ammonia. Amm + ACN is eluted by 40mM ammonia in 30% ACN. SP is detected from supernatant flow.

5.4.2 TiO_2 enrichment

5.4.2.1 TiO_2 resins in different formats

Pre-packed TiO_2 resins in different forms (magnetic beads, TiO pipette tips, and TiO Spin tip) were employed to determine their P-peptide enrichment capabilities. Trypsin digested peptides from *S. solfataricus* P2 grown on standard glucose (0.4%) media were used to evaluate these TiO_2 workflows and results are shown in Fig 5.3. Obviously, different TiO_2 formats resulted in different numbers of detected P-peptides, but the distribution pattern of P-sites on Ser/Thr/Tyr was similar, as detailed in Table TS 5.2 in the Appendix.

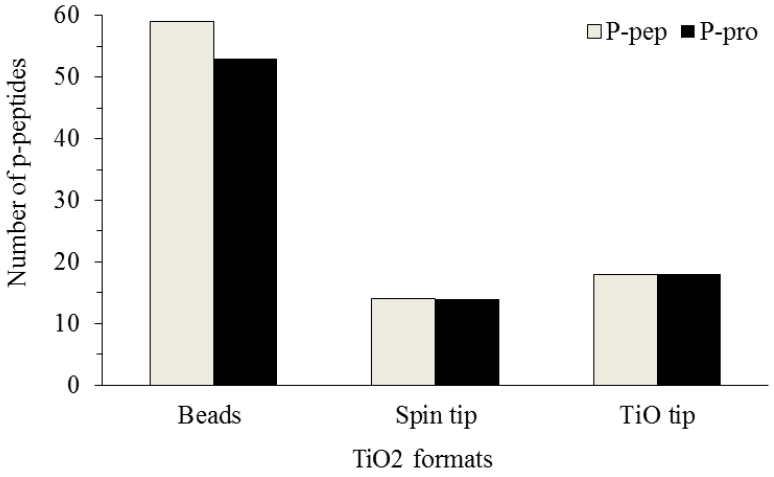

Fig 5.3 Total identification numbers of unique P-peptides and P-proteins by using different TiO_2 formats.

It is clear from Fig 5.3 that the use of magnetic beads led to a better result in terms of number/robustness of identified P-peptides/proteins. At a peptide/TiO_2 ratio of 20/1, 53

P-proteins were determined from 59 P-peptides with 71 P-sites identified, while the identification number of P-peptides and P-sites were only 14 and 18, 16 and 25 for spin tips and pipette tips, respectively. The P-peptide selectivity from beads was the highest compared to the others. One of the explanations is that: 4 cycles of incubation procedure of beads results in binding and elution of more P-peptides. In contrast, only one incubation process was applied to spin tips and TiO pipette tips. In addition, there was a report that physical properties of TiO_2 beads had some effects on their enrichment specificity and capacity [144, 172]. In addition, a high surface/volume ratio such as a mesoporous surface benefits more efficient binding of P-peptides with TiO_2 resins than smooth surface [144, 172]. The enrichment capacity might result from undocumented modifications of the TiO_2 materials. It is reasonable to infer that a continuously votexed incubation (30 min) for magnetic beads allows for sufficient reaction and therefore contributes to the higher identification number.

Different concentrations of ACN and trifluoroacetic acid (TFA) in loading buffer were evaluated for spin tips, TiO tips and beads: 0.4% TFA in 80% ACN *vs* 0.1% TFA in 3% ACN *vs* 5% TFA in 80% ACN, respectively. Higher concentrations of ACN and organic acids such as TFA, acetic acid and formic acid in loading buffers was reported to help to reduce non-specific interactions with the beads [278], therefore, it is one of the reasons for various enrichment efficiencies among these experiments.

Only 18.6% of multiple P-peptides was characterised from beads, the percentage of multi P-peptides determined from spin tip and TiO tips were 9 (14.3%) and 30 (33.3%) respectively. Considering sample handling and phosphoproteome coverage, a combination of magnetic beads and IMAC-Fe seems to be suitable for a large scale global phosphoproteome study.

5.4.2.2 Effect of different acids on non-P-peptides binding

We next attempted to optimise series of acids as non-P-peptide inhibitors in the sample loading buffer for phospho enrichment using magnetic beads, since this has been shown to play an important role in enrichment efficiency [157]. Using the same amount digest peptides as mentioned above, we tested the performances of glycolic acid, lactic acid and citric acids, as well as glutamic acid in the loading buffer. Different concentrations of citric acids were prepared according to [176]. The results are shown in Fig 5.4A and

summarised in Table TS5.3 in the Appendix. Furthermore, percentages of mono- and multiple- P-sites were calculated and depicted in Table5.11C. Only 27 P-sites were identified and 27 unique P-peptides were determined from 26 P-proteins, corresponding to a phosphorylation ratio of 5.8% without any enrichment treatment (Control). The number of detected P-peptides was greatly improved by incorporating acids in the loading buffer of TiO_2 enrichment as reported by [176, 283, 284], and the average phosphorylation ratio (as defined by [179], number of P-peptides/total number of identified peptides x 100%) reached up to 14.0%. It clearly indicates the necessity of enrichment and the importance of non-P-peptides excluders should be used for complex samples.

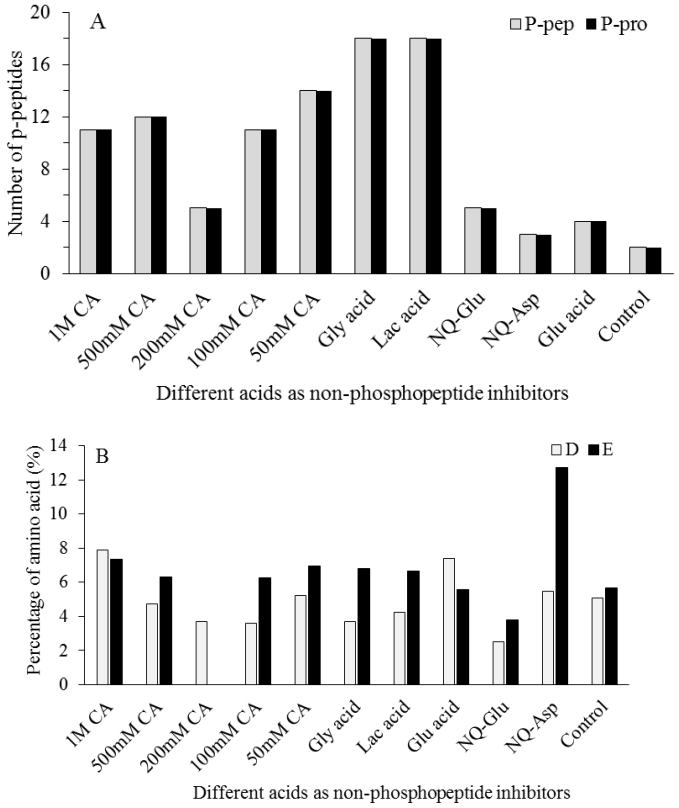

Fig 5.4 Total numbers of unique P-peptides and P-proteins detected when different acids as non-P-peptides inhibitor (A). Percentage of acidic acids in P-peptides sequence when different acids were applied (B). CA is citric acid. Gly acid is glycolic acid. Lac ac is lactic acid. NQ-Glu is lactic acid with addition of glutamine. NQ-Asp is lactic acid with addition of asparagine. Glu acid is glutamic acid. D is short for aspartic acid. E is short for glutamic acid.

From Fig 5.4A, we can see that glycolic acid and lactic acid resulted in a similar number of the highest P-peptide and p-protein identification and P-sites determination numbers (in Table TS 5.4 in Appendix). It is in agreement with Jensen and Larsen's [157] work,

which was performed on a standard peptide mixture using MALDI-MS. The performance of various concentrations of citric acid was different, and the P-peptide identification number of 200 mM CA was the lowest, which supported Zhao, *et al.*'s observation [176]. The identification numbers of other four CA concentrations were similar, and there was no significant increase of detection number when CA concentration increased up to 500 mM and 1M. There was no significant difference for amino acid ratio of N and Q in both P-peptide and non-P-peptide (results are shown in Table TS5.5 in the Appendix), that explained why N/Q-Glu and N/Q-Asp expereiments (lactic acid as non-P-peptide inhibitor) did not show better performance compared to lactic acid. The percentage of acidic acids (aspartic acid and glutamic acid) in P-peptide sequence were calculated, and results are depicted in Fig 5.4B and Table TS5.5 in the Appendix. Fig 5.4B shows that, compared to the control (5.1%), the percentage of Asp (D) were reduced in almost all the acids except 1M CA and Glu. In contrast, the effect of acids on and Glu (E) was different. The addition of all acids except 200 mM CA and lactic acid together with Asp shows a negative effect on reducing Glu binding.

The ratios of single and multiple P-peptides were also calculated as shown in Table 5.2C. It shows clearly that mono- P-peptides were predominantly detected from all acids treatment. A possible reason will be discussed in Section 5.4.4.

5.4.2.3 Effect of peptide-to-beads ratio on P-peptides enrichment

The ratio of peptide to TiO_2 beads has been found to play a significant role in P-peptides enrichment efficiency [179, 180]. The amount of TiO_2 beads slurry applied for efficient P-peptides enrichment was investigated by testing different ratios of peptide to TiO_2 beads (w/v) ranging from 80/1, 40/1, 20/1 and 10/1. In total, 145 P-peptides corresponding 124 P-proteins were determined from TiO_2 beads (from all four comparisons of peptide/beads ratios). The numbers of identified P-peptides from each ratio of peptide/beads experiments are shown in Fig 5.5 and detailed information about the distribution of Ser/Thr/Tyr sites, and single- or multiple- P-peptides are listed Table 5.2D and Table TS5.6 in the Appendix. From Fig 5.5, we can see that numbers of unique P-peptides increased with the decrease of peptide-to-beads ratios such as 57, 46, 59 and 43 for 80/1, 40/1, 20/1 and 10/1 respectively, and the highest number of detected P-peptides was corresponding to the ratio of 20/1. The corresponding P-proteins are detailed in Table TS 5.6. There were 51, 43, 53 and 42 P-proteins, and P-sites were 69, 63, 71 and

60 for ratios of 80/1, 40/1, 20/1 and 10/1, respectively. Although the ratio of 20/1 outperformed to the others, (40/1, 20/1 and 10/1) the difference was not significant. However, a slightly higher number of P-peptides with more than two P-sites was observed from the deficient beads usage (40/1 compared 20/1 and 10/1): 16 *vs* 11 *vs* 14. The detection number was similar to them (11) when insufficient beads (ratio of 80/1) was used. It agrees with a previous study reporting that the reduced use of beads benefits multiple P-peptide binding, but P-peptides selectivity will decrease if excessive or inadequate beads was used [179]. Taken all together, the optimised ratio 20/1 was used for future quantitative phosphoproteomics experiments.

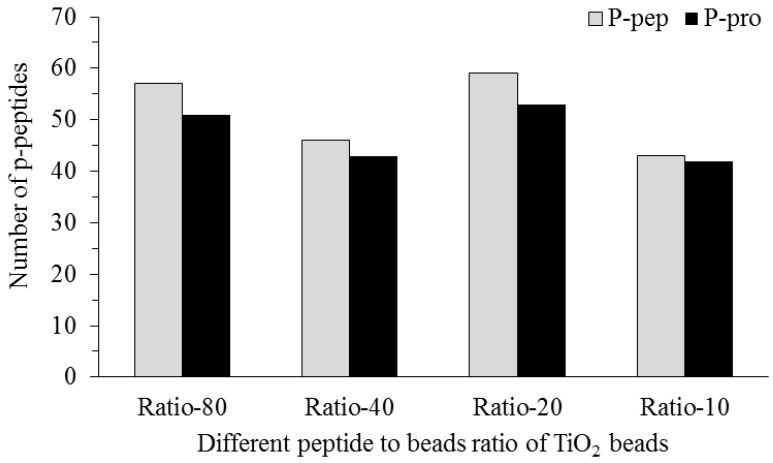

Fig 5.5 Total numbers of unique P-peptides and P-proteins corresponding to different ratios of peptide to beads.

In summary, peptides-to-TiO_2 ratios affected the enrichment efficiency of TiO_2 beads. Here, for *S. solfataricus* P2, the optimum peptide-to-TiO_2 ratio (w/v) of 20/1 was determined, however, the phosphorylation ratio was lower than other studies (30% for all the tested ratios compared to 80% of HeLa cell lysate [179]). It might be explained by high phosphorylation level (30%) in Eukaryotes compared to *S. solfataricus* (17%) as reported by [7]. Furthermore, a higher phosphoproteome coverage might be achieved by using efficient elution such as the application of a pH stepwise elution and pH step gradient elution [181, 182]. In addition, compared to conventional use of ammonium hydroxide, which favours the elution of short P-peptides (1-1.5KDa), bis-tris propane tends to elute long P-peptides (1-4KDa) [177].

5.4.2.4 Consecutive incubations by TiO_2 beads

Other factors such as consecutive incubations of TiO_2 beads have been found to increase total numbers of identified P-peptides [179]. We investigated four consecutive

incubations for each specific ratio to determine both the best ratio of peptides-to-beads and number of incubated cycles. Detailed results are shown in Figs 5.6 and Table 5.2E, and Table TS 5.7 in Appendix. The overlap number between different cycles is shown in Venn diagram Fig 5.7 A and B. It can be seen that 23 P-peptides were identified from the 1st incubation by using peptides to beads ratio of 20/1, and P-peptides increased slightly to 28 for the 2nd cycle. Of these, 11 P-proteins were overlapped from the two incubations. A decrease was observed from 2nd to the 3rd incubation, and 9 P-peptide from 9 P-protein were overlapped.

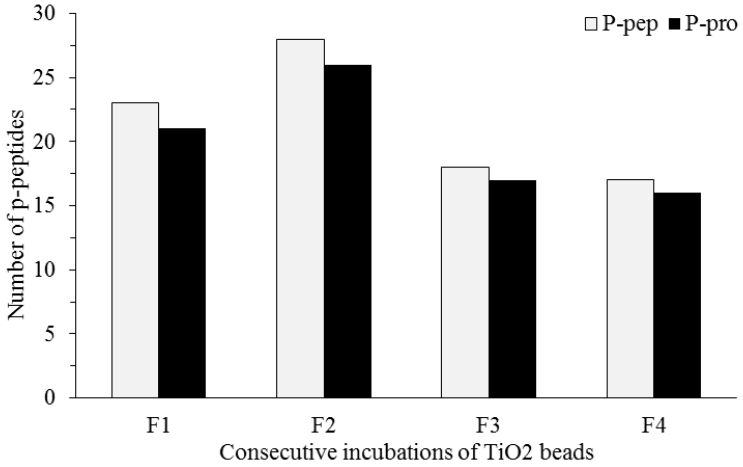

Fig 5.6 Total numbers of unique P-peptides and P-proteins in relationship to consecutive incubation of beads.

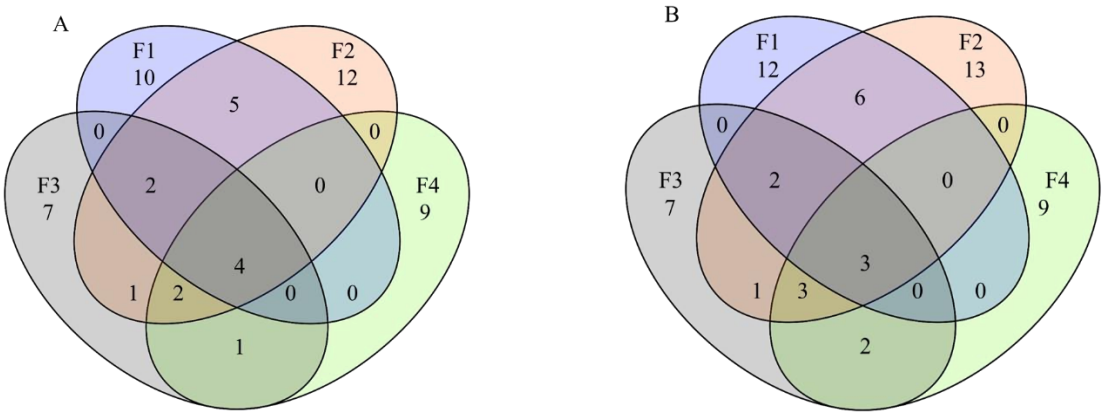

Fig 5.7 Venn diagrams showing the overlap numbers between consecutive incubations using TiO_2 beads: identified P-proteins (A) and P-peptides (B).

In addition, there are still 16 P-proteins detected from 4th incubations, but the overlap protein percentage is 43.75% (7/16) with the 3rd incubation. So, no further test was performed by considering instrument time and experiment budget. In comparison to the

one-step incubation, a 256% (23 *vs* 59) increment in the total P-peptides and 252% (21 *vs* 53) of p-protein identification were achieved by incorporation of four cycles of continuous incubation. The increase of incubation cycles from 1 to 4 led to increased numbers of total P-peptides detected (Table TS 5.7).

5.4.3 SIMAC

Fig 5.1C shows 5 elutions, which were run from one biological sample using the modified SIMAC technique. A total of 31 P-peptides corresponding to 30 P-proteins were detected, with 38 P-sites determined. A detailed identification number for each step was shown in Fig 5.8 and supplementary Table TS 5.8. The distribution of single and multiple phosphorylation ratio can be found in Table 5.2F. As a result, 28 P-peptides were detected from ammonium elution. Furthermore, 7 P-peptides corresponding to 7 P-proteins were detected from the 1st cycle of TiO_2 enrichment of TFA elution. And 8 P-peptides corresponding to 7 phosphorylated proteins were detected from the 2nd cycle of TiO_2 enrichment using TFA as elution buffer. However, a high number of non-phosphopeptides were also detected from the 2nd cycle of TFA elution. It indicates that one cycle TiO_2 treatment for TFA elution is good enough. After ammonium hydroxide and TFA elution, some P-peptides were also detected from supernatant flow treated by two cycle of TiO_2, 12 and 8 P-peptides were detected, respectively. However, a high number of non-P-peptides was also found from the 2nd runs, corresponding to a phosphorylation ratio of 2.3%. Therefore, a balance between P-peptides recovery and enrichment selectivity should be considered for global phosphoproteomic analysis.

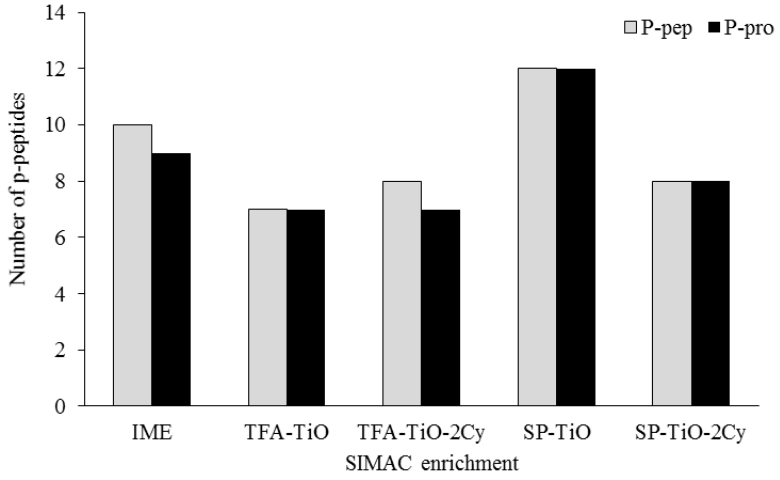

Fig 5.8 Total numbers of unique P-peptides and P-proteins detected using SIMAC technique. IME is the 1st elution using IMAC. TFA-TiO is TFA elution and goes to TiO_2 enrichment. TFA-TiO-2Cy is the 2nd

cycle of TiO_2 enrichment with TFA elution. SP-TiO is supernatant flow submit to TiO_2 enrichment. SP-TiO-2Cy is 2nd cycle of TiO2 enrichment of supernatant flow.

5.4.4 IMAC *vs* TiO_2 *vs* SIMAC

Numbers of detected P-peptides and P-proteins from IMAC, SIMAC and TiO_2 (ratio 20/1) techniques are shown in Fig 5.9 and Table 5.2G, and detailed in Table TS 5.9 in the Appendix. Comparisons were made based on the same amount of proteins used for P-peptides enrichment (500 µg of proteins and a biological triplicate applied for each condition). A high percentage of triply charged peptides were observed in P-peptides from all experiments and details can be found in Table TS5.10 in the Appendix. The overlap of P-peptides and P-proteins identified between different methods are shown in Fig 5.10 A-B. Among these P-peptides, 59 of them were detected from TiO_2 conditions (peptide/beads ratio of 20/1 and 4 cycles of incubation), whereas only 13 were found from a single IMAC enrichment and 31 were detected from the modified SIMAC technique, corresponding to a different percentage of P-peptides: 6.1% *vs* 3.8% *vs* 5.2%, respectively. It has been shown that P-peptides enrichment using TiO_2 resulted in higher coverage of P-peptides (for 1 cycle of incubation: P-proteins from 23 P-peptides). It supports the previous report that TiO_2 is more selective than IMAC for P-peptide enrichment [278].

In addition, only seven P-proteins and eight P-peptide were overlapped in three techniques (Fig 5.10 A and B). A higher number of unique P-peptides were observed than that of P-peptides (23 compared to 51) for SIMAC and TiO_2 enrichment experiments. In total, 7.7% (1 of 13) of P-proteins and 15.4% (2 of 13) of P-proteins determined from IMAC were overlapped with those detected from either SIMAC or TiO_2.

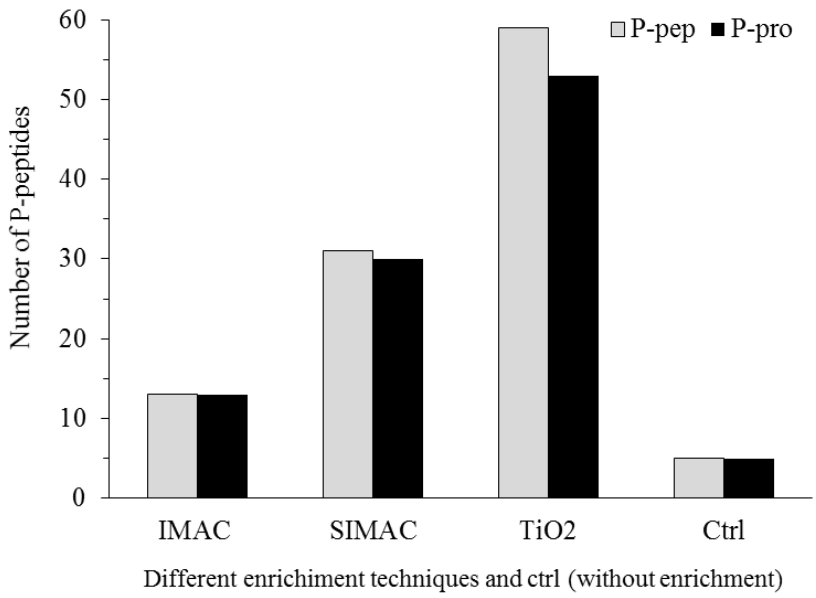

Fig 5.9 Total numbers of unique P-peptides and P-proteins identified by different treatments. Ctrl is without enrichment.

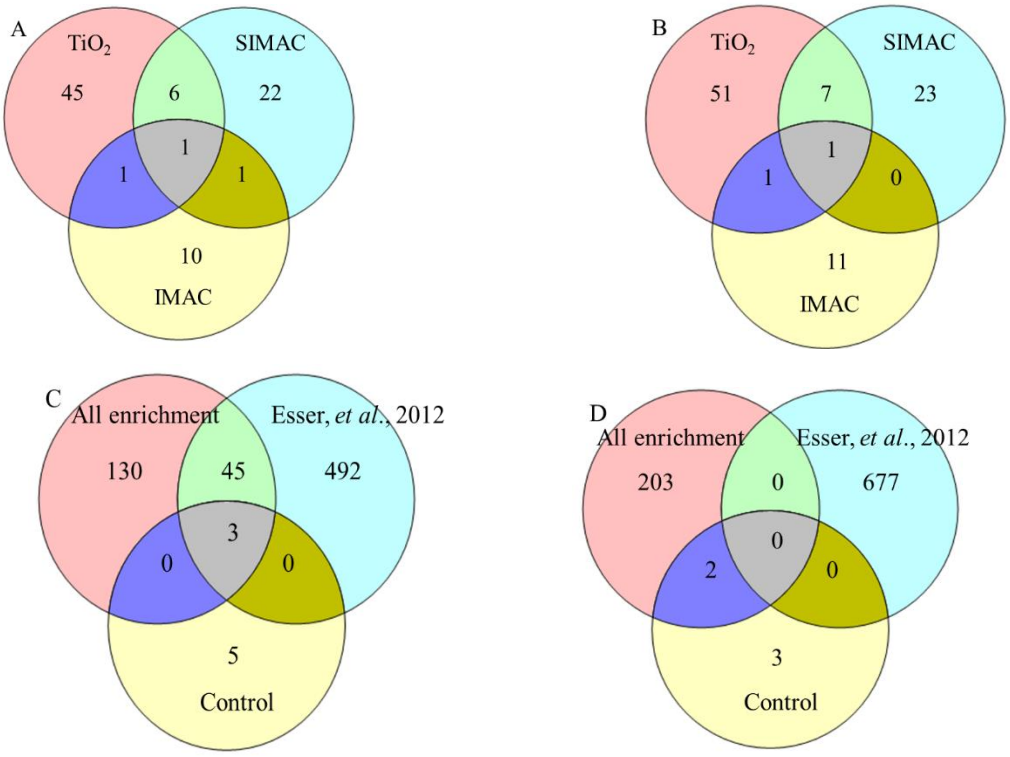

Fig 5.10 Venn diagrams showing the overlap between different enrichment experiments: identified P-proteins (A) and P-peptides (B). In total, 59, 31 and 13 P-peptides corresponding to 53, 30 and 13 P-proteins were determined using TiO_2, SIMAC and IMAC techniques respectively. (C) Venn diagram showing the overlap between previous study [7], all enrichment and control (without enrichment) experiments identified using maXis-UHR-TOF of unique P-proteins (C) and P-peptides (D).

An improvement of P-peptides recovery by using SIMAC could have resulted from the contributions of both IMAC and TiO_2, which supports the previous report [183]. It

supports the previous statement that a combination use of IMAC and MOAC is necessary for comprehensive phosphoproteome analysis [184]. Altogether, 82 unique P-peptides from 76 P-proteins with a determination of 100 P-sites were identified by a combination of magnetic beads with the optimum peptide/beads ratio of 20/1 and 4 consecutive incubations and modified SIMAC technique (including 7 common P-proteins and 8 P-peptides). The optimised strategy will be applied for further global quantitative enrichment optimisation.

The distribution of single, double, triple and higher P-sites from different enrichment experiments are shown in Table 5.2 A-F. Compared to multi-P-peptides, a large number of mono-P-peptides (almost 80% of identified P-peptides ≤ 2 P) were identified from all experiments. It was possibly due to the suppression of multi-P-peptides by abundance of single- or non- P-peptides in MS analysis process [166], since LC/ESI-MS/MS was reported to be efficient for singly charged P-peptides detection [278]. Overall, of the 59 P-peptides identified by TiO_2 magnetic beads (peptide to beads ratio of 20/1), 48 were mono-phosphorylated, 10 were doubly-phosphorylated and only one triply-phosphorylated. In comparison, the SIMAC technology detected 25 mono P-peptides (1P), 5 doubly phosphorylated (2P) and only one triply phosphorylated (3P).

Table 5.2 Comparisons of different P-peptides enrichment strategies

Table 5.2A IMAC.

	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	4P	1P	≥ 2P
Amm	12	1	1	1	92.3	7.7
Amm + ACN	7	1	0	0	87.5	12.5
SP	9	2	1	0	75.0	25.0

Table 5.2B Different TiO_2 formats.

	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	4P	1P	≥2P
Beads	48	10	1	0	81.4	18.6
Spin tip	12	2	0	0	85.7	14.3
TiO tip	12	5	1	0	66.7	33.3

Table 5.2C Various acids as non-phosphopeptides excluders for TiO_2 beads.

	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)			
	1P	2P	3P	4P≥	1P	2P	3P	4P≥
1 M CA	7	3	0	1	64	27	0	9
500 mM CA	3	1	1	0	25	8	8	0
200 mM CA	2	1	1	1	40	20	20	20
100 mM CA	6	4	1	0	55	36	9	0
50m M CA	6	6	1	1	43	43	7	7
Glycolic acid	11	3	1	3	61	17	6	17
Lactic acid	9	6	1	2	50	33	6	11
NQ-Glu	2	1	2	0	40	20	40	0
NQ-Asp	2	1	0	0	67	33	0	0
Glutamic acid	2	2	0	0	50	50	0	0
Control	0	2	0	0.0	0	100	0	0

Table 5.2D different peptide-TiO_2 beads ratio.

Peptide/beads ratio	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	4P	1P	≥2P
Ratio-80	46	10	1	0	80.7	19.3
Ratio-40	30	15	1	0	65.2	34.8
Ratio-20	48	10	1	0	81.4	18.6
Ratio-10	31	13	1	0	72.1	27.9

Table 5.2E Consecutive incubations of TiO_2 beads.

TiO_2 Incubations	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	≥ 4P	1P	≥2P
F1	38	17	2	0	66.7	33.3
F2	32	12	2	0	68.1	29.8
F3	30	15	4	0	61.2	38.8
F4	25	14	4	0	58.1	41.9

Table 5.2F SIMAC.

SIMAC	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	4P	1P	≥2P
IME	8	2	0	0	80.0	20.0
TFA-TiO	5	1	1	0	71.4	28.6
TFA-TiO-2Cy	7	1	0	0	87.5	12.5
SP-TiO	12	0	0	0	100.0	0.0
SP-TiO-2Cy	7	1	0	0	87.5	12.5
SIMAC	25	5	1		80.6	19.4

Table 5.2G Different enrichment strategies.

	Number of P-peptides with various P-sites				Ratio of P-peptides with various P-sites (%)	
	1P	2P	3P	≥ 4P	1P	≥2P
TiO_2	48	10	1	0	81.4	18.6
IMAC	12	1	1	0	92.3	7.7
SIMAC	25	5	1	0	80.6	19.4
Control	4	1	0	0	80.0	20.0

Note: P indicates phosphorylation.

The numbers of P-peptides/proteins detected using affinity-based IMAC-Fe and TiO_2 enrichments were compared with the study reported by Esser [7] (without P-peptides enrichment), and results are shown in Fig 5.10C. A total of 48 P-proteins were overlapped from both studies. In addition, only one P-peptide with phosphorylation site was detected in both studies. In total, the P-protein number was 675, which accounts for 22.5% of the proteome of *S. solfataricus*.

P-proteins from all enrichment experiments (identified using maXis-UHR-Q-TOF) were distributed into 20 out of 26 arCOG functional categories in *S. solfataricus* with regard to their biological functions based on arCOG classification (http://archaea.ucsc.edu/arcogs/). Results are shown in Fig 5.11.

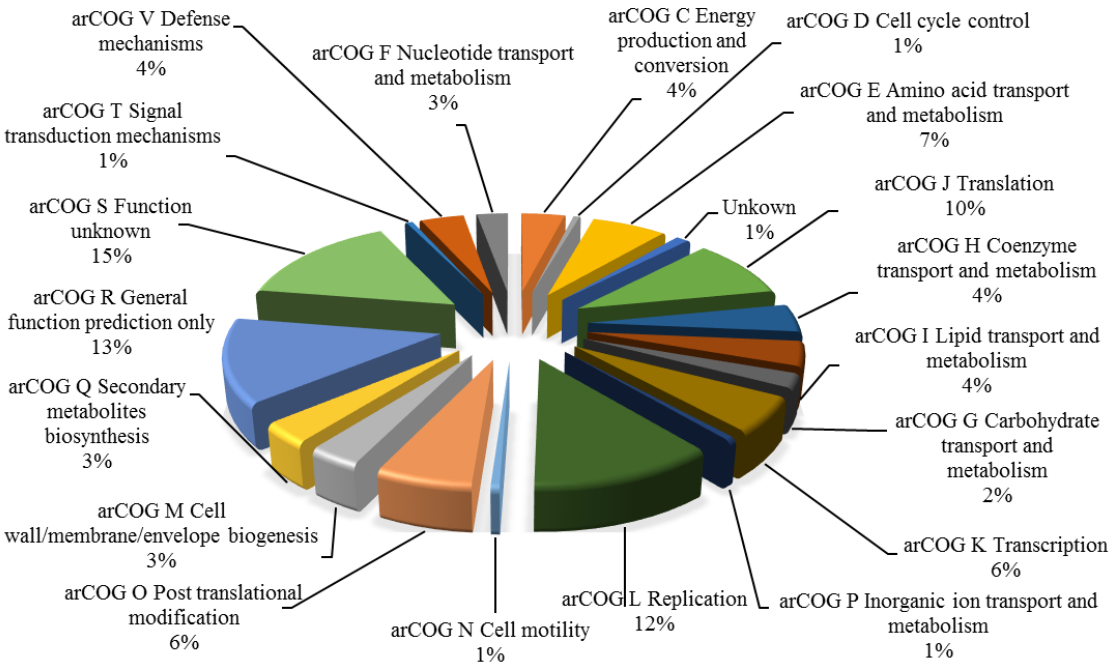

Fig 5.11 Pie chart of functional categories of all enriched P-proteins (178) identified from 205 unique P-peptides.

Fig 5.11 shows the two largest group of enriched phosphorylated proteins were proteins with unknown function (arCOG S, 15.2%) and general function prediction only (arCOG R, 13.5%), followed by arCOG replication (arCOG L, 11.8%), , and amino acid transport and metabolism (arCOG E, 6.7%).

The unique P-proteins identification number were compared from different techniques when similar carbon source (0.4% glucose) was supplied for *S. solfataricus* P2 and results are depicted in Fig 5.12. Compared to which without incorporating any enrichment strategies, a high number was observed by using enrichment techniques (205 *vs* 311). Also, a high number of P-proteins involved in replication (arCOG L) and post translational modification (arCOG O). The distribution into arCOG functional groups were less (20 *vs* 21). No P-proteins were classified into arCOG intracellular trafficking, secretion and vesicular transport (arCOG U). But similar distributions on cellular metabolism arCOG (F, I, K) were identified by using enrichment and without enrichment techniques [7].

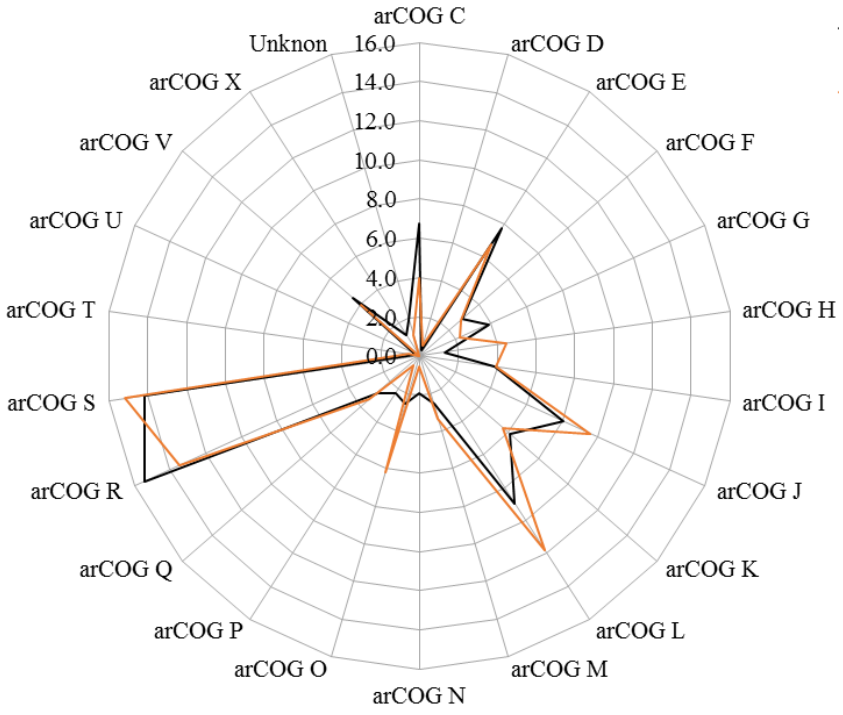

Fig 5.12 Comparison of the classification of arCOG functional codes investigated by using with (orange) and without (black) enrichment techniques [7] reveals some differences.

In Fig 5.13, it gave an example of P-site annotation. The neutral loss of H_3PO_4 was observed with a good coverage of b- and y- ions for identification.

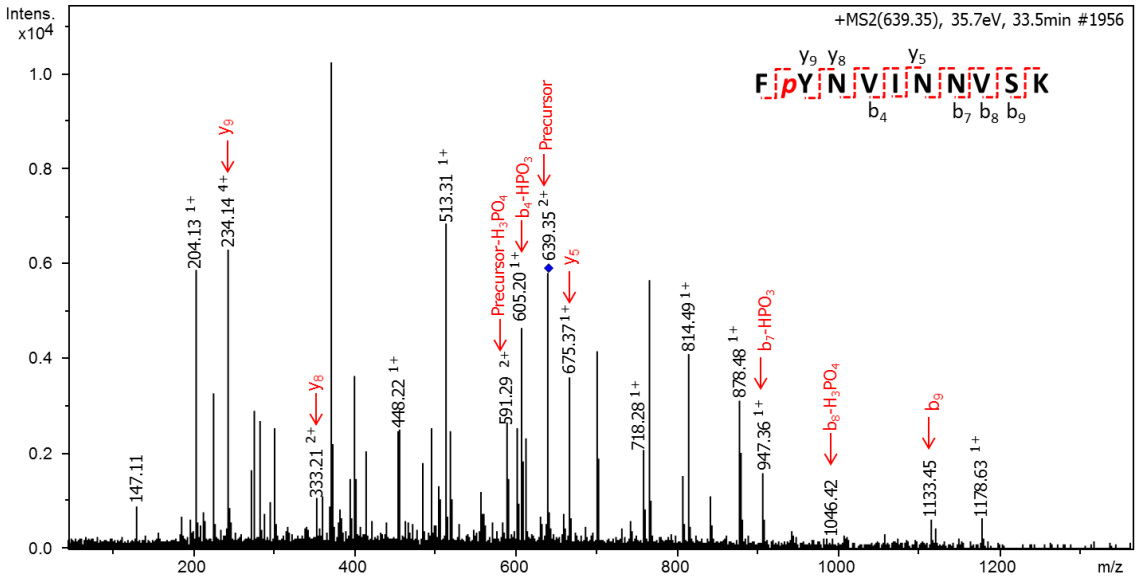

Fig 5.13 An example of sptetrum annotation for phosphopeptide (including b- and y- series ions and neutral loss) determined from P-protein gi 13814416 identified using Phenyx search engine.

5.5 Conclusions

Using a maXis UHR-TOF LC-MS/MS we have been able to assess the P-peptide enrichment effectiveness from the cell lysates of *S. solfataricus* using IMAC, TiO_2 and SIMAC. Compared to enrichment-free PAcIFIC technique. The advantage of enrichment strategies include less MS time. A higher percentage (almost twice more than that in total peptides) of triply charged P-peptides were observed from all enrichment results detected by the maXis Q-TOF as detailed in Table TS 5.10 in the Appendix. It indicates that the possibility of improvement on phosphoproteome coverage by using both CID and ETD, where the latter has proven to be more efficient for detection of highly-charged ≥3+ charge) peptides [205]. In addition, P-peptides with one P-sites were predominantly identified using TiO_2 (81.4%), it shows the necessity for the combination of other enrichment technical workflows.

We can conclude that:

- Firstly, magnetic beads led to higher number of P-peptides/proteins detected compared to pipette tips and spin tips, although it is still unclear why one of these formats resulted

in a better result than another for P-peptides enrichment. It might be due to undocumented modifications of the resins materials made by their respective manufacturer(s).

- Secondly, both SIMAC and consecutive TiO_2 showed their benefits compared to single IMAC use in terms of P-peptides and p-protein identification.

- Thirdly, the optimised peptide-to-TiO_2 beads ratio of 20/1 and 4 consecutive incubations for magnetic beads were chosen for further experiments. In total, 71 P-sites from 53 P-proteins were detected with an identification of 59 non-redundant P-peptides. Together, 145 unique P-peptides were identified from 124 P-proteins, with a determination of 187 P-sites on Ser (55, 29.4%), Thr (41, 21.9%) and Tyr (91, 48.75%) from the peptides/ beads ratio experiments.

- Finally, the combind use of SIMAC with TiO_2 provide complementary information which benefits a better phosphoproteome coverage. Together, we have optimised an efficient strategy that can be used for large-scale phosphopeoteomic experiments of *S. solfataricus* P2: peptide/magnetic beads ratio of 20/1 and 4 consecutive incubations and the modified SIMAC. However, phosphorylation enrichment from complex cell lysates with a high dynamic range is still a challenge. Large amounts of proteins are still recommend for global quantitative phosphoproteome studies.

Chapter 6 Expanding phosphoproteomic study from identification to quantitation in *Sulfolobus solfataricus*

6.1 Abstract

Based on information presented in in Chapter 5, P-peptides (phospho peptides) were enriched from whole cell lysates by a combination of TiO_2 magnetic beads and SIMAC (IMAC and TiO_2) before being fractionated using a HILIC technique. A total of 73 unique P-peptides corresponding to 61 P-proteins (phospho proteins) were detected with 91 P-sites (phospho sites) observed. The P-sites distribution of 36.3%/28.6%/35.2% were observed on Ser/Thr/Tyr residues. All quantified P-proteins were widely distributed into 16 arCOG functional categories and 31 KEGG pathways. Of these P-proteins, only 4 of them were differentially regulated in response to the carbon source changing from glucose to tryptone using 0.01Da as MS tolerance. A half of these regulated P-proteins are enzymes involved in amino acid metabolic pathways. The data provide valuable quantitative phosphoproteomic information for the understanding of how *S. solfataricus* P2 responded to different carbon sources (glucose *vs* tryptone). Most of the quantified P-proteins involved in carbohydrate metabolism are unaffected in expression level, while P-proteins functional in amino acid metabolism were up or down regulated.

6.2 Introduction

Recognition of important roles of protein phosophorylation in signal transduction in almost all cellular processes [94, 95, 285] has led to studies on the phosphoproteome of bacteria and eukaryotes. Previous studies within archaea demonstrated that protein phosphorylation [96, 108] mainly occurs on Ser, Thr or Tyr residues. Recent large scale phosphoproteome studies in the halophilic archaeon *Halobacterium salinarum* [286] and hyperthermophilic *Sulfolobus* species [7, 47] have revealed that protein phosphorylation sites are widely distributed in various cellular processes and especially enriched in metabolic processes.

The first large set of protein phosphorylation on Ser/Thy/Tyr and the initial genome-wide phosphoproteomic study in archaea was performed on *H. salinarum,* which resulted in the determination of 81 P-sites from 69 P-proteins [286]. *S. solfataricus* able to grow under different carbon sources (from glucose to tryptone) has resulted in phosphoproteomic level changes, from which 1318 P-sites were identified from 540 unique P-proteins [7]. Furthermore, 809 unique P-proteins were detected in phosphoproteomics experiments of

three *S. acidocaldarius* strains (one wild and two mutants) [47]. However, these phosphoproteomic studies were carried out in terms of P-proteins identification without enrichment only. Until now, no quantitative phosphoproteomics studies in archaea have been performed.

In this Chapter, we have studied the global quantitative phosphoproteome of *S. solfataricus* P2 grown on two different carbon sources by a combination of P-peptide enrichment techniques (TiO_2 and SIMAC). We have identified a number of (61) unique P-proteins that are involved in 31 of the 78 KEGG pathways and their biological significance was determined by pathway mapping. This is the first quantitative phosphoproteome study in the third domain of life - the archaea.

6.3 Materials and methods

6.3.1 Microorganism growth conditions and protein digestion

Cell growth of *S. solfataricus* strain P2 on 0.4% glucose or 0.2% tryptone was described in detail in Chapter 3. Cell pellets collection, cell lysates, protein extraction and digestion were performed following procedure detailed in Chapter 3. In total, 10 mg of protein was used per phenotype. An SDS-PAGE gel was run to test trypsin digestion efficiency.

6.3.2 Phosphopeptides enrichment

The resultant peptides were desalted by C_{18} materials (Discovery-C18, Sigma) before being enriched by PHOS-Select Iron Affinity Gel (IMAC-Fe) and TiO_2 (modified SIMAC) as well as 4 cycles of continuous incubation of TiO_2 magnetic beads with a peptide/beads ratio of 20/1. P-peptides enriched by a combination of SIMAC and TiO_2 beads from whole cell lysates were combined together respectively and desalted thoroughly to remove ammonium before iTRAQ labelling.

6.3.3 Isobaric p-peptide labelling

One set of 4-plex iTRAQ reagents were applied (114 and 115 are biological replicates of peptides from 0.4% glucose growth condition, while 116 and 117 are biological replicates for 0.2% tryptone). The labelling process was performed according to the manufacturer's procedure (ABSciex, USA). Dried labelled p-peptides were fractioned using a HILIC column as described in Chapter 4 [224]. A total of 32 HILIC fractions were collected. These

P-peptide fractions were dried and C_{18} (Nest, USA) desalted and frozen at -80°C for further MS analysis.

6.3.4 LC-MS/MS analysis and data analysis

MS analysis was performed on ESI maXis-UHR-TOF MS (Bruker, Germany) coupled with on-line LC as described in Chapter 5. A 90 min LC gradient was used for iTRAQ samples and the MS set up procedure were used as described in Chapter 5. Only 26 of HILIC fractions were run due to an instrument performance problem.

All MS/MS spectra files were firstly converted into .mgf files using DataAnalysis (V4.0, Bruker Daltonics, Germany) before submitting to an in-house Phenyx search engine (v.2.6, Geneva Bioinformatics, Switzerland). The same peptide sequence of concatenated target/decoy database *S. solfataricus* strain P2 was applied as mentioned in Chapter 5. Similar search parameters were used for P-protein identification as described in Chapter 5. Briefly, tolerance of 0.01 Da and 0.1 Da were used for MS and MS/MS (respectively) search. And a charge of +1, +2, +3 and +4. Acceptance parameters are: minimums peptide length of 6. Peptides Z score, P-value and AC score do not have a clear optimum, so these values are set as 4, 10^{-4}, and 4 for the following process with in-house software FDRslide for FDR estimation. Trypsin was used as the proteolytic enzyme with up to two missed cleavages. Cys_CAM, iTRAQ _K, iTRAQ_Ntermi as fixed modification and oxidation plus phosphorylation (STY) as variable modifications. In addition, deamidation (caused by basic elution of IMAC or TiO_2 enrichment) as variable peptide modification were incorporated for P-peptide identification [158]. Also, proteins with more than 1 peptide were used for quantification. Results were extracted to Excel 2013 (Microsoft 2013, USA) for further analyses. Data were then submitted into in-house techniques for quantification determination [224]. The inhouse FDRslide tool kit was used and 5% of FDR was applied for all data before performing quantification analysis.

6.4 Results and discussions

In this research, a 4-plex iTRAQ experiment was applied to investigate the global phosphoproteome responses of *S. solfataricus* P2 to carbon source change from glucose to tryptone. Trypsin digested peptides from *S. solfataricus* grown on 0.4% glucose and tryptone were labelled with iTRAQ reagents 114 and 115 (for glucose), and 116 and 117 (for tryptone) (biological duplicate for each condition).

6.4.1 Search parameters for Phenyx search engine

In addition, besides 0.01Da as MS tolerance, a wider MS tolerance (0.05Da) was also tested since this parameter was used in our group [282] and other study where the same instrument was used for global proteomic analysis [280]. In total, 259 P-proteins were quantified when 0.05Da MS tolerance was applied and 18 P-proteins showed abundance change. However, only the data with strict MS tolerance (0.01Da) were discussed in the following parts.

A total of 600 proteins from 1853 unique peptides were detected in this experiment, however, only 61 unique P-proteins were detected from 73 P-peptides. A total of 91 P-sites were identified from 73 non-redundant P-peptides. A brief summary of P-proteins detected and quantified until recently is listed in Table 6.2.The P-peptide ratio (number of P-peptides/all detected peptides, as defined by [179]) is only 3.94% (73/1853) and P-protein ratio (P-protein/all identified proteins) is 9.90% (61/600), which is lower than expected. It indicates that further optimisation for quantitative level phosphoproteome studies needs to be carried out. Furthermore, only one injection was run for all the HILIC fractionations and few rich intensity fractions were lost due to an instrument problem as well and no technical replicates were applied. Therefore, we believed that if all fractions were run with technical replicates, results will be improved. This thus represents a proof of principle.

Among these quantified proteins, 62 proteins overlapped with proteins detected from the enrichment technical optimisation (data from Chapter 5). Only 9 quantified P-proteins overlapped with those identified from previous enrichment experiment (data from Chapter 5). A comparison of both quantified P-proteins (A) and P-peptides (B) from this experiment (a) with other P-proteins identified from enrichment technique (d) (178 P-proteins from Chapter 5), and without enrichment (PAcIFIC) technique (b and c) (311 and 311 for glucose and tryptone respectively as reported elsewhere [7]) are shown in Figs 6.2 A and B. From Fig 6.2A, we can see that only 2 P-proteins overlapped from all of these experiments. Among the 21 shared proteins between P-identification (data from Chapter 5 and Esser, *et al* [7]) and P-quantification (this chapter): 6 were only identified by incorporating enrichment technique (data from Chapter 5) and 5 of them were only detected using a PAcIFIC strategy [7] in the glucose feed condition, and in total 4 of them were only detected in tryptone media respectively.

Table 6.2 Number of P-proteins identified from *S. solfataricus* P2

Experiments	Glucose	Tryptone	
PAcIFIC[1)]	311	311	
Enrichment[2)]	178	-	
Control[2)]	5	-	
Sum	820		Identification
iTRAQ[3)]	61		Quantitation

1) Esser et al., 2012. 2) Data from Chapter 5. 3) Data from this Chapter

In contrast, no P-peptides overlapped from both techniques (iTRAQ quantified and PAcIFIC). And no P-peptides overlapped between enrichment (Chapter 5) and quantitative (this Chapter) experiments.

The percentage of P-peptides carrying a triple charged in the P-peptide enrichment experiments was higher than that in triply charged non-P-peptides (data from Chapter 5). However, the percentage of +3 charged peptides for the iTRAQ experiment is almost the same between P-peptides (203/400, 50.8 %) and non-P-peptides (1148/2797, 41.0%).

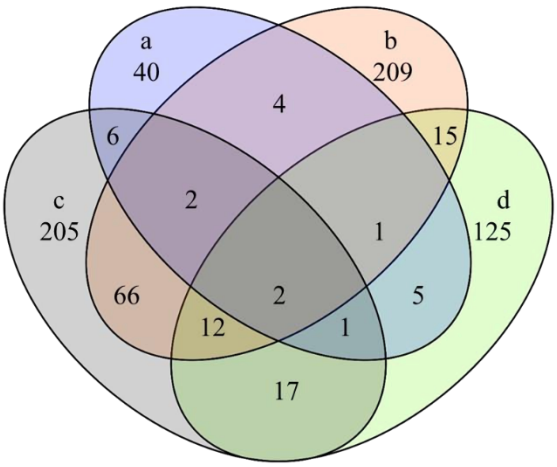

Fig 6.1 Venn diagram of overlapped P-proteins between different experiments: a) Quantified number from this iTRAQ experiment. b) Cells grown on glucose and c) Cells grown on tryptone using (PAcIFIC) technique [7]. d) Identified number from P-peptide enrichment experiment (Chapter 5).

In total, 91 P-sites were identified, with a distribution of 33 (36.3%), 26 (28.6%) and 32 (35.2%) on Ser/Thr/Tyr residues. The distribution ratio of P-sites on Ser/Thr/Tyr (36.3%/28.6%/35.2%) in this quantitative study differed from the ratio in Chapter 5 (31.4%/25.0%/43.6%) and previous phosphorylation study (25.8%/20.6%/53.6%) as reported elsewhere [7].

Furthermore, most of the detected P-peptides (77.5%) were mono P-peptides, and only 22.5% of the quantified P-peptides carried up to three P-sites. In total, of these 71 P-peptides identified, 55, 12 and 4 were mono-, doubly-, and triply-phosphorylated respectively, and no P-peptides carried more than four P-sites. There were some P-peptides carrying more than triply P-sites observed from different experiments such as from TiO_2 (34.2%), or SIMAC (35.8%), as mentioned in Chapter 5. It is less with enrichment results (5.6% for data in this Chapter), which may be caused by a strong ion suppression that the detection of multiple P-peptides were greatly suppressed by the high abundance of non- or single- P-peptides [166], because the ESI-LC-MS/MS has been observed to be beneficial to detect singly charged P-peptides detection [278]. It also might be a result from the loss of samples during multiple C_{18} desalting steps that are required by the iTRAQ labelling procedure (to avoid ammonium interference from P-peptides elution) and after HILIC fractionation (for efficiency of LC-MS/MS running).

6.4.2 Functional classification of quantified P-proteins

Of these 61 quantified P-proteins, 49.2%, 6.6% and 42.6% were located in cytoplasmic, cytoplasmic membrane and unknown categories respectively. The quantified P-proteins were classified into 16 of total 26 arCOG functional categories and a percentage distribution of these results is shown in Fig 6.3. We can see that most of quantified P-proteins are involved in metabolism pathways. This finding is well agreed with previous phosphoproteomics study of *S. solfataricus* P2 [7]. The largest group belonged to general function prediction (arCOG R, 18%) followed by uncharacterized conserved protein (arCOG S, 11.5%s) since almost a half of the genome has not been annotated yet [6]. Similare percentage (11.5%, and 9.84%) of of P-proteins were classified into energy production and conversion and amino acid metabolism (arCOG C and E), respectively.

The percentage of P-proteins quantified in this Chapter (61 proteins), from enriched optimisation work (178 proteins) and P-proteins from Esser [7]: without enrichment with their distributions based on arCOG are shown in Fig 6.4. We can see that most of the detected P-proteins were distributed into arCOG R and arCOG S. Compared to P-proteins from identification (data from Esser [7] and Chapter 5), percentages of two categories (arCOG C and M) increased in the quantitative phosphoproteomic analysis: 2.57% *vs* 3.37% *vs* 8.20%.

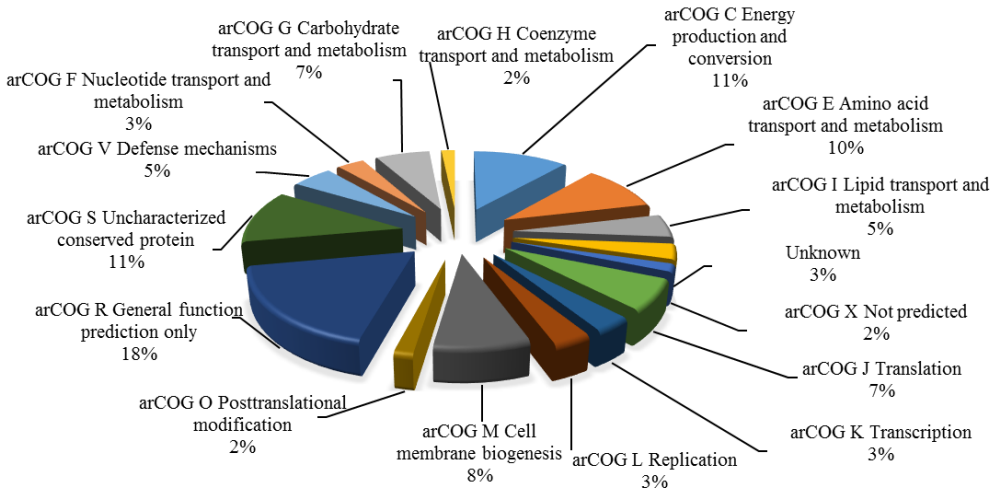

Fig 6.3 Classification of quantified P-proteins based on arCOG functions in *S. solfataricus* P2.

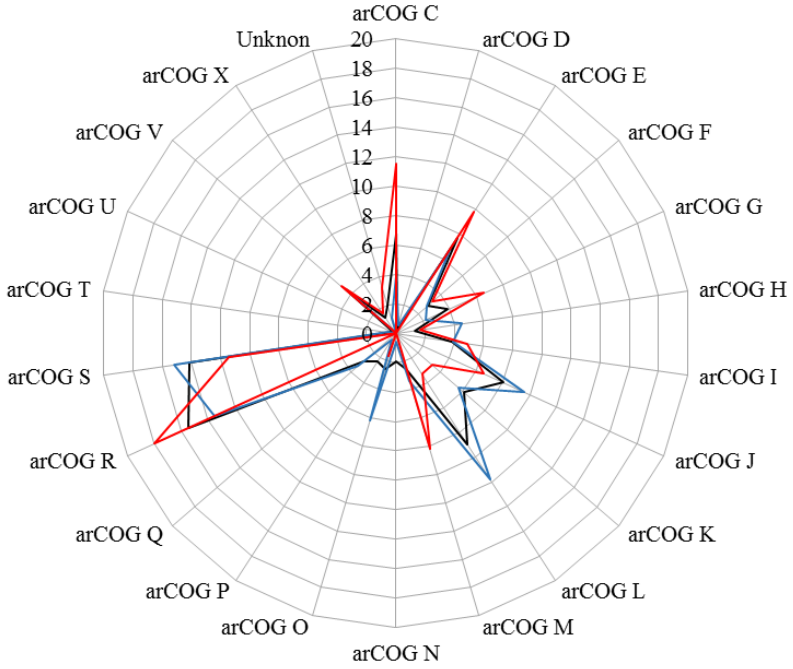

Fig 6.4 Percentage distribution of quantified P-proteins (61 proteins, red line) *vs* P-proteins (311 proteins, blue line) from Esser et al. [7], and P-proteins from enrichment optimisation (178, black line).

6.4.3 P-proteins involving KEGG pathways

There are 27 of P-proteins were annotated as hypothetical proteins. Quantitated P-proteins were plotted against *S. solfataricus* KEGG pathways (http://www.genome.jp/kegg/pathway.html). As a result, 18 out of 61 P-proteins were distributed into 31 of 78 KEGG pathways, as listed in Table 6.2A-C. Fifteen of these P-proteins are involved in multiple metabolic pathways.

Table 6.2 P-proteins involving in different KEGG pathways

Table 6.2A Carbohydrate metabolic pathways

Carbohydrate metabolism	**P-proteins**
Citrate cycle (TCA cycle)	SSO2815
Glycolysis / Gluconeogenesis	SSO0286
Pentose phosphate pathway	SSO0666
Pyruvate metabolism	SSO2496
Glyoxylate and dicarboxylate metabolism	SSO2496
Starch and sucrose metabolism	SSO0810
Propanoate metabolism	SSO2496
Butanoate metabolism	SSO2496
Amino sugar and nucleotide sugar metabolism	SSO0810
Pentose and glucoronate interconversions	SSO0810

Table 6.2B Amino acid pathways

Amino acid metabolism	**P-protein**
Ala, Asp and Glu metabolism	SSO1930
Gly, Ser and Thr metabolism	SSO0248
Arg and Pro metabolism	SSO1930
Lys degradation	SSO2496
Phe metabolism	SSO2996
Trp metabolism	SSO0452, SSO2017, SSO2122
β-Ala metabolism	SSO2720
Val, Leu and iso-Leu biosynthesis	SSO0248 SSO0504
Val, Leu and iso-Leu degradation	SSO2496

Table 6.2C Other metabolic pathways

	KEGG pathways	**P-proteins**
Nucleotide metabolism	Pyrimidine metabolism	SSO0976
	Purine metabolism	SSO0241
Energy metabolism	Carbon fixation pathways in prokaryotes	SSO2815, SSO2496
Translation	Aminoacyl-tRNA biosynthesis	SSO0504, SSO0938
Xenobiotics biodegradation and metabolism	Benzoate degradation	SSO2496
Metabolism of terpenoids and polyketides	Terpenoid backbone biosynthesis	SSO2496
Lipid metabolism	Fatty acid metabolism	SSO2496

Note: P-protein with underline indicates abundance change

6.4.3.1 Carbohydrate metabolism

In total, only 6 quantified P-proteins involved in the CCM was observed. The reconstruction of the CCM in relationship with P-proteins are shown in Fig 6.4 A. Combining the information from Table 6.2 and Fig 6.4A, it can be seen that P-proteins are involved in multiple metabolic pathways. For instance, SSO2496 is involved in glycolysis, pyruvate metabolism and TCA cycle, which agrees well with literature observations [7]. Furthermore,

only one of the detected P-proteins SSO0810 in pentose and glucuroate interconversions as well as starch and sucrose metabolism pathways were differentially regulated. The regulation of phosphorylation level was not significantly changed, which might be due to the negative affect by other PTMs, as mentioned in Section 6.4.4. It provides useful information for phosphorylation in *S. solfataricus* P2.

6.4.3.2 Amino acid metabolism

In total, four P-proteins SSO0248, SSO0504, SSO1930 and SSO2496 involved in amino acid metabolism including Val, Leu and iso-Leu biosynthesis, Gly, Ser andThr metabolism, Ala, Asp and Glu metabolism respectively. However, none of them show abundance change.

6.4.4 Regulated P-proteins in responding to changing carbon sources

Quantitative P-proteomic analysis of *S. solfataricus* P2 results in responding to different carbon sources showed that most of quantified P-proteins were involved in metabolic pathways, which agrees well with previous phosphoproteomics study at the P-proteins identification level. In response to change carbon source from glucose to tryptone, 4 P-proteins showed an abundance change: 1 of them were down-regulated and 3 protein was up-regulated (as listed in Table 6.3). The arCOG annotation and are listed in Table 6.3.

Table 6.3 P-proteins with their abundance changes

GI	arCOG annotation	Log2 ratio	STD
SSO0625	GTP-binding protein	1.15	0.01
SSO0637	hypothetical protein	2.02	0.23
SSO0810	UDP-glucose 6-dehydrogenase (ugd)	2.16	0.04
SSO2815	2-oxoacid--ferredoxin oxidoreductase, alpha chain	0.79	0.09

Note: log2 ratio <1 indicates down-regulation, and log2 ratio >1 indicates up-regulation. – indicates no KEGG pathway related.

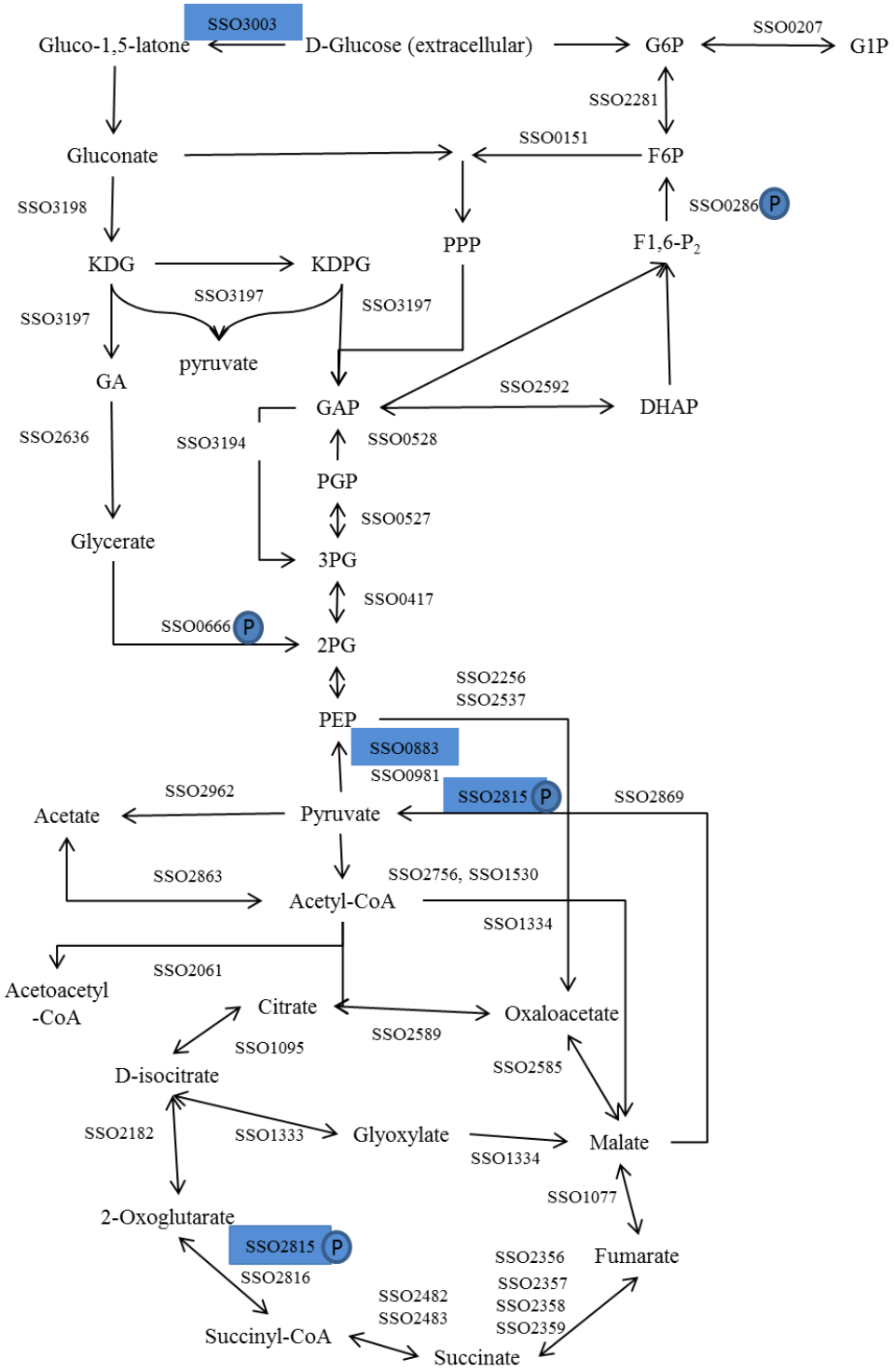

Fig 6.4 Central carbohydrate metabolism (CCM) in *S. solfataricus* P2 (Glycolysis, TCA and pyruvate metabolism) (A). G1P: glucose 1-phosphate; G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; F1,6P: fructose 1,6-bisphosphate; DHAP: dihydroxyacetone 3-phosphate; KD(P)G: 2-keto-3-deoxy-(6-phospho)gluconate; GAP: glyceraldehyde 3-phosphate; PGP: 1,3-bisphosphoglycerate; 3-PG: 3-phosphoglycerate; 2-PG: 2-phosphoglycerate; GA: glyceraldehyde; PEP: posphoenolpyruvate; PPP: pentose phosphate pathway. Blue background is down-regulated proteins. P is P-protein.

SSO0625 is GTP (guanosine triphosphate)-binding protein, which is predicted to be guanosine triphosphatase (GTPase) in *S. solfataricus* P2 [7]. GTPase binds and hydrolyses GTP, which works like a molecular switch that have the 'ON' GTP-bound activated state and 'OFF' GDP (guanosine diphosphate) -bound inactivated state. GTPases are widely distributed in three domain of life and are well-studied in bacteria and eukaryote, which are proposed to play important roles in signal transduction, cell division, cell viability and they are stress-related proteins [287, 288]. There are three GTP-binding proteins in *S. solfataricus*: SSO0269, SSO0581 and SSO0625 [7]. Hitherto, SSO0269 was classified into Hflx (high frequency of lysogenization) family based on crystal structure [289]. It shares common Hflx in N-terminal and G-domain in C-terminal with Hflx of *E. coli*, which is a ribosome-related GTPase but hints a role in translation rather than ribosome biogenesis [290]. However, the biological function of any *S. solfataricus* GTP-binding proteins (SsGBPs) is largely unknown except the investigation of GTP hydrolysing activity and thermal stabilities of SSO0269 through mutations study [291]. The biological function of Hflx in *E. coli* was remained unexplained until the observation of GTP-mediated autophosphorylation recently by Kaur and co-workers [292]. A preference of Mn^{2+} to Mg^{2+} as cofactor was found for Hflx and further investigation indicates a role in Mn homeostasis [292]. In addition, Mn-treated *ΔHflx* mutant cells exhibit filamentation phenotype and SOS response (global response to DNA damage) [292]. Autophosphorylation of Hflx was proposed to be a start point of signal transduction for Mn homeostasis and studies are still going on [292]. Autophosphorylation of leucine-rich repeat kinase 2 (LRRK2) on Thr^{1343} was reported, which is a protein harbor both kinase and GTPase activities [293]. This protein has a Ras-like GTP-binding domain and a kinase domain, and mutations at either domain lead to Parkinson's disease [293]. Multiple Ser or Thr residues were found to be potential phosphosites and two novel sites were identified within the ROC/GTPase domain *in vitro*. In addition, it indicates autoregulation of LRRK2 [293]. The determination of regulated phosphoprotein SsGBP (SSO0625) upon carbon source change propose the interest to reveal the mystery of GTPases in *S. solfataricus.*

Moreover, it can be noticed that one of the most abundant proteins: thermosome β-subunit (SSO0282) was quantified and it showed 'light' down regulation when lose parameters were applied (0.1Da for MS tolerance). A further study on SSO0282 seems promising to unravel the cross-talk between phosphorylation and acetylation in archaea, since N-terminal protein acetylation of SSO0282 and other 16 proteins in *S. solfataricus* have been reported [294]. A

list of proteins with their abundance changes and PTMs modifications are shown in Table 6.4. Phosphorylation-dependent binding was proposed to be impaired by lysine acetylation of 14-3-3 protein, which was investigated through mutation lysine sites (K50, K69 K118+K123) [88]. Furthermore, the predicted down regulation of Zn-dependent alcohol dehydrogenase (ADH-2, SSO0764) may be of interest due to its potential function in alcohol and acetone metabolism [35, 40, 41] in *S. solfataricus.* However, the abundance change cannot be distinguished from protein or phosphorylation level, further validation work still needs to be carried out.

Table 6.4 Proteins with different PTMs modifications

Protein	**Acetylation** [a]	**Phosphorylation-identification** [b]	**Phosphorylation-Quantitation** [c]	
SSO0278	+	-	-	Up
SSO0317	-	+	+	Up
SSO0352	+	-	-	Down
SSO0862	+	+	+	Down
SSO0962	+	-	-	Down
SSO1457	+	-	-	Up

Note: a) Data from [294]. b) Data from Chapter 5. c) Data from this Chapter.

6.5 Conclusions

The application of P-peptide enrichment techniques coupled with quantitative proteomics enabled us to access the quantitative phosphoproteome network of *S. solfataricus* P2 in response to carbon source change from glucose to tryptone. Among the determined 600 proteins, 61 of them were phosphorylated. The quantified P-proteins were widely distributed into 17 of 26 arCOG categories. Evidence was provided at the quantitative level that P-proteins are widely involved in amino acids, carbohydrate metabolism and other cellular processes. However, only 4 of these quantified P-proteins were differentially regulated under glucose *vs* tryptone conditions (with the dataset in hand). Four of these regulated P-proteins are involved in carbohydrate or amino acid metabolism and the others have not been characterised. The majority of P-proteins were not affected by changing carbon sources (glucose or tryptone). Here is the first time the quantitative analysis of P-proteins in *S. solfataricus* has been reported, offering potential application for other achaeal studies.

Chapter 7 Global metabolome analysis on response of *Sulfolobus solfataricus* to different carbon sources

7.1 Abstract:

The lack of 46 genes in *S. solfataricus* PBL2025 (a spontaneous mutant) results in quantitative proteome level change compared to model strain P2, also, many changes occurred at proteome level when changing carbon source from glucose to tryptone for both these strains. It would be predicted that altered genetic level would impact on metabolites and therefore metabolome. However, little information is known concerning the alteration of global metabolome. Metabolomics results based on GC-MS experiments indicated that 163 metabolites could be identified from all experiments and 113 of them were assigned a KEGG C number (used for pathway mapping), while only 47 metabolites involved in KEGG pathways. Amino acid metabolisms were significantly affected and most of the detected amino acids showed high abundances, which supports for quantitative proteomic data in Chapter 4. The regulation of *S. solfataricus* in responding to carbon source change from glucose to tryptone might follow feedback-inhibition mechanism.

7.2 Introduction

S. solfataricus P2 is a widely used model strain in archaeal research. *S. solfataricus* PBL2025 has been a valuable strain as genetic tool through gene deleting in biological study [241]. By changing the carbon source from glucose to tryptone, Esser et al [7] observed a significant phosphorylation level change of proteins, especially those involved in CCM of *S. solfataricus* P2. Our previous proteomic studies (data from Chapter 4) have revealed interesting differences related to central carbohydrate metabolism (CCM), amino acids and energy metabolism between two strains (P2 and PBL2025) grown on glucose media and the same strain grown on different carbon sources (glucose *vs* tryptone). Proteomics techniques have been widely performed with regarding to interpreting the unique metabolic pathways of this archaeon [4]. By contrast, metabolomic study is still very limited except the study on CCM response of P2 to optimal (80°C) and suboptimal (75°C) living temperatures under standard glucose media [38] and CCM changes in metabolites levels between 2-keto-3-deoxygluconate kinase deletion mutant strain PBL2025⊿3195 and PBL2025 grown on Brock media at 76°C [48].

By incorporating the GC-MS technique, this work describes the application of metabolomics methods in *S. solfataricus* study to characterise global metabolites changes between different

strains as well as different carbon sources. Same biological samples (for proteomic/phosphoproteomic investigation in Chapter 4, 5 and 6) grown on same carbon source (glucose or tryptone) were used. The objective is to obtain preliminary data to provide global metabolomic information to model P2 strain and two mutant strains PBL2025 and PBL2073. It is also anticipated that these data together with our previous proteomic data may provide fundamental information to unravel the -omics changes (proteomic and metabolomic) in responding to different growth conditions of *S. solfataricus*.

7.3 Materials and methods

7.3.1 Sample preparation

S. solfataricus strains P2, PBL2025 and PBL2073 were cultivated separately in 250ml flasks at 80°C, pH 3.0 and 120 rpm in a horizontal shaking thermal incubator (Thermotron, Infors, UK) in Brock media with 4 g/L glucose or 2 g/L tryptone as sole carbon source. A volume of 50ml of cultured cells was harvested in late-exponential growth phase (OD_{650} = 1.0 ± 0.1) based on growth curves (see Chapter 3) and centrifuged at 5,000 x g for 10min at 4°C. Cells pellets were frozen using liquid nitrogen and then stored at -80°C until required. All chemicals were purchased from Sigma-Aldrich (Gillingham, UK) unless otherwise stated. Biological triplicates were applied.

For extraction of metabolites, the cell pellet was washed by 1.5ml of 0.9% (w/v) NaCl and then distilled water (5,000 x g, 5min, 4°C). Subsequently, cells were re-suspended in 500 µl ice cold (-20°C) methanol containing an equivalent volume of acid washed glass beads (425−600µm) (compared to cell pellet volume). Global metabolites were extracted using a disruptor (Genie Vortex, USA) with 7 cycles alternately 45 sec of disruption and 2 min of incubation on ice between each run. Supernatants containing the soluble crude extracts were centrifuged at 13,000 × g for 30 min at 4°C, transferred and dried by a vacuum centrifugation (Eppendorf Concentrator 5301, Germany). Global metabolites were derivatized as described in details elsewhere [295]. Briefly, dried samples were re-suspend in 25 µl of dimethylformamide contain 0.1% pyrimide. Later, 25 µl of N-methyltrifluoroacetmide was added. Samples are incubated at 80°C for 60 min and are ready for further analysis. Nor-valine was used as external standard; its derivatization followed the above procedure. Samples and external standard was mixed before GC-MS injection.

7.3.2 GC-MS analysis

GC-MS settings for global metabolome analyses were used as follows: analysis was performed on a GC coupled to a Finnigan Trace DSQ single Quadrupole GC-MS coupled to an auto-sampler (model AS3000, Thermo Electron, Waltham, USA) equipped with a 30 m × 0.25 mm × 0.25 μm df stabilwax fused silica column (Thames Restek, Bucks, UK). Hot needle injections with a 0.1 μl injection volume were employed. Helium was used as carrier gas at a constant flow rate of 1 mL/min. The splitless injector was set to 250 °C, running on the splitless mode. The oven temperature profiles were programmed to trace global metabolite: 40 °C held for 0.5 min, followed by a 30 °C/min ramping up to 200 °C and held at 200 °C for 1 min.

The GC-MS was operated in the full scan mode with mass range covering from 50 to 650 m/z. The ionization temperature was applied at 230°C and 10^5 eV was also used for ionization process. An external standard (Nor-valine) was added to all phenotypes for normalizing data purposes. A biological triplicates and a technical duplicates were performed for each sample.

7.3.3 Data processing and analysis

All chromatograms (raw data) were firstly converted into .mzXML format by MSConvert of ProteoWizard (version 3.0 4624) and submitted to online open source XCMS [296] for statistical analysis and visualized cloud plot data [297]. Cloud plot figures were generated based on significantly different extraction total ion chromatography (TIC) features, and criteria was set up as fold change ≥1.5 and p-value < 0.01. However, the information extracted from XCMS are only used for statistic purpose, since the data was searched against plant specific METLIN database.Identification of metabolites was achieved by search against NIST main library and extracted data was modified (to remove derivation information) using in house Macros program (from Dr.Vaidyanathan). The mass spectrum (Xcalibur .raw file) was analysed using the Automated Mass Spectral Deconvolution and Identification System (AMDIS) (V2.70) to identify the compounds by matching detected data with the mass spectral library, which linked to the National Institute of Standards and Technology (NIST) MS search 2.0 NIST 2011 (NIST 11, USA). A deconvolution issue need to be considered which was resulted from the shared fragment ions by co-eluting compounds. To achieve a better identification of even low abundant peaks in the TIC various and to determine the maximum number of components with smallest false-positive hits, optimised parameters were used [298]. The deconvolution settings for AMDIS were: scan direction: high to low; instrument type: quadrupole; adjacent peak subtraction, 0; resolution, high; sensitivity, high; shape requirements,

medium with target compounds from main library; component width: 12 (signal to noise ratio will be better with small component width) with minimum match factor 60. After deconvolution, the match factor >60% and probability >20% were considered for which the experimental spectrum matches with NIST reference spectrum [299]. The relative intensity ('area' extracted from AMDIS) was normalized against the area of Nor-valine. Thus metabolite abundance were expressed as ratios relatives to external standard. Therefore, the abundance change between different samples can be compared in terms of 'relative change'.

Each metabolites was matched with unique C number (entry number of metabolites in KEGG COMPOUND database), which was obtained by manually search against KEGG database (http://www.kegg.jp/kegg/kegg2.html). The related pathways were mapped by searching against database of *S. solfataricus* [300].

7.4 Results and discussions

Until now, there has been no single software or search engine that can be solely used for global metabolic data analysis. To have a basic understanding the response of *S. solfataricus* strains to different carbon sources, all the GC-MS data was firstly submitted for statistical analysis using online XCMS [296] and results are listed in Table 7.1, whereas cloud plot figures are depicted in Fig 7.1A-F. However, the database information used in XCMS are not specific for microorganism, so further metabolites identification was achieved by searching against the NIST library purchased from AMDIS. The identification numbers of metabolites from different experiments are listed in Table 7.2. Comparisons between different carbon sources are shown in Fig 7.2, and differences between various strains are shown in Fig 7.3

7.4.1 Statistic analysis using XCMS

Statistic features of metabolites changed between different comparisons are summarised in Table 7.1, and quantitative proteome data (withdrawn from Chapter 4) are also shown in this Table. For strains grown on the same carbon source, a similar change could be observed between significantly changed metabolomic features from GC-MS and abundance changed proteins from LC-MS/MS (data from Chapter 4). For instance, for cells grown on glucose media, a high number of proteins showed their abundance changes for P2 and PBL 2025 comparison (GA *vs* GB) than PBL2025 and PBL2073 (GB *vs* GC) comparison: 158 proteins *vs* 81 proteins. Correspondingly, more changed features was detected at metabolome level: 690 *vs* 219 for GA *vs* GB and GB *vs* GC comparisons respectively. For GA *vs* GB, the absence of 47 genes (*SSO3004-SSO3050*) in PBL2025 resulted in a higher number change than which

only one gene difference in both proteome and metabolome levels. Furthermore, it can be seen that the effect from changing carbon source seems different between different strains (P2 and PBL2025). The regulation of 159 proteins were affected and 533 features changed from metabolic level for P2 grown on tryptone *vs* glucose, whereas a high number proteins (208) showed their increase or decrease in abundance relative to PBL2025 but a small number metabolome features (425) were found. However, a solid conclusion cannot be withdrawn for other comparisons, since only 5 comparisons of proteomic data are available due to technical limitation for data analysis between different iTRAQ sets.

Examples of significantly changed metabolomics features are shown in cloud plot Figs 7.1 A-F. A single bubble in cloud plot (Fig 7.1) indicates a feature. Different colours indicates the up or down regulation of the feature: green for up- red for down-regulation. The size of bubbles indicates log (fold change). A darker color indicates a smaller p-value. The bubble with a black outline indicates the feature was identified from METLIN database, while bubble without a black outline indicates no identification was found. It can be seen from Fig 7.1A, 533 regulated features were detected from tryptone *vs* glucose comparison of P2, 333 of them were up-regulated and 200 were down regulated with fold change ≥1.5 and p-value <0.01.

Table 7.1 Statistic analysis using online XCMS

	Comparisons	**Changed/All identified features [a]**	**Changed/All quantified proteins [b]**
Comparison 1 (Carbon source: tryptone *vs* glucose)	P2	533/3740	159/583
	PBL2025	425/3502	208/609
	PBL2073	430/3660	-
	Merge data of P2, PBL2025 and PBL2073	652/3100	-
Comparison 2 Different strains grown on glucose	P2 *vs* PBL2025	690/3740	158/583
	P2 *vs* PBL2073	830/3772	-
	PBL2025 *vs* PBL2073	219/3573	81/609
	P2 *vs* PBL2025 and PBL2073	787/3531	-
Comparison 3 Different strains grown on tryptone	P2 *vs* PBL2025	105/3353	-
	P2 *vs* PBL2073	356/3580	-
	PBL2025 *vs* PBL2073	152/3313	43/450
	P2 *vs* PBL2025 and PBL2073	310/3392	-

a) *Statistically significant changed features, P<0.01, and fold change ≥1.5 were applied. b) Significantly changed proteins with P<0.01 (data from Chapter 4).*

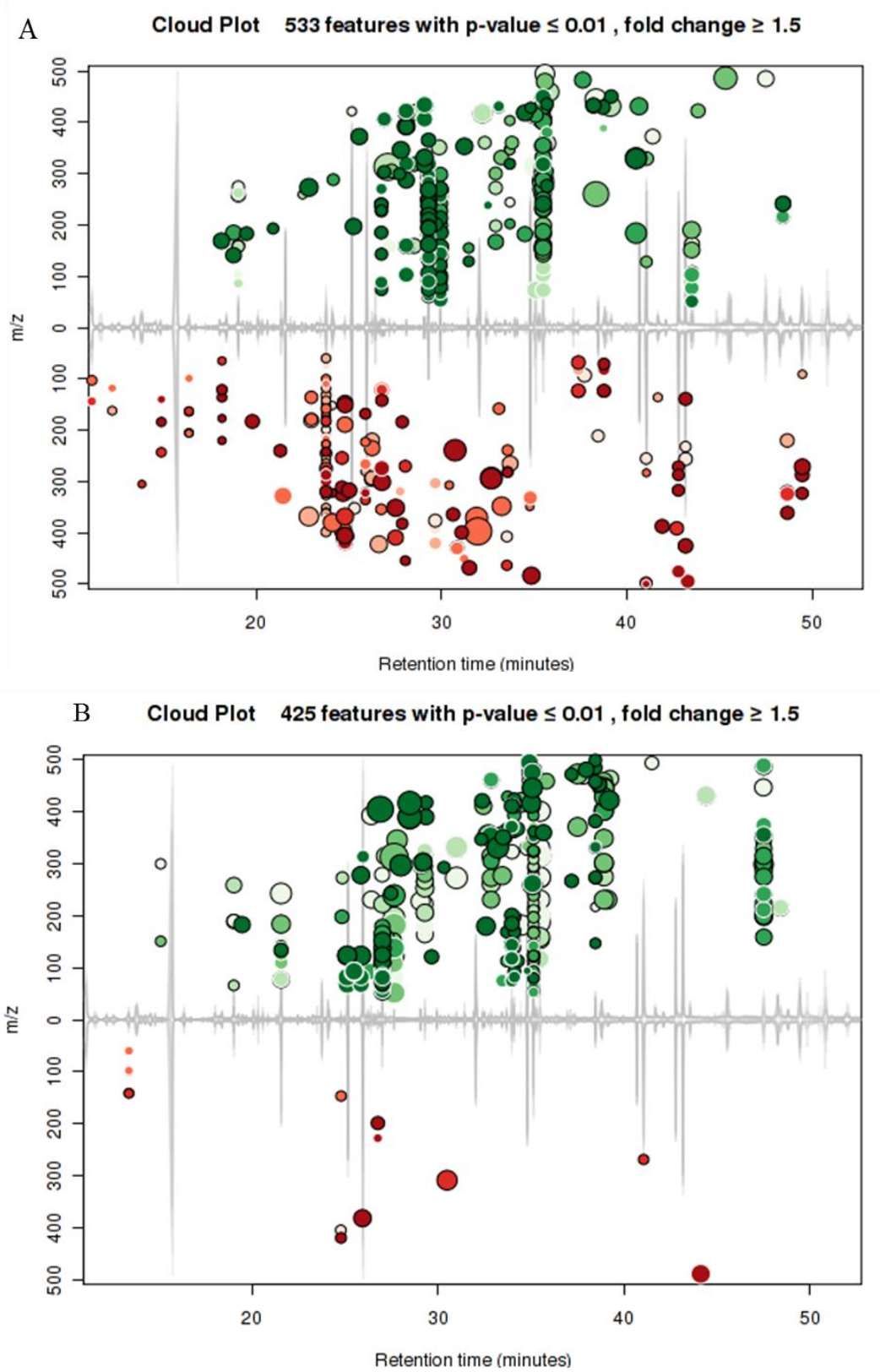

Fig 7.1 Cloud plot from XCMS statistical analysis for (A) P2 grown on glucose compared to tryptone. (B) PBL2025 grown on glucose compared to tryptone.

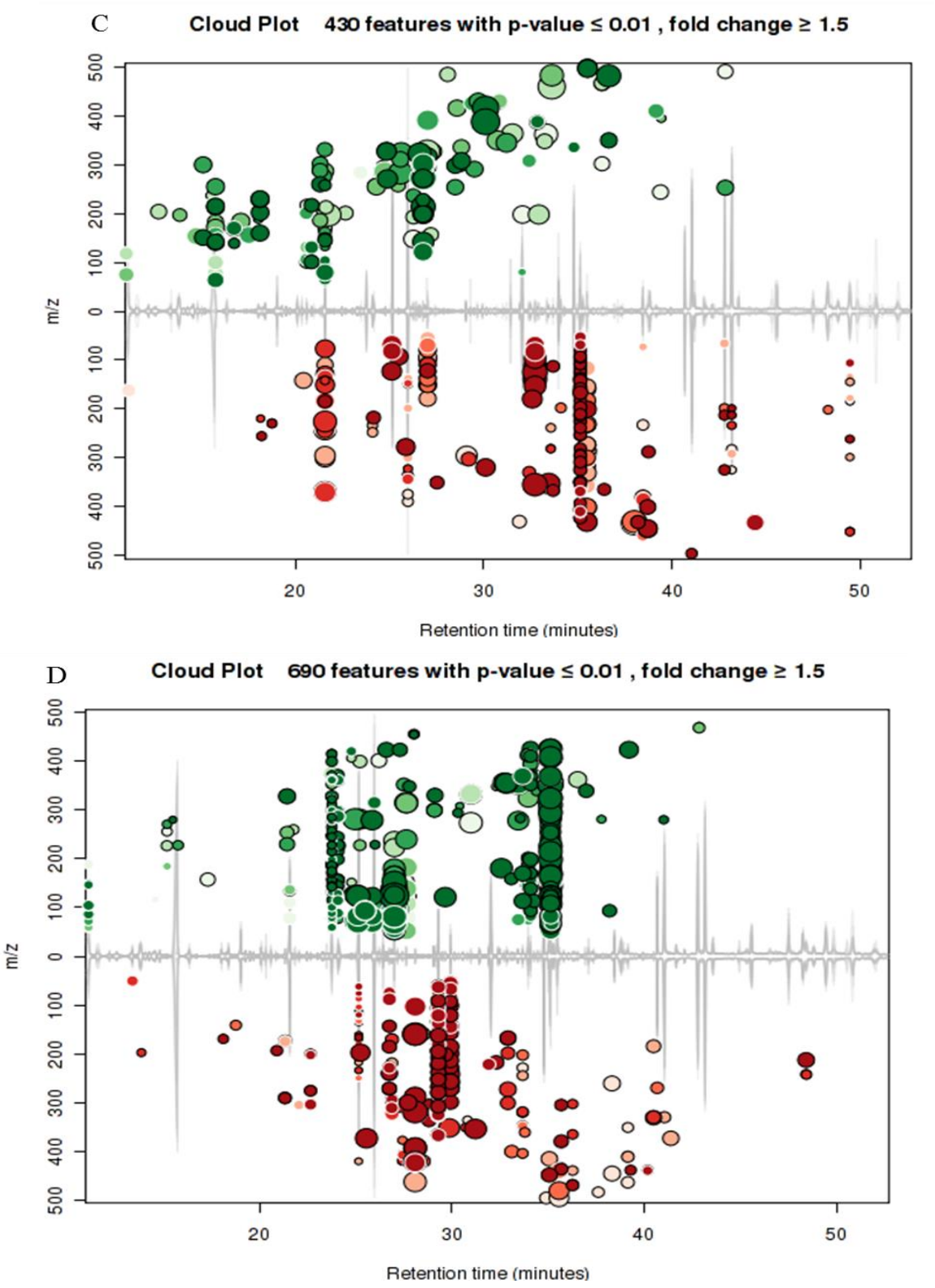

Fig 7.1 Cloud plot from XCMS statistical analysis for (C) PBL2073 grown on glucose compared to tryptone. (D) Strain PBL2025 compared to P2 both grown on glucose.

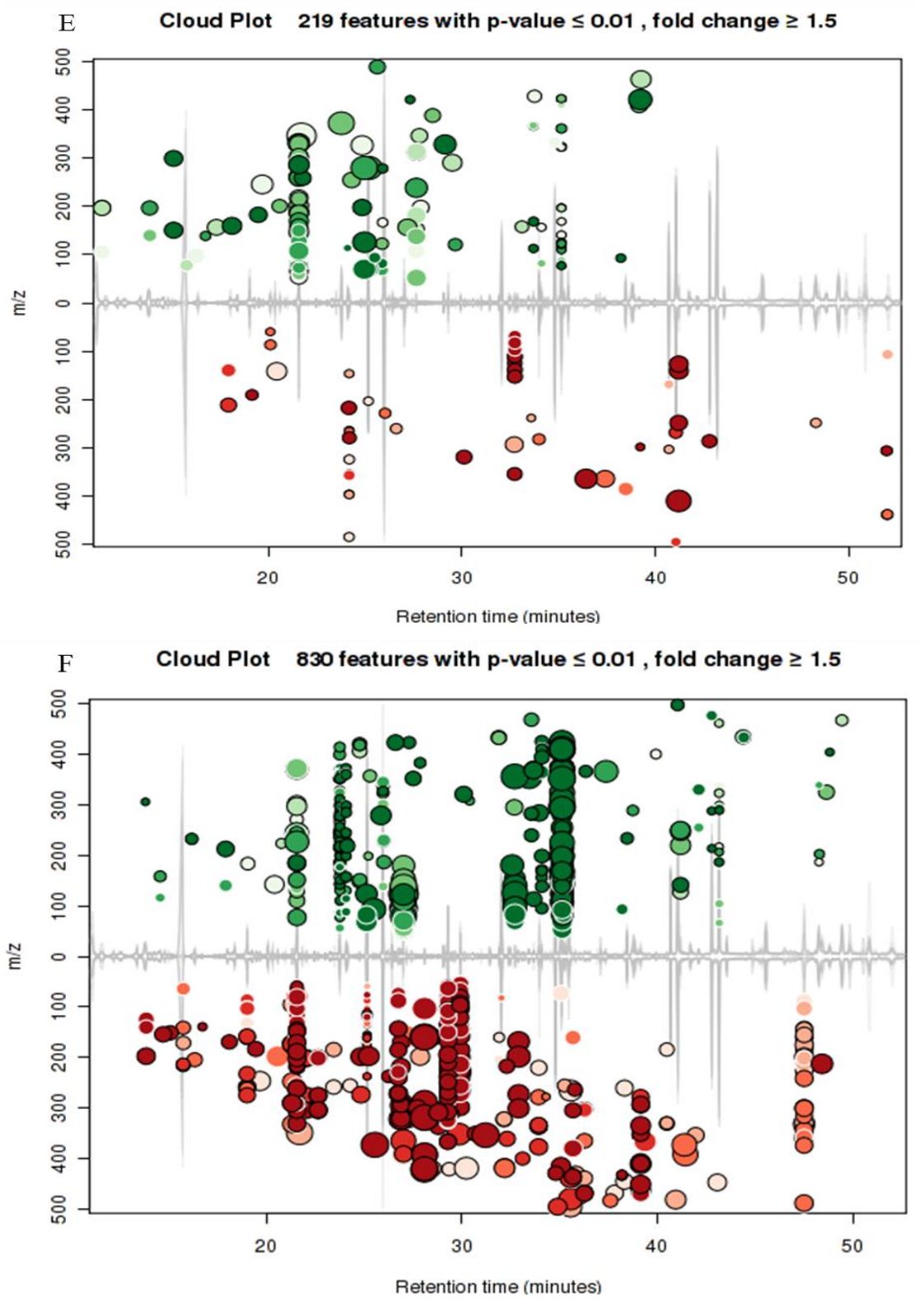

Fig 7.1 Cloud plot from XCMS statistical analysis for (E) strain PBL2025 compared to PBL2073 both grown on glucose. (F) Strain PBL2073 compared to P2 both grown on glucose.

7.4.2 Compounds identification

There is no software available that can provide both metabolite annotation and statistical analysis for GC-MS data from microorganism. Therefore, annotation of metabolites was achieved by using GC-MS raw data searched against AMDIS. The raw results were saved in .txt format. Further analysis were performed using Excel (Microsoft 2013, USA). These results were firstly simplified by Macros program to remove deviration. To gain high confidence identification, only compounds detected from all biological triplicates per phenotype were used (as shown in Fig 6.2), and others were discarded. In total, 163 compounds were identified from all experiments and 113 of them were assighed a KEGG C numer (results are listed in Table SP 6.1 in supplementary materials), where only 15 of them have been reported in previous study [38].

Table 7.2 A summary number of compounds identified.

Carbon source	Strains	Replicates				Merge data	Merge data of 3 strains
		3	4	5	6		
Glucose	P2	2	7	81	11	101	
	PBL2025	2	20	18	84	124	140
	PBL2073	0	20	74	0	94	
Tryptone	P2	5	11	20	92	128	
	PBL2025	2	29	68	0	99	141
	PBL2073	3	21	92	0	116	

Notes: Biological triplicates and technical duplicates were run for each experiment. The number 3-6 indicates that metabolites were detected from 3 (biological triplicates), or 4, or 5, or 6 GC-MS runs. Metabolites determined only from 1 or 2 biological samples are not considered.

7.4.2.1 *S.solfataricus* strains grown on different carbon sources

A similar coumpounds detection number was observed for P2, PBL2025 and PBL2073 strains: 140, 133 and 140, respectively as shown in Fig 7.2 A-D. Among these, most of the compounds (84%) were common (as shown in Fig 7.2 A) between glucose and tryptone growth media. The unique number of compounds detected from glucose *vs* tryptone varied between different strains (as indicated in Figs 7.2 B-D): a percentage of 12%, 27.4%, and 26% for P2, PBL2025 and PBL2073 respectively.

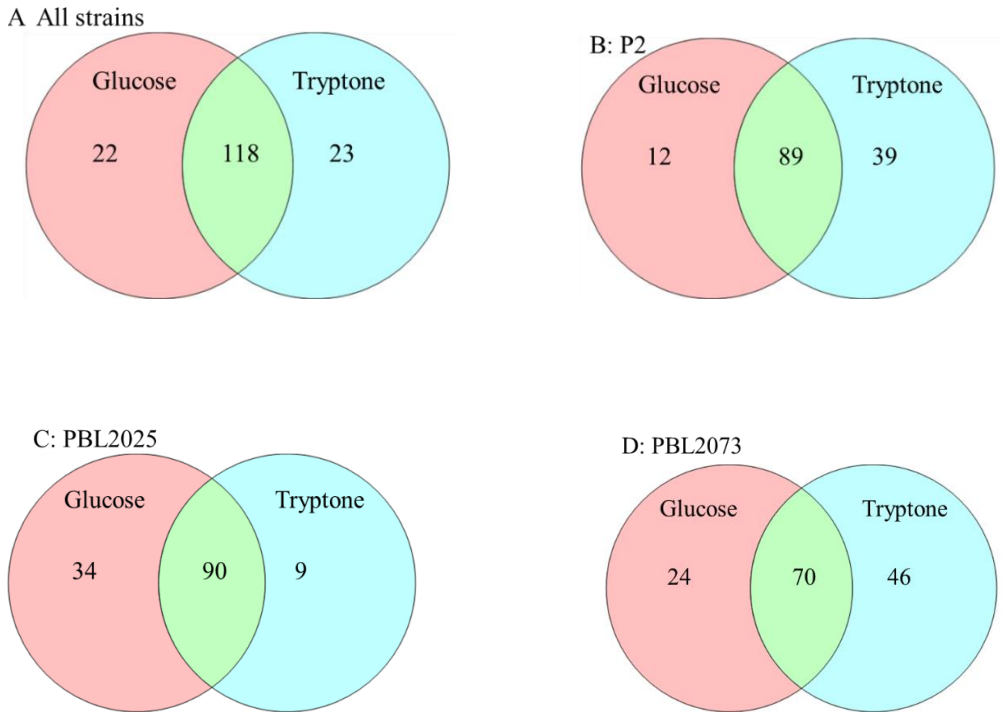

Fig 7.2 Number of compounds determined from different carbon sources. (A) summary of all strains. (B) for P2 strain. (C) for PBL2025 strain and (D) for PBL2073 strain.

7.4.2.2 Different strains grown on glucose or tryptone

The numbers of compounds determined for different strains grown either on glucose or tryptone are listed and interpreted in Venn diagram formats that can be found in Figs 7.3 A-C. For instance, compared to P2, a high percentage ((12+19+8)/140=27.8%, Fig 7.3B) of unique compounds were detected in PBL2025 and PBL2073 strains when all the strains were grown on glucose. In contrast, only 9.2% unique compounds were detected on tryptone from these strains (Fig 7.3C). It the directly utilisation of amino acids

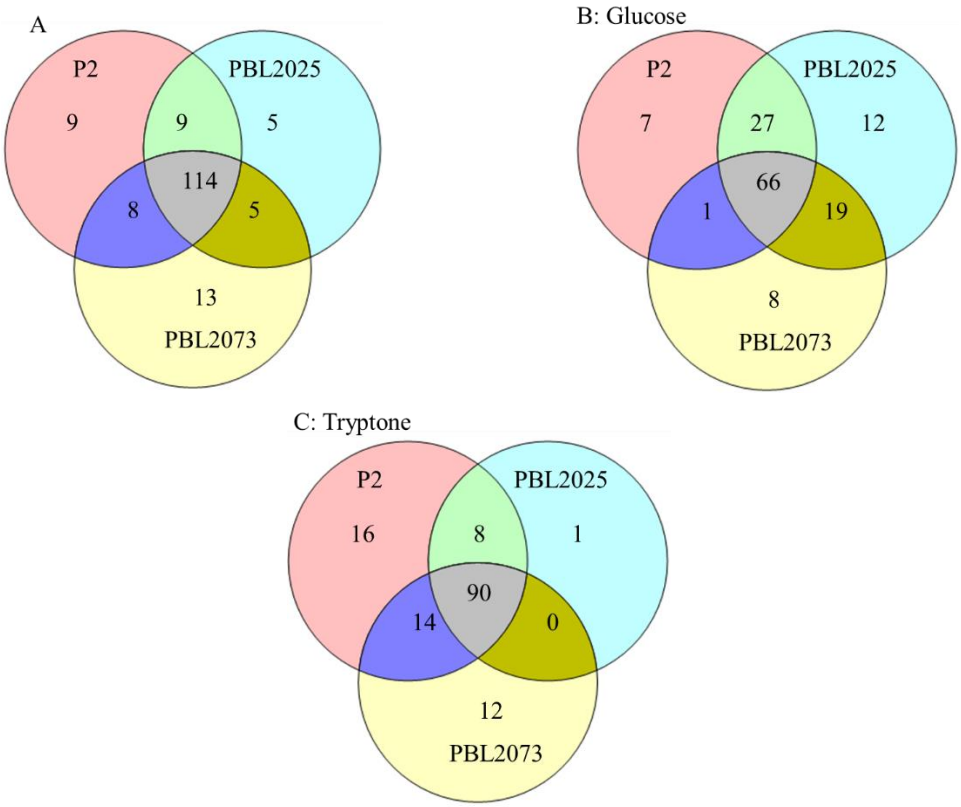

Fig 7.3 Compounds identified from different *S. solfataricus* strains (A) when cells grown on glucose and tryptone, (B) strains grown on glucose only and (C) cells grown on tryptone only.

7.4.3 Metabolites plotted into KEGG pathway

Among 163 compounds detcetd from all strains, only 113 of them were assigned C numbers, which were plotted against KEGG pathway. 47 of them were involved in metabolic pathways. Our experiments identified many new metabolites (listed in Table TS 6.1 in the Appendix) that has not been reported in *S solfataricus* [38]. However, these results are for identification only, further validation needs to be carried out.

7.4.3.1 Responses of *S.solfataricus* grown on glucose *vs* tryptone

Among 140 compounds detected from tryptone *vs* glucose for P2 strain, 24 reduced their abundances and most of them 116 (82.9%) increased their abundances when cultivated in tryptone. A total of 15 amino acids and a few belonging to KEGG metabolic pathways were detected by GC-MS, as listed in Table 7.3.

Most of detected amino acids except glycine increased their abundance. In combination with the proteomic data from Chapter 4 (Section 4.4.4.2), it can be reasonable to hypothesis that amino acid followed a negative feedback-inhibition pattern. For instance, the accumulation of

Gln inhibited glutamine synthetase (encoded by *SSO0366*, down regulation of this enzyme was detected in Chapter 4) activity. A high amount of intermedia metabolites (methylglutamic acid and pyroglutamic acid) might indicate the acceleration of reversible conversion between glutamate and α-ketoglutaric acid. It is supported by proteomic data (Fig 4.7A in Chapter 4), where the up regulation of NAD specific glutamate dehydrogenase (gdhA-2 and gdhA-4 encoded by *SSO1907* and *SSO2044*) was also observed. The presence of α-ketoglutaric acid was not detected for cells grown on tryptone media in late exponential growth phase. As a consequence, cell could not survive due to lack of available carbon/energy source. Examples of amino acids involving in KEGG pathways (aminoacyl-tRNA biosynthesis and ABC transporters) are shown in Figs 7.4 A and B.

Table 7.2 Comparisons of different carbon sources for different strains

Category	Metabolities	Glucose			Tryptone			Tryptone/Glucose		
		2025/P2	2073/ 2025	2073/ P2	2073/ 2025	2073/P2	2025/ P2	P2	PBL2025	PBL2073
Metabolic intermediates	Acetyl-L-alanine	1.79	0	0	~	3.33	0	0.99	~	0
	3-Iodo-L-tyrosine	2.21	0	0	NF	NF	0	3.01	0	NF
	Methylglutamic acid	NF	NF	NF	0.94	1.17	1.25	~	~	~
	Pyroglutamic acid	6.64	0	0	0.94	4.58	4.87	5.93	4.35	~
	1-Methylguanine	4.63	0.15	0.678	0.78	1.47	1.88	2.31	0.93	4.99
Amino acids	Alanine	6.21	0	0	0.82	1.11	1.36	2.76	0.6	~
	Glutamic acid	3.84	0	0	1.82	4.3	2.37	4.67	2.88	~
	Glutamine	5.22	0.03	0.14	0	NF	0	7.05	0	0
	Glycine	2.58	0	0	0.21	1.03	4.89	0.82	1.55	~
	Leucine	2.18	0	0	0.79	0.84	1.06	2.09	1.02	~
	Phenylalanine	1.37	0	0	~	1.12	0	2.01	0	~
	Proline	2.18	0	0	0.44	0.84	1.92	2.09	~	1.84
	Serine	2.46	0	0	0.32	0.1	0.3	13.74	1.68	~
	Threonine	6.89	0	0	NF	0	0	2.42	0	NF
	Tyrosine	2.23	0	0	2.29	3.01	1.31	2.99	1.76	~
	Tryptophan	0	0	0	~	1.16	0	2.46	NF	~
Other metabolites	Acetic acid	2.51	0.24	0.61	0.51	1.77	3.47	1.01	1.39	2.93
	Glyceraldehyde	0	NF	0	NF	NF	NF	0	NF	NF
	α-ketoglutaric acid	3.56	0.28	1	NF	NF	NF	0	0	0
	Lactic acid	5.89	0.13	0.76	0.41	1.48	3.6	2.61	1.6	5.11
	Urea	2.24	0.29	0.65	~	~	NF	0	0	0.91

Note: 0 indicates that metabolites only not detected in 'divisor' condition. ~ indicates that metabolite is not detected in 'dividended' condition. NF indicates not found. The 'area' was normalized using external standard Nor-valine and ratio was calculated.

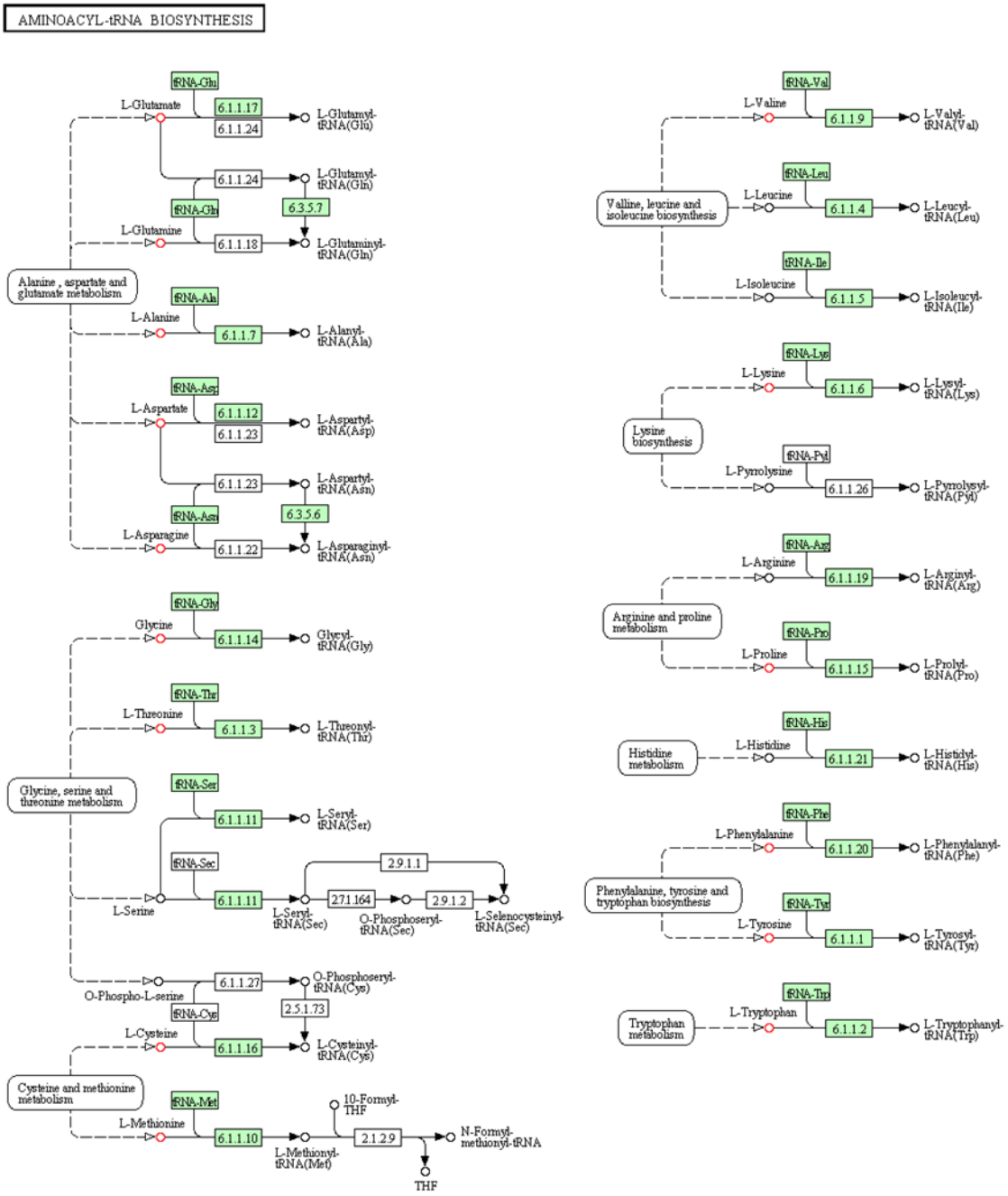

Fig 7.4 Detected metabolites involving in (A) Aminoacyl-tRNA biosynthesis in *S. solfataricus* P2 strain for tryptone *vs* glucose. Figure was obtained by plotting C number into KEGG pathways [300]. Red circle indicates metabolites detected in this Chapter.

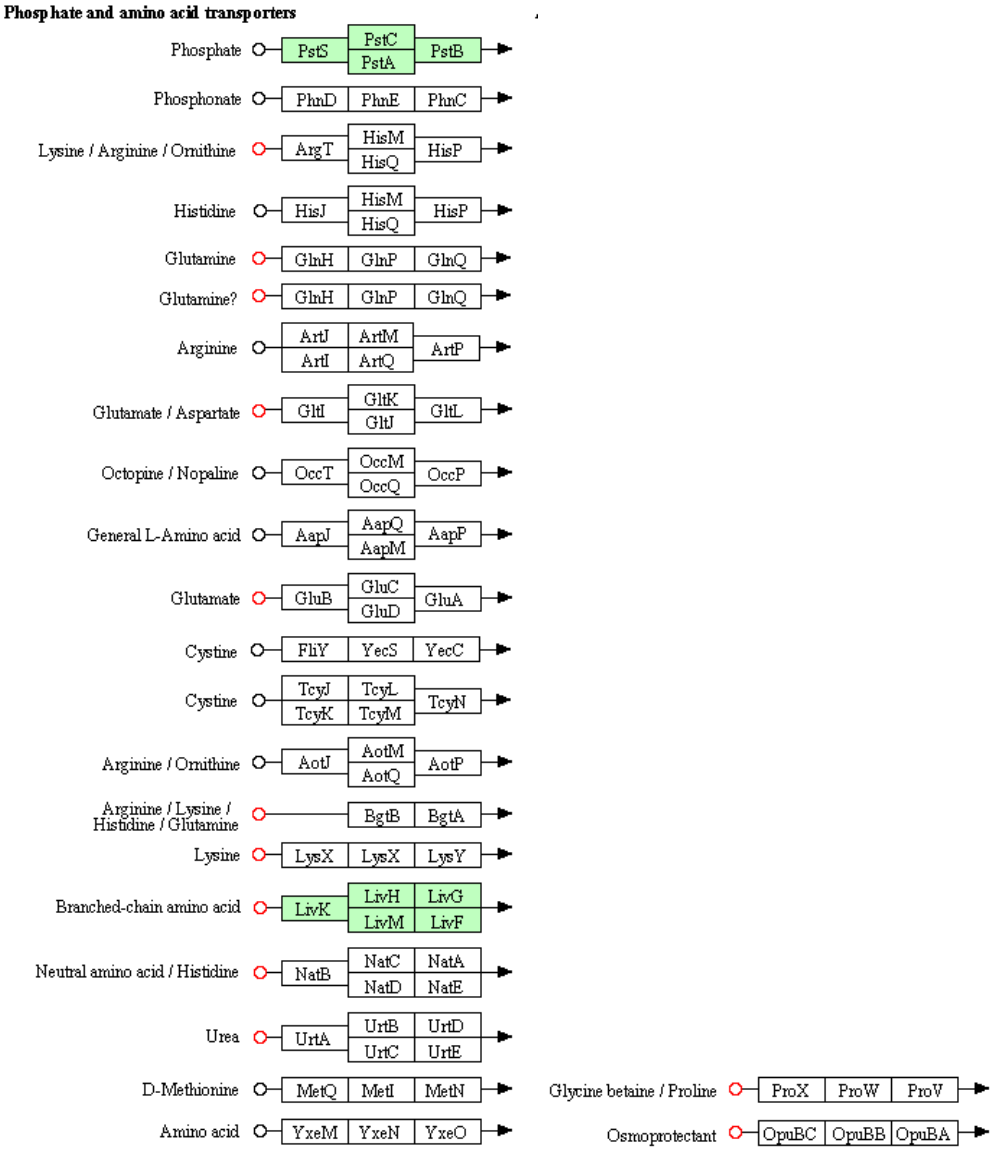

Fig 7.4 Metabolites involved in ABC transporters (B). Figure was reproduced from KEGG pathway.

7.5 Conclusions

Global metabolomics study was performed by GC-MS to investigate the affected metabolites in *S. solfataricus* by changing carbon source from glucose to tryptone. Furthermore, the changes between different strains grown on the same carbon sources were also investigated. Most of detected metabolites were overlapped between different experiments. Most of identified amino acids increased their abundances such as Glu, Pro, Ser, and Tyr when cells were grown on tryptone compared to glucose. High abundances of detected metabolic intermideate, for instance, methyglutamic acid and pyroglutamic acid were observed when tryptone was supplied for all the strains. The down regulation of amino acid biosynthesis for tryptone *vs* glucose of *S. solfataricus* P2 and PBL2025 (proteomic data from Chapter 4) can be explained by the accumulation of end products (amino acids), which will activate the negative

feedback-inhibition regulation mechanism. Besides, almost all the detected amino acids except Trp and listed metabolic intermediates show high abundances in PBL2025 compared to P2, when glucose was the sole carbon source for both strains. Interestingly, a reduced abundances of the above mentioned metabolites were observed for PBL2073 *vs* P2 and PBL2073 *vs* PBL2025 comparisons with the presence of glucose.

Chapter 8 Conclusions and future work

8.1 Conclusions

This thesis investigated the responses of three different *S. solfataricus* strains (P2, PBL2025 and PBL2073) to various carbon sources (glucose, tryptone, acetone, ethanol, iso-propanol). However, -omics techniques including proteomic (iTRAQ) and metabolomic (GC-MS) strategies were only applied to unravel the respectively change from glucose to tryptone of three strains due to limit growth of mutant strains (PBL2025 and PBL2073) on the other tested carbon sources. Moreover, P-peptide enrichment protocol was established for *S. solfataricus* P2. Compared to other approaches, for instance, PAcIFIC [8], the MS running and corresponding data analysis time is shorter using enrichment strategy: 2 weeks *vs* 8 weeks for sample running. Furthermore, a combination use with label-based quantitative proteomics technologies such as iTRAQ made the quantitative information of phosphoproteins available. As a result, quantitative phosphoproteomics study was performed on model strain P2 to investigate the protein phosphorylation changes.

Response of *S. solfataricus* strains to carbon sources change from glucose *vs* tryptone were investigated in proteomics (Chapter 4) and metabolomics (Chapter 7) level. iTRAQ analyses led to the quantification of 583, 609 and 450 proteins with more than 2 unique peptides for 1st, 2nd, and 3rd iTRAQ experiments respectively. However, the comparisons between different iTRAQ sets were not analysed due to bioinformatic technique limitation. An investigation of the global proteome level changes between PBL2025 and P2 was carried out, as a result, 158 proteins (27% of quantified proteins) showed their different abundances when both strains grown on standard glucose media. Of these proteins, more than half (61%) of regulated proteins involving in carbon fixation, butanoate and so forth showed their up regulation, whereas the others (39%) belonging to carbohydrate or amino acid metabolic pathways showed their down regulation. Moreover, 208 and 159 (27% and 34% of quantified proteins) proteins were affected under tryptone *vs* glucose for two strains PBL2025 and P2 respectively. The down regulation of pyruvate and glycolysis metabolism agreed a previous report that sugar degradation was blocked when tryptone was supplied in P2 [7]. However, the glycolysis pathway seemed to be up-regulated in PBL2025 strain in the presence of tryptone compared to glucose. Moreover, the down regulation of detected amino acid pathways was significant in both strains in responding to carbon sources change from glucose to tryptone. This study

provides useful quantitative proteomics data for metabolic pathways manipulation of the model archaea *S. solfataricus* P2 and PBL2025.

For global metabolomic analysis, based on GC-MS technique, 113 metabolites were assigned with C numbers from all experiments, whereas only 47 of them were involved in KEGG pathways. High abundances of detected amino acids and metabolic intermediate such as Glu, Pro, Ser, and Tyr, methyglutamic acid and pyroglutamic were observed when cells were grown on tryptone compared to glucose from 3 strains. Almost all detected amino acids except Trp showed their high abundance in PBL2025 compared to P2 when glucose was used as a sole carbon source for both strains. Global metabolic and proteomic datasets showed that, amino acid metabolic pathways were down regulated when tryptone was supplied, which is inferred to follow a feedback-inhibited pattern. More solid conclusions cannot be made in -omics level for other comparisons due to limit detection of metabolites involving in carbohydrate pathways as well as unavailability of comparisons between iTRAQ sets at the moment.

Furthermore, P-peptides enrichment strategies were optimised for global quantitative phospohproteomic study. TiO_2 beads outperformed other tested formats (spin and pipette tips). Also, 4 cycles of incubation of beads with a 20/1 (w/v) peptide-to-TiO_2 beads ratio were optimal condition for P-peptide enrichment of *S. solfataricus*. In addition, majority of P-peptides with single P-sites were detected using TiO_2 (81.4%), the combination use of TiO_2 was recommended than the usage of single one. Moreover, almost twice more triply charged peptides were observed for P-peptides than that for all identified peptides by all of the tested MS instruments: ion trap (HCT-Ultra), the Amazon and the maXis UHR-Q-TOF. It indicated that the possibility to get a better phosphoproteome coverage by using both CID and ETD fragmentations, since ETD has been shown to be beneficial to highly-charged peptides detection [205].

Optimised enrichment methods were applied for quantitative phosphoproteomic analysis of *S. solfataricus* P2 grown on glucose compared to tryptone conditions. A total of 91 P-sites (phospho sites) were determined from 61 P-proteins (phospho proteins) contributed from 73 unique P-peptides. The distribution of P-sites on Ser/Thr/Tyr was 36%/28%/35%. Of these quantified P-proteins, 19 were distributed into 31 KEGG pathways and 4 of these quantified P-proteins were differentially regulated in responding to different growth conditions (from glucose to tryptone). The regulated P-protein: Zn-dependent alcohol dehydrogenase (ADH-2,

encoded by *SSO0764*) may be of biological as well as industrial interest. ADH enzymes are important catalysts in alcohol and acetone metabolism, it shows great commeicial value.

8.2 Future work

Biological samples (cell pellets) were obtained from the same growth conditions. Therefore, -omics (proteomics, phosphoproteomics and metalobolomics) data generated in this thesis are consistant, which can be used for further modelling work in regarding to the response of this hyperthomophilic archaeon under various carbon sources (glucose *vs* tryptone). The link between carbon source variance and temperature changes will broaden the CCM network studies for community [301].

P-peptide enrichment efficiency was less than 30%, which is still quite low, compared to which from human samples. From technical aspects for phosphoproteome coverage improvement, the application of Ti^{4+}-IMAC, or micro size TiO_2 beads, ERLIC separation or a combination with different MS fragmentation strategies such as CID with ETD may be useful. However, the number of biological replicates need to be considered for iTRAQ or TMT quantitative studies. For instance, only three of 4-plex [61], five of 8-plex iTRAQ labels [62] and four of 6-plex TMT tags [63] give rise to unique reporter ions under ETD mode that would allow for quantification, respectively. Moreover, localization of all identified P-proteins from enrichment optimisation (178 P-proteins in Chapter 5), control (5 P-proteins) and Esser, *et al*., (540 P-proteins), 19.6% of them were located in membrane (16.8% of cytoplasmic membrane and 2.7% of outer membrane). In contrast, only 13% of quantified P-proteins were located in membrane in our study (Chapter 6). Improvements of phosphoproteome coverage maybe achieved from protein extraction aspects by incorporating different extraction buffers such as SDC, which have been proven to benefit for quantitative membrane proteome study as investigated by [302]. The successful application of P-peptides enrichment with iTRAQ labelling technique in *S. solfatricus* P2 indicates that further work can be perfomed on other *Sulfolobus* species, such as *S. acidocaldarirus*, since hundreds of P-proteins were determined from wild and mutant strains [47]. Furthermore, the application in other archaeal researches might be possible if the successful application in *Sulfolobus* species is achieved.

Among the encoding genes from *SSO3004 to SSO3050* that were not presented in PBL2025 strain, 7 of them were quantified in *S. solfataricus* P2, including SSO3003, SSO3007, SSO3009, SSO3019, SSO3038, SSO3039, and SSO3042, and three of these proteins (carbon monoxide

dehydrogenase encoded by *SSO3009*, α-xylosidase SSO3022, and uncharacterised protein SSO3038) were observed as P-proteins. It would be interesting to investigate the phospho proteome difference between PBL2025 and P2, which will benefit for biofilm studies. The protein-protein interactions between the quantified P-proteins with identified P-proteins from [7] and our enrichment optimisation experiments would enable the studies on uncharacterised, predicted or hypothetical proteins. Some regulated P-proteins such as SSO0625 is promising candidate for further studies since SSO0625 predicted bo be GTPase, for which the biological function of any *S. solfataricus* GTP-binding proteins (SsGBPs) is unknown.

In addition, further research on cross-talks between different PTMs such as phosphorylation and acetylation will be of interest. Impaired effect on phosphorylation from acetylation have been reported in 14-3-3 proteins by Lys sites mutation strategy [88]. The acetylation of 17 proteins [294] and phosphorylation of hundreds of proteins [7] in *S. solfataricus* have been reported. Proteins reported to be both phosphorylated and acetylated will be mainly focus on PTM cross-talk studies in the future.

However, quantitative phosphoproteome data from Chapter 6 will need further validations. For instance, Western blot, anti phosphotyrosine antibody-based immunohistochemistry, radioactive P^{32} or synthetic P-peptides, before going to biological mutation work in genomic level. Furthermore, not many investigations on signal pathways in *S. solfataricus* have been performed, it will be fascinating to construct signal maps in this archaeon based on phosphorylation studies.

In addition, comparisons between different iTRAQ datasets are not available, which relies on bioinformatics tools. Co-operation work with researchers who have had background is ongoing, it would be fascinating to interpret the links and difference between different comparisons, for instance, PBL2025 compared to P2 grown on tryptone, or comparison of three different strains utilising the same carbon source. Moreover, not many metabolites detected in this thesis belonging to central carbohydrate metabolic pathways, therefore, to improve metabolome coverage, further work can be carried out using different extraction buffers coupled with LC-MS, for instance, high resolution UHPLC coupled with MS or NMR detections.

References

1. Brock, T.D., et al., *Sulfolobus: a new genus of sulfur-oxidizing Bacteria living at low pH and high temperature.* Arch Mikrobiol 1972. **84**(1): p. 54-68.
2. Zillig, W., et al., *The Sulfolobus-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases* Arch. Microbiol, 1980. **125**(3): p. 259-269.
3. She, Q., et al., *The complete genome of the crenarchaeon Sulfolobus solfataricus P2.* Proc Natl Acad Sci U S A., 2001. **98**(14): p. 7835-7840.
4. Kort, J.C., et al., *A cool tool for hot and sour Archaea: Proteomics of Sulfolobus solfataricus.* Proteomics, 2013. **13**(18-19): p. 2831-2850.
5. She, Q., et al., *The complete genome of the crenarchaeon Sulfolobus solfataricus P2.* Proc Natl Acad Sci U S A, 2001. **98**(14): p. 7835-7840.
6. Esser, D., et al., *Functional curation of the Sulfolobus solfataricus P2 and S. acidocaldarius 98-3 complete genome sequences.* Extremophiles, 2011. **15**(6): p. 711-712.
7. Esser, D., et al., *Change of carbon source causes dramatic effects in the phospho-proteome of the Archaeon Sulfolobus solfataricus.* J Proteome Res, 2012. **11**(10): p. 4823-4833.
8. Panchaud, A., et al., *Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean.* Analytical chemistry, 2009. **81**(15): p. 6481-6488.
9. Hunter, T., *Signaling—2000 and Beyond* Cell, 2000. **100**(1): p. 113-127.
10. Li, D.-C., et al., *Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, Sulfolobus solfataricus P2.* Cell Stress and Chaperones, 2012. **17**(1): p. 103-108.
11. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proceedings of the National Academy of Sciences, 1977. **74**(11): p. 5088.
12. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proceedings of the National Academy of Sciences, 1990. **87**(12): p. 4576.
13. Gribaldo, S. and C. Brochier-Armanet, *The origin and evolution of Archaea: a state of the art.* Philos Trans R Soc Lond B Biol Sci, 2006. **361**(1470): p. 1007-1022.
14. Barns, S.M., et al., *Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences.* Proc Natl Acad Sci U S A, 1996. **93**(17): p. 9188-9193.
15. Brock, T.D., et al., *Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature.* Archiv für Mikrobiologie, 1972. **84**(1): p. 54-68.
16. Leigh, J.A., et al., *Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales.* FEMS microbiology reviews, 2011. **35**(4): p. 577-608.
17. Chen, L., et al., *The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota.* Journal of bacteriology, 2005. **187**(14): p. 4992-4999.
18. Kawarabayasi, Y., et al., *Complete genome sequence of an aerobic thermoacidophilic.* DNA Res, 2001. **8**(4): p. 123-140.
19. Frock, A.D. and R.M. Kelly, *Extreme thermophiles: moving beyond single-enzyme biocatalysis.* Current Opinion in Chemical Engineering, 2012.
20. Grogan, D., *Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains.* J Bacteriol, 1989. **171**: p. 6710-6719.
21. Hochstein, L.I. and H. Stan-Lotter, *Purification and properties of an ATPase from Sulfolobus solfataricus.* Archives of biochemistry and biophysics, 1992. **295**(1): p. 153-160.

22. Rolfsmeier, M. and P. Blum, *Purification and characterization of a maltase from the extremely thermophilic crenarchaeote Sulfolobus solfataricus.* Journal of bacteriology, 1995. **177**(2): p. 482.
23. Cavicchioli, R., *Archaea — timeline of the third domain.* Nat Rev Microbiol, 2011. **9**(1): p. 51-61.
24. Kristjansson, K.J. and G.O. Hreggvidsson, *Ecology and habitats of extremophiles.* . World J.Microbiol.Biotechnol, 1995. **11**(1): p. 17-25.
25. Egorova, K. and G. Antranikian, *Industrial relevance of thermophilic Archaea.* Curr Opin Microbiol, 2005. **8**(6): p. 649-655.
26. Antranikian, G., C.E. Vorgias, and C. Bertoldo, *Extreme environments as a resource for microorganisms and novel biocatalysts.* Adv Biochem Eng Biotechnol, 2005. **96**: p. 219-262.
27. de Miguel Bouzas, T., J. Barros-Velázquez, and T.G. Villa, *Industrial applications of hyperthermophilic enzymes: a review.* Protein Pept Lett., 2006. **13**(7): p. 645-651.
28. Orell, A., et al., *Life in blue: copper resistance mechanisms of bacteria and archaea used in industrial biomining of minerals.* Biotechnol Adv, 2010. **28**(6): p. 839-848.
29. Joo, W.A. and C.W. Kim, *Proteomics of Halophilic archaea.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **815**(1-2): p. 237-250.
30. Gray, N.D., et al., *Occurrence and activity of Archaea in aerated activated sludge wastewater treatment plants.* Environ Microbiol, 2002. **4**(3): p. 158-168.
31. Blumer-Schuette, S.E., et al., *Extremely thermophilic microorganisms for biomass conversion: status and prospects.* Current Opinion in Biotechnology, 2008. **19**(3): p. 210-217.
32. Morana, A., et al., *Gene cloning and expression in Escherichia coli of a bi-functional β-D-xylosidase/α-L-arabinosidase from Sulfolobus solfataricus involved in xylan degradation.* Extremophiles, 2007. **11**(1): p. 123-132.
33. Bustard, M.T., et al., *Biodegradation of high-concentration isopropanol by a solvent-tolerant thermophile, Bacillus pallidus.* Extremophiles., 2002. **6**(4): p. 319-323.
34. Radianingtyas, H. and P.C. Wright, *2-Propanol degradation by Sulfolobus solfataricus.* Biotechnol Lett., 2003. **25**(7): p. 579-583.
35. Chong, P.K., et al., *Translational and transcriptional analysis of Sulfolobus solfataricus P2 to provide insights into alcohol and ketone utilisation.* Proteomics, 2007. **7**(3): p. 424-435.
36. Izzo, V., et al., *The thermophilic archaeon Sulfolobus solfataricus is able to grow on phenol.* Research in microbiology, 2005. **156**(5): p. 677-689.
37. Christen, P., et al., *Phenol biodegradation by the thermoacidophilic archaeon Sulfolobus solfataricus 98/2 in a fed-batch bioreactor.* Biodegradation, 2011. **22**(3): p. 475-484.
38. Zaparty, M., et al., *"Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.* Extremophiles, 2010. **14**(1): p. 119-142.
39. Snijders, A.P., et al., *Reconstruction of central carbon metabolism in Sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis.* Proteomics, 2006. **6**(5): p. 1518-1529.
40. Chong, P.K., et al., *Proteome analysis of Sulfolobus solfataricus P2 propanol metabolism.* J Proteome Res, 2007. **6**(4): p. 1430-1439.
41. Chong, P.K., et al., *Proteome and transcriptional analysis of ethanol-grown Sulfolobus solfataricus P2 reveals ADH2, a potential alcohol dehydrogenase.* Journal of proteome research, 2007. **6**(10): p. 3985-3994.
42. Zolghadr, B., et al., *Appendage-mediated surface adherence of Sulfolobus solfataricus.* J Bacteriol 2010. **192**(1): p. 104-110.
43. Koerdt, A., et al., *Complementation of Sulfolobus solfataricus PBL2025 with an α-mannosidase: effects on surface attachment and biofilm formation.* Extremophiles, 2012. **16**(1): p. 115-125.

44. De Rosa, M., et al., *Glucose metabolism in the extreme thermoacidophilic archaebacterium Sulfolobus solfataricus.* Biochem J., 1984. **224**(2): p. 407-414.
45. Chong, P.K., et al., *Proteome and transcriptional analysis of ethanol-grown Sulfolobus solfataricus P2 reveals ADH2, a potential alcohol dehydrogenase.* J Proteome Res., 2007. **6**(10): p. 3985-3994.
46. Uhrigshardt, H., et al., *Evidence for an operative glyoxylate cycle in the thermoacidophilic crenarchaeon< i> Sulfolobus acidocaldarius</i>*. FEBS letters, 2002. **513**(2): p. 223-229.
47. Reimann, J., et al., *Archaeal Signal Transduction: Impact of Protein Phosphatase Deletions on Cell Size, Motility, and Energy Metabolism in Sulfolobus acidocaldarius.* Molecular & Cellular Proteomics, 2013. **12**(12): p. 3908-3923.
48. Kouril, T., et al., *Unraveling the function of the two Entner–Doudoroff branches in the thermoacidophilic Crenarchaeon Sulfolobus solfataricus P2.* FEBS Journal, 2013. **280**(4): p. 1126-1138.
49. Heinemann, J., et al., *Expanding the paradigm of thiol redox in the thermophilic root of life.* Biochimica et Biophysica Acta (BBA)-General Subjects, 2014. **1840**(1): p. 80-85.
50. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Molecular & cellular proteomics, 2004. **3**(12): p. 1154-1169.
51. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Analytical chemistry, 2003. **75**(8): p. 1895-1904.
52. Ong, S.-E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Molecular & cellular proteomics, 2002. **1**(5): p. 376-386.
53. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nature biotechnology, 1999. **17**(10): p. 994-999.
54. Mann, M., et al., *Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome.* Trends Biotechnol, 2002. **20**(6): p. 261-268.
55. Kanshin, E., S. Michnick, and P. Thibault. *Sample preparation and analytical strategies for large-scale phosphoproteomics experiments*. in *Seminars in cell & developmental biology*. 2012. Elsevier.
56. Loroch, S., et al., *Phosphoproteomics—More than meets the eye.* Electrophoresis, 2013. **34**(11): p. 1483-1492.
57. Palumbo, A.M., et al., *Tandem mass spectrometry strategies for phosphoproteome analysis.* Mass Spectrom Rev, 2011. **30**(4): p. 600-625.
58. Polat, A.N. and N. Özlü, *Towards single-cell LC-MS phosphoproteomics.* Analyst, 2014. **139**(19): p. 4733-4749.
59. Janek, K., et al., *Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry.* Rapid Communications in Mass Spectrometry, 2001. **15**(17): p. 1593-1599.
60. Villén, J., S.A. Beausoleil, and S.P. Gygi, *Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis.* Proteomics, 2008. **8**(21): p. 4444-4452.
61. Phanstiel, D., et al., *Peptide and protein quantification using iTRAQ with electron transfer dissociation.* Journal of the American Society for Mass Spectrometry, 2008. **19**(9): p. 1255-1262.
62. Phanstiel, D., et al., *Peptide quantification using 8-plex isobaric tags and electron transfer dissociation tandem mass spectrometry.* Analytical chemistry, 2009. **81**(4): p. 1693-1698.

63. Viner, R.I., et al., *Quantification of post-translationally modified peptides of bovine α-crystallin using tandem mass tags and electron transfer dissociation.* J Proteomics, 2009. **72**(5): p. 874-885.
64. Haseltine, C., et al., *Extragenic pleiotropic mutations that repress glycosyl hydrolase expression in the hyperthermophilic archaeon Sulfolobus solfataricus.* Genetics, 1999. **152**(4): p. 1353.
65. Schelert, J., et al., *Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.* Journal of bacteriology, 2004. **186**(2): p. 427.
66. Esser, D., et al., *Functional curation of the Sulfolobus solfataricus P2 and S. acidocaldarius 98-3 complete genome sequences.* Extremophiles., 2011. **[Epub ahead of print]**.
67. Lamble, H.J., et al., *Metabolic pathway promiscuity in the archaeon Sulfolobus solfataricus revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase.* Journal of Biological Chemistry, 2003. **278**(36): p. 34066.
68. Albers, S.-V., et al., *Insights into ABC transport in archaea.* Journal of bioenergetics and biomembranes, 2004. **36**(1): p. 5-15.
69. Gogliettino, M., et al., *A highly selective oligopeptide binding protein from the archaeon Sulfolobus solfataricus.* J Bacteriol 2010. **192**(12): p. 3123-3131.
70. Lalithambika, S., et al., *Carbohydrate hydrolysis and transport in the extreme thermoacidophile Sulfolobus solfataricus.* Appl. Environ. Microbiol, 2012. **78**(22): p. 7931-7938.
71. Koerdt, A., et al., *Complementation of Sulfolobus solfataricus PBL2025 with an α-mannosidase: effects on surface attachment and biofilm formation.* Extremophiles, 2011: p. 1-11.
72. Chong, P.K., et al., *Proteome analysis of Sulfolobus solfataricus P2 propanol metabolism.* Journal of proteome research, 2007. **6**(4): p. 1430-1439.
73. Brouns, S.J.J., et al., *Identification of the missing links in prokaryotic pentose oxidation pathways.* Journal of Biological Chemistry, 2006. **281**(37): p. 27378.
74. Nunn, C.E., et al., *Metabolism of pentose sugars in the hyperthermophilic archaea Sulfolobus solfataricus and Sulfolobus acidocaldarius.* J Biol Chem, 2010. **285**(44): p. 33701-33709.
75. Baldomà, L. and J. Aguilar, *Involvement of lactaldehyde dehydrogenase in several metabolic pathways of Escherichia coli K12.* J Biol Chem, 1987. **262**(29): p. 13991-13996.
76. Sridhara, S. and T.T. Wu, *Purification and properties of lactaldehyde dehydrogenase from Escherichia coli.* J Biol Chem, 1969. **244**(19): p. 5233-5238.
77. Inoue, Y., et al., *Metabolism of 2-oxoaldehydes in yeasts. Purification and characterization of lactaldehyde dehydrogenase from Saccharomyces cerevisiae.* Eur J Biochem, 1985. **153**(2): p. 243-247.
78. Abu-Qarn, M., J. Eichler, and N. Sharon, *Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea.* Curr Opin Struct Biol, 2008. **18**(5): p. 544-550.
79. Calo, D., L. Kaminski, and J. Eichler, *Protein glycosylation in Archaea: sweet and extreme.* Glycobiology, 2010. **20**(9): p. 1065-1076.
80. Jarrell, K.F., et al., *S-layer glycoproteins and flagellins: reporters of archaeal posttranslational modifications.* Archaea, 2010. **2010**(pii: 612948).
81. Larkin, A. and B. Imperiali, *The expanding horizons of asparagine-linked glycosylation.* Biochemistry, 2011. **50**(21): p. 4411-4426.
82. Mischerikow, N. and A.J. Heck, *Targeted large-scale analysis of protein acetylation.* Proteomics, 2011. **11**(4): p. 571-589.
83. Soppa, J., *Protein acetylation in archaea, bacteria, and eukaryotes.* Archaea, 2010. **2010**(pii: 820681).

84. Maupin-Furlow, J.A., et al., *Proteasomes: perspectives from the Archaea.* Front Biosci, 2004. **9**: p. 1743-1758.
85. Maupin-Furlow, J.A., et al., *Proteasomes from structure to function: perspectives from Archaea.* Curr Top Dev Biol, 2006. **75**(125-169).
86. Jers, C., et al., *Bacillus subtilis BY-kinase PtkA controls enzyme activity and localization of its protein substrates. Molecular Microbiology.77(2): 287-299.* Mol Microbiol 2010. **77**(2): p. 287-299.
87. Hunter, T., *The age of crosstalk: phosphorylation, ubiquitination, and beyond.* Mol Cell, 2007. **28**(730-738).
88. Choudhary, C., et al., *Lysine acetylation targets protein complexes and co-regulates major cellular functions.* Science, 2009. **325**(5942): p. 834-840.
89. Kirpekar, F., S. Douthwaite, and P. Roepstorff, *Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry.* RNA, 2000. **6**(2): p. 296-306.
90. Kvaratskhelia, M. and M.F. White, *Two Holliday junction resolving enzymes in Sulfolobus solfataricus.* J Mol Biol, 2000. **297**(4): p. 923-932.
91. Moracci, M., et al., *Identification and molecular characterization of the first alpha -xylosidase from an archaeon.* J Biol Chem., 2000. **275**(29): p. 22082-22089.
92. Jungblut, P.R. and H. Schlüter, *Towards the analysis of protein species: an overview about strategies and methods.* Amino Acids., 2011. **41**(2): p. 219-222.
93. Salzano, A.M. and M. Crescenzi, *Mass spectrometry for protein identification and the study of post translational modifications.* Ann Ist Super Sanita, 2005. **41**(4): p. 443-450.
94. Cozzone, A.J., *Protein phosphorylation in prokaryotes.* Annual Review of Microbiology, 1988. **42**: p. 97-125.
95. Stock, J.B., A.J. Ninfa, and A.M. Stock, *Protein phosphorylation and regulation of adaptive responses in bacteria.* Microbiological Reviews, 1989. **53**(4): p. 540-590.
96. Kennelly, P.J., *Protein Ser/Thr/Tyr Phosphorylation in the Archaea.* Journal of Biological Chemistry, 2014. **289**(14): p. 9480-9487.
97. Fischer, E.H., et al., *Structure of the site phosphorylated in the phosphorylase b to a reaction.* J Biol Chem, 1959. **234**(7): p. 1698-1704.
98. Cohen, P., *The origins of protein phosphorylation.* Nat Cell Biol, 2002. **4**(5): p. 127-130.
99. Kennelly, P.J., *Protein kinases and protein phosphatases in prokaryotes: a genomic perspective.* FEMS Microbiol Lett, 2002. **206**(1): p. 1-8.
100. Attwood, P.V., P.G. Besant, and M.J. Piggott, *Focus on phosphoaspartate and phosphoglutamate.* Amino Acids, 2011. **40**(4): p. 1035-1051.
101. Moorhead, G.B., et al., *Evolution of protein phosphatases in plants and animals.* Biochem J. , 2009. **417**(2): p. 401-409.
102. Ray, W.K., et al., *A Phosphohexomutase from the archaeon Sulfolobus solfataricus Is covalently modified by phosphorylation on serine.* J Bacteriol, 2005. **187**(12): p. 4270-275.
103. Aertsen, A. and C.W. Michiels, *Stress and how bacteria cope with death and survival.* Crit Rev Microbiol., 2004. **30**(4): p. 263-273.
104. Marles-Wright, J. and R.J. Lewis, *Stress responses of bacteria.* Curr Opin Struct Biol, 2007. **17**(6): p. 755-760.
105. Zhu, Q. and A.A. Wani, *Histone modifications: crucial elements for damage response and chromatin restoration.* J Cell Physiol, 2010. **223**(2): p. 283-299.
106. Kanamaru, K., *Roles of the His-Asp phosphorelay signal transduction system in controlling cell growth and development in Aspergillus nidulans.* Biosci Biotechnol Biochem, 2011. **75**(1): p. 1-6.

107. Duan, X. and Z.G. He, *Characterization of the specific interaction between archaeal FHA domain-containing protein and the promoter of a flagellar-like gene-cluster and its regulation by phosphorylation* Biochem Biophys Res Commun, 2011. **107**(1): p. 242-247.
108. Eichler, J. and M.W. Adams, *Posttranslational protein modification in Archaea.* Microbiology and Molecular Biology Reviews, 2005. **69**(3): p. 393-425.
109. Lipmann, F.A. and P.A. Levene, *Serinephosphoric acid obtained on hydrolysis of vitellinic acid.* Journal of Biological Chemistry, 1932. **98**(1): p. 109.
110. Burnett, G. and E.P. Kennedy, *The enzymatic phosphorylation of proteins.* J Biol Chem, 1954. **211**(2): p. 969-980.
111. Eckhart, W., M.A. Hutchison, and T. Hunter, *An activity phosphorylating tyrosine in polyomaT antigen immunoprecipitates.* Cell, 1979. **18**: p. 925-933.
112. Pawson, T. and J.D. Scott, *Protein phosphorylation in signaling—50 years and counting.* Trends Biochem. Sci. , 2005. **30**: p. 286-290.
113. Cohen, P., *The regulation of protein function by multisite phosphorylation – a 25 year update.* Trends Biochem Sci, 2000. **25**(12): p. 596-601.
114. Macek, B., et al., *The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis.* Mol. Cell. Proteomics, 2007. **6**: p. 697-707
115. Macek, B., et al., *Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation.* Mol Cell Proteomics, 2008. **7**(2): p. 299-307.
116. Kobir, A., et al., *Protein phosphorylation in bacterial signal transduction.* Biochim Biophys Acta, 2011. **1810**(10): p. 984-994.
117. Schmidl, S.R., et al., *The phosphoproteome of the minimal bacterium Mycoplasma pneumoniae: analysis of the complete known Ser/Thr kinome suggests the existence of novel kinases.* Mol Cell Proteomics. , 2010. **9**(6): p. 1228-1242.
118. Imamura, H., et al., *Towards the systematic discovery of signal transduction networks using phosphorylation dynamics data.* BMC Bioinformatics., 2010. **11**: p. 232.
119. Spudich, J.L., *Light-regulated retinal-dependent reversible phosphorylation of Halobacterium proteins.* J Biol Chem, 1980. **255**(12): p. 5501-5503.
120. Smith, S.C., P.J. Kennelly, and M. Potts, *Protein-tyrosine phosphorylation in the Archaea.* J Bacteriol 1997. **179**(7): p. 2418-2420.
121. Aivaliotis, M., et al., *Ser/Thr/Tyr protein phosphorylation in the archaeon Halobacterium salinarum-A representative of the third domain of life.* PLoS ONE, 2009. **4**(3): p. e4777.
122. Kennelly, P.J., *Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry.* Biochem. J., 2003. **370**: p. 373-389.
123. Besant, P.G. and P.V. Attwood, *Detection and analysis of protein histidine phosphorylation.* Mol Cell Biochem, 2009. **329**(1-2): p. 93-106.
124. Boyer, P.D., *Phosphohistidine.* Science, 1963. **140**(3586): p. 1147-1153.
125. Grebe, T.W. and J.B. Stock, *The histidine protein kinase superfamily.* Adv Microb Physiol, 1999. **41**: p. 139-227.
126. Oka, A., H. Sakai, and S. Iwakoshi, *His-Asp phosphorelay signal transduction in higher plants: receptors and response regulators for cytokinin signaling in Arabidopsis thaliana.* Genes Genet Syst, 2002. **77**(6): p. 383-391.
127. Mizuno, T., *Two-component phosphorelay signal transduction systems in plants: from hormone responses to circadian rhythms.* Biosci Biotechnol Biochem, 2005. **69**(12): p. 2263-2276.
128. Rudolph, J. and D. Oesterhelt, *Chemotaxis and photoaxis require a CheA histidine kinase in the archaeon Halobacterium salinarium.* EMBO J, 1995. **14**(4): p. 667-673.

129. Rudolph, J., et al., *Phosphorylation in halobacterial signal transduction.* EMBO J, 1995. **14**(17): p. 4249-4257.
130. West, A.H. and A.M. Stock, *Histidine kinases and response regulator proteins in two-component signaling systems.* Trends Biochem Sci., 2001. **26**(6): p. 369-376.
131. Klumpp, S. and J. Krieglstein, *Phosphorylation and dephosphorylation of histidine residues in proteins.* Eur J Biochem, 2002. **269**(4): p. 1067-1071.
132. Klumpp, S. and J. Krieglstein, *Reversible phosphorylation of histidine residues in vertebrate proteins.* Biochim Biophys Acta, 2005. **1754**(1-2): p. 291-295.
133. Kowluru, A., *Emerging roles for protein histidine phosphorylation in cellular signal transduction: lessons from the islet beta-cell.* J Cell Mol Med, 2008. **12**(5B): p. 1885-1908.
134. Attwood, P.V., et al., *Focus on phosphohistidine.* Amino Acids 2007. **32**(145-156).
135. Singh, R.K. and A. Gunjan, *Histone tyrosine phosphorylation comes of age*. Epigenetics., 2011. **6**(2): p. 153-160.
136. McAllister, T.E., M.G. Nix, and M.E. Webb, *Fmoc-chemistry of a stable phosphohistidine analogue.* Chem Commun, 2011. **47**: p. 1297-1299.
137. Matthews, H.R., *Protein kinases and phosphatases that act on histidine, lysine, or arginine residues in eukaryotic proteins: A possible regulator of the mitogen-activated protein kinase cascad.* Pharmacol Ther, 1995. **67**(3): p. 323-350.
138. Skórko, R., *Protein phosphorylation in the archaebacterium Sulfolobus acidocaldarius.* Eur J Biochem, 1984. **145**(3): p. 617-622.
139. Humbard, M.A., S.M.J. Stevens, and J.A. Maupin, *Posttranslational modification of the 20S proteasomal proteins of the archaeon Haloferax volcanii*. J Bacteriol, 2006. **188**(21): p. 7521-7530.
140. Humbard, M.A., et al., *Phosphorylation and methylation of proteasomal proteins of the haloarcheon Haloferax volcanii*. Archaea., 2010. **2010**: p. 481725.
141. Moll, R. and G. Schäfer, *Chemiosmotic H+ cycling across the plasma membrane of the thermoacidophilic archaebacterium Sulfolobus acidocaldarius.* FEBS Lett, 1988. **232**(2): p. 359-363.
142. van de Vossenberg, J.L., et al., *Ion permeability of the cytoplasmic membrane limits the maximum growth temperature of bacteria and archaea.* Mol Microbiol, 1995. **18**(5): p. 925-932.
143. Konings, W.N., et al., *The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments.* Antonie Van Leeuwenhoek, 2002. **81**(1-4): p. 61-72.
144. Fíla, J. and D. Honys, *Enrichment techniques employed in phosphoproteomics.* Amino acids, 2012. **43**(3): p. 1025-1047.
145. Beltran, L. and P.R. Cutillas, *Advances in phosphopeptide enrichment techniques for phosphoproteomics.* Amino acids, 2012. **43**(3): p. 1009-1024.
146. Trufelli, H., et al., *An overview of matrix effects in liquid chromatography-mass spectrometry.* Mass Spectrom Rev, 2011. **30**(3): p. 491-509.
147. Liu, H., R.G. Sadygov, and J.R.r. Yates, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.* Anal. Chem, 2004. **76**: p. 4193-4201.
148. Switzar, L., M. Giera, and W.M. Niessen, *Protein digestion: an overview of the available techniques and recent developments.* J Proteome Res, 2013. **12**(3): p. 1067-1077.
149. Neubauer, G. and M. Mann, *Mapping of phosphorylation sites of gel-isolated proteins by nanoelectrospray tandem mass spectrometry: potentials and limitations.* Analytical chemistry, 1999. **71**(1): p. 235-242.
150. Bian, Y., et al., *Improve the coverage for the analysis of phosphoproteome of HeLa cells by a tandem digestion approach.* Journal of proteome research, 2012. **11**(5): p. 2828-2837.

151. Bisle, B., et al., *Quantitative profiling of the membrane proteome in a halophilic archaeon.* Molecular & Cellular Proteomics, 2006. **5**(9): p. 1543-1558.
152. Pham, T.K., et al., *Quantitative proteomic analysis of Sulfolobus solfataricus membrane proteins.* Journal of proteome research, 2010. **9**(2): p. 1165-1172.
153. Schlosser, A., et al., *Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry.* Analytical chemistry, 2001. **73**(2): p. 170-176.
154. Dickhut, C., et al., *Impact of digestion conditions on phosphoproteomics.* J Proteome Res, 2014.
155. Liu, F., et al., *Integration of cell lysis, protein extraction and digestion into one step for ultrafast sample preparation for phosphoproteome analysis.* Analytical chemistry, 2014.
156. Lee, A.M., et al., *Proteomics of Pyrococcus furiosus, a hyperthermophilic archaeon refractory to traditional methods.* J Proteome Res, 2009. **8**(8): p. 3844-3851.
157. Jensen, S.S. and M.R. Larsen, *Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques.* Rapid Communications in Mass Spectrometry, 2007. **21**(22): p. 3635-3645.
158. Worthington, J., P.R. Cutillas, and J.F. Timms, *IMAC/TiO2 enrich for peptide modifications other than phosphorylation: Implications for chromatographic choice and database searching in phosphoproteomics.* Proteomics, 2011. **11**(23): p. 4583-4587.
159. Ino, Y. and H. Hirano, *Mass spectrometric characterization of proteins transferred from polyacrylamide gels to membrane filters.* FEBS Journal, 2011. **278**(20): p. 3807-3814.
160. Granvogl, B., M. Plöscher, and L.A. Eichacker, *Sample preparation by in-gel digestion for mass spectrometry-based proteomics.* Anal Bioanal Chem, 2007. **389**(4): p. 991-1002.
161. Hellman, U., et al., *Improvement of an "In-Gel" digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing.* Anal Biochem, 1995. **224**(1): p. 451-455.
162. Luque-Garcia, J.L. and T.A. Neubert, *On-membrane tryptic digestion of proteins for mass spectrometry analysis*, in *Protein Blotting and Detection*. 2009, Springer. p. 331-341.
163. Lee, C.H., et al., *On-membrane digestion of beta-casein for determination of phosphorylation sites by matrix-assisted laser desorption/ionization quadrupole/time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 2001. **15**(3): p. 191-202.
164. Yang, C., X. Zhong, and L. Li, *Recent advances in enrichment and separation strategies for mass spectrometry-based phosphoproteomics.* Electrophoresis, 2014.
165. Kinoshita, E., et al., *Phosphate-binding tag, a new tool to visualize phosphorylated proteins.* Molecular & Cellular Proteomics, 2006. **5**(4): p. 749-757.
166. Thingholm, T.E., et al., *SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides.* Molecular & Cellular Proteomics, 2008. **7**(4): p. 661-671.
167. Dunn, J.D., G.E. Reid, and M.L. Bruening, *Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry.* Mass spectrometry reviews, 2010. **29**(1): p. 29-54.
168. Negroni, L., et al., *Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography.* Journal of Chromatography B, 2012. **891**: p. 109-112.
169. Ye, J., et al., *Optimized IMAC− IMAC Protocol for Phosphopeptide Recovery from Complex Biological Samples.* J Proteome Res, 2010. **9**(7): p. 3561-3573.
170. Thingholm, T.E., O.N. Jensen, and M.R. Larsen, *Analytical strategies for phosphoproteomics.* Proteomics, 2009. **9**: p. 1451-1468.
171. Engholm-Keller, K. and M.R. Larsen, *Titanium dioxide as chemo-affinity chromatographic sorbent of biomolecular compounds—Applications in acidic modification-specific proteomics.* Journal of proteomics, 2011. **75**(2): p. 317-328.

172. Vilasi, A., et al., *Enrichment specificity of micro and nano-sized titanium and zirconium dioxides particles in phosphopeptide mapping.* J Mass Spectrom 2013. **48**(11): p. 1188-1198.
173. Larsen, M.R., et al., *Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns.* Mol Cell Proteomics., 2005. **4**(7): p. 873-876.
174. Thingholm, T.E., et al., *Highly selective enrichment of phosphorylated peptides using titanium dioxide.* Nature protocols, 2006. **1**(4): p. 1929-1935.
175. Sugiyama, N., et al., *Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications.* Mol Cell Proteomics, 2007. **6**(6): p. 1103-1109.
176. Zhao, X., et al., *Citric acid-assisted two-step enrichment with TiO2 enhances the separation of multi-and mono-phosphorylated peptides and increases phosphoprotein profiling.* J Proteome Res, 2013.
177. Fukuda, I., et al., *Optimization of Enrichment Conditions on TiO2 Chromatography Using Glycerol As an Additive Reagent for Effective Phosphoproteomic Analysis.* J Proteome Res, 2013. **12**(12): p. 5587-5597.
178. Kanshin, E., S.W. Michnick, and P. Thibault, *Displacement of N/Q-rich peptides on TiO2 beads enhances the depth and coverage of yeast phosphoproteome analyses.* J Proteome Res, 2013.
179. Li, Q.-r., et al., *Effect of peptide-to-TiO_2 beads ratio on phosphopeptide enrichment selectivity.* J Proteome Res, 2009. **8**(11): p. 5375-5381.
180. Montoya, A., et al., *Characterization of a TiO_2 enrichment method for label-free quantitative phosphoproteomics.* Methods, 2011. **54**(4): p. 370-378.
181. Park, S.-S. and S. Maudsley, *Discontinuous pH gradient-mediated separation of TiO2-enriched phosphopeptides.* Anal Biochem, 2011. **409**(1): p. 81-88.
182. Simon, E.S., et al., *Improved enrichment strategies for phosphorylated peptides on titanium dioxide using methyl esterification and pH gradient elution.* Anal. Biochem, 2008. **377**(2): p. 234-242.
183. Sun, Z., K.L. Hamilton, and K.F. Reardon, *Evaluation of quantitative performance of sequential immobilized metal affinity chromatographic enrichment for phosphopeptides.* Analytical biochemistry, 2014. **445**: p. 30-37.
184. Bodenmiller, B., et al., *Reproducible isolation of distinct, overlapping segments of the phosphoproteome.* Nature methods, 2007. **4**(3): p. 231-237.
185. Zarei, M., et al., *Comparison of ERLIC-TiO2, HILIC-TiO2, and SCX-TiO2 for global phosphoproteomics approaches.* J Proteome Res, 2011. **10**(8): p. 3474-3483.
186. Alpert, A.J., *Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides.* Anal Chem, 2008. **80**(1): p. 62-76.
187. Hao, P., T. Guo, and S.K. Sze, *Simultaneous analysis of proteome, phospho-and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography.* PloS one, 2011. **10**(12): p. 5568-5574.
188. Zarei, M., et al., *Combinatorial use of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and strong cation exchange (SCX) chromatography for in-depth phosphoproteome analysis.* Journal of proteome research, 2012. **11**(8): p. 4269-4276.
189. Engholm-Keller, K., et al., *TiSH-a robust and sensitive global phosphoproteomics strategy employing a combination of TiO_2, SIMAC, and HILIC.* J Proteomics, 2012. **75**(18): p. 5749-5761.
190. Zhou, H., et al., *Toward a comprehensive characterization of a human cancer cell phosphoproteome.* J Proteome Res, 2012. **12**(1): p. 260-271.

191. D'Ambrosio, C., et al., *Analytical methodologies for the detection and structural characterization of phosphorylated proteins.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**(1-2): p. 163-180.
192. Collins, M.O., L. Yu, and J.S. Choudhary, *Analysis of protein phosphorylation on a proteome-scale.* Proteomics, 2007. **7**(16): p. 2751-2768.
193. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-9533.
194. Jones, A.W. and H.J. Cooper, *Dissociation techniques in mass spectrometry-based proteomics.* Analyst, 2011. **136**(17): p. 3419-3429.
195. Klink, F., *Peptide sequencing with Electrospray LC/MS Part 1: Ion types and nomenclature.* Seperation Science, 2011.
196. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom., 1984. **11**(11): p. 601.
197. Johnson, R.S., et al., *Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine.* Anal Chem., 1987 **59**(21): p. 2621-2625.
198. Wiesner, J., T. Premsler, and A. Sickmann, *Application of electron transfer dissociation (ETD) for the analysis of posttranslational modifications.* Proteomics., 2008. **8**(21): p. 4466-4483.
199. Mikesh, L.M., et al., *The utility of ETD mass spectrometry in proteomic analysis.* Biochim Biophys Acta, 2006. **1764**(12): p. 1811-1822.
200. Molina, H., et al., *Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry.* Proc Natl Acad Sci U S A, 2007. **104**(7): p. 2199-2204.
201. Swaney, D.L., et al., *Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors.* Anal Chem, 2007 **79**(2): p. 477-485.
202. Liu, C.W. and C.C. Lai, *Effects of electron-transfer coupled with collision-induced dissociation (ET/CID) on doubly charged peptides and phosphopeptides.* J Am Soc Mass Spectrom, 2011. **22**(1): p. 57-66.
203. Kleinnijenhuis, A.J., et al., *Analysis of histidine phosphorylation using tandem MS and ion-electron reactions.* Anal Chem. , 2007. **79**(19): p. 7450-7456.
204. Boersema, P.J., S. Mohammed, and A.J. Heck, *Phosphopeptide fragmentation and analysis by mass spectrometry.* J Mass Spectrom., 2009. **44**(6): p. 861-878.
205. Kim, M.S., et al., *Systematic evaluation of alternating CID and ETD fragmentation for phosphorylated peptides.* Proteomics, 2011. **11**(12): p. 2568-2572.
206. Eng, J.K., A.L. McCormack, and J.R. Yates, III, , *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-989.
207. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-3567.
208. Colinge, J., et al., *OLAV: towards high-throughput tandem mass spectrometry data identification.* Proteomics, 2003. **3**(8): p. 1454-1463.
209. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra.* Bioinformatics, 2004. **20**(9): p. 1466-1467.
210. Diament, B.J. and W.S. Noble, *Faster SEQUEST searching for peptide identification from Tandem Mass Spectra.* J. Proteome Res., 2011. **10**(9): p. 3871-3879.
211. Dagda, R.K., T. Sultana, and J. Lyons-Weiler, *Evaluation of the consensus of four peptide identification algorithms for tandem Mass Spectrometry based proteomics.* J Proteomics Bioinform, 2010. **3**: p. 39-47.

212. Kapp, E.A., et al., *An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis.* Proteomics, 2005. **5**(13): p. 3475-3490.
213. Searle, B.C., M. Turner, and A. Nesvizhskii, *Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies.* J Proteome Res. , 2008. **7**(1): p. 245-253.
214. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nature methods, 2007. **4**(3): p. 207-214.
215. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nature biotechnology, 2006. **24**(10): p. 1285-1292.
216. Savitski, M.M., et al., *Confident phosphorylation site localization using the Mascot Delta Score.* Molecular & Cellular Proteomics, 2011. **10**(2).
217. Taus, T., et al., *Universal and confident phosphorylation site localization using phosphoRS.* J Proteome Res, 2011. **10**(12): p. 5354-5362.
218. Aldridge, B.B. and K.Y. Rhee, *Microbial metabolomics: innovation, application, insight.* Curr Opin Microbiol, 2014. **19**: p. 90-96.
219. Zhang, A., et al., *Modern analytical techniques in metabolomics analysis.* Analyst, 2012. **137**(2): p. 293-300.
220. Bräsen, C., et al., *Carbohydrate metabolism in archaea: current insights into unusual enzymes and pathways and their regulation.* Microbiology and Molecular Biology Reviews, 2014. **78**(1): p. 89-175.
221. Ahmed, H., et al., *The semi-phosphorylative Entner–Doudoroff pathway in hyperthermophilic archaea: a re-evaluation.* Biochemical Journal, 2005. **390**(Pt 2): p. 529.
222. Zaparty, M., et al., *"Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.* Extremophiles., 2010. **14**(1): p. 119-142.
223. Strober, W., *Monitoring cell growth.* Curr Protoc Immunol., 2001. **Appendix 3**: p. Appendix 3A.
224. Ow, S.Y., et al., *iTRAQ underestimation in simple and complex mixtures:"the good, the bad and the ugly".* J Proteome Res, 2009. **8**(11): p. 5347-5355.
225. Macek, B., et al., *The serine/threonine/tyrosine phosphoproteome of the model bacterium Bacillus subtilis.* Molecular & Cellular Proteomics, 2007. **6**(4): p. 697-707.
226. Srivastava, O.P. and K. Srivastava, *Purification of gamma-crystallin from human lenses by acetone precipitation method.* Curr Eye Res, 1998 **17**(11): p. 1074-1081.
227. Simpson, D.M. and R.J. Beynon, *Acetone precipitation of proteins and the modification of peptides.* J. Proteome. Res, 2010. **9**(1): p. 4444-450.
228. Mohammed, S. and A.J. Heck, *Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications.* Curr Opin Biotechnol, 2011. **22**(1): p. 9-16.
229. Hao, P., et al., *Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) versus SCX for Fractionation of iTRAQ-Labeled Peptides.* J Proteome Res, 2011. **Oct 20, 2011, Epub**
230. Edelmann, M.J., *Strong cation exchange chromatography in analysis of posttranslational modifications: innovations and perspectives.* BioMed Research International, 2011. **2011**.
231. Kettenbach, A.N. and S.A. Gerber, *Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments.* Anal Chem, 2011. **Epub ahead of print**.

232. Park, S.S. and S. Maudsley, *Discontinuous pH gradient-mediated separation of TiO2-enriched phosphopeptides.* Anal Biochem, 2011. **409**(1): p. 81-88.
233. McLachlin, D.T. and B.T. Chait, *Analysis of phosphorylated proteins and peptides by mass spectrometry*. Current opinion in chemical biology, 2001. **5**(5): p. 591-602.
234. Sezonov, G., D. Joseleau-Petit, and R. D'Ari, *Escherichia coli physiology in Luria-Bertani broth.* J. Bacteriol., 2007. **189**(23): p. 8746-8749.
235. Thongboonkerd, V., et al., *Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation.* Kidney Int, 2002. **62**(4): p. 1461-1469.
236. Febbraio, F., et al., *Thermal stability and aggregation of sulfolobus solfataricus beta-glycosidase are dependent upon the N-epsilon-methylation of specific lysyl residues: critical role of in vivo post-translational modifications.* J Biol Chem, 2004. **279**(11): p. 10185-10194.
237. Botting, C.H., et al., *Extensive lysine methylation in hyperthermophilic crenarchaea: potential implications for protein stability and recombinant enzymes.* Archaea, 2010. **pii: 106341**.
238. Taouatas, N., et al., *Strong cation exchange-based fractionation of Lys-N-generated peptides facilitates the targeted analysis of post-translational modifications.* Mol Cell Proteomics, 2009. **8**(1): p. 190-200.
239. Hao, P., T. Guo, and S.K. Sze, *Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography.* PLoS One, 2011. **6**(2): p. e16884.
240. Rolfsmeier, M. and P. Blum, *Purification and characterization of a maltase from the extremely thermophilic crenarchaeote Sulfolobus solfataricus.* J Bacteriol 1995. **177**(2): p. 482-485.
241. Berkner, S. and G. Lipps, *Genetic tools for Sulfolobus spp.: vectors and first applications.* Archives of microbiology, 2008. **190**(3): p. 217-230.
242. Schelert, J., et al., *Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption.* J Bacteriol 2004. **186**(2): p. 427-437.
243. Gusev, Y., *Computational methods for analysis of cellular functions and pathways collectively targeted by differentially expressed microRNA.* Methods, 2008. **44**(1): p. 61-72.
244. Elferink, M.G.L., et al., *Sugar transport in Sulfolobus solfataricus is mediated by two families of binding protein-dependent ABC transporters.* Molecular microbiology, 2001. **39**(6): p. 1494-1503.
245. Albers, S.-V., et al., *Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein.* Journal of bacteriology, 1999. **181**(14): p. 4285-4291.
246. Gogliettino M, et al., *A highly selective oligopeptide binding protein from the archaeon Sulfolobus solfataricus.* J Bacteriol, 2010. **192**(12): p. 3123-3131.
247. Konings, W.N., et al., *The cell membrane plays a crucial role in survival of bacteria and archaea in extreme environments.* Antonie Van Leeuwenhoek, 2002. **81**(1-4): p. 61-72.
248. White, D., *The Physiology and Biochemistry of Prokaryotes.* . 1995, New York: Oxford University Press.
249. Owen, O.E., S.C. Kalhan, and R.W. Hanson, *The key role of anaplerosis and cataplerosis for citric acid cycle function.* Journal of Biological Chemistry, 2002. **277**(34): p. 30409-30412.
250. Pham, T.K., et al., *A quantitative proteomic analysis of biofilm adaptation by the periodontal pathogen Tannerella forsythia.* Proteomics, 2010. **10**(17): p. 3130-3141.
251. Bessarabova, M., et al., *Knowledge-based analysis of proteomics data.* BMC bioinformatics, 2012. **13**(Suppl 16): p. S13.
252. Albers, S.V., et al., *Insights into ABC transport in archaea.* Journal of bioenergetics and biomembranes, 2004. **36**(1): p. 5-15.

253. Fukushima, A., et al., *Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach.* BMC systems biology, 2011. **5**(1): p. 1.
254. Bar-Even, A., E. Noor, and R. Milo, *A survey of carbon fixation pathways through a quantitative lens.* Journal of experimental botany, 2012. **63**(6): p. 2325-2342.
255. Haferkamp, P., et al., *An additional glucose dehydrogenase from Sulfolobus solfataricus: fine-tuning of sugar degradation?* Biochemical Society Transactions, 2011. **39**(1): p. 77.
256. Alberts, B., et al., *Chloroplasts and Photosynthesis.* 2002.
257. Berg, I.A., et al., *Autotrophic carbon fixation in archaea.* Nature Reviews Microbiology, 2010. **8**(6): p. 447-460.
258. Berg, I.A., et al., *A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea.* Science, 2007. **318**(5857): p. 1782-1786.
259. Brosnan, J.T., *Interorgan amino acid transport and its regulation.* The Journal of nutrition, 2003. **133**(6): p. 2068S-2072S.
260. Umbarger, H.E., *Amino acid biosynthesis and its regulation.* Annual review of biochemistry, 1978. **47**(1): p. 533-606.
261. Adelberg, E.A. and H.E. Umbarger, *Isoleucine and valine metabolism in Escherichia coli.* Journal of Biological Chemistry, 1953. **205**(1): p. 475-482.
262. Umbarger, H.E. and B. Brown, *Isoleucine and valine metabolism in Escherichia coli.* Journal of Biological Chemistry, 1958. **233**(2): p. 415.
263. Knöchel, T., et al., *The crystal structure of anthranilate synthase from Sulfolobus solfataricus: functional implications.* Proceedings of the National Academy of Sciences, 1999. **96**(17): p. 9479-9484.
264. Brosnan, J.T. and M.E. Brosnan, *Branched-chain amino acids: enzyme and substrate regulation.* The Journal of nutrition, 2006. **136**(1): p. 207S-211S.
265. Park, C.B. and S.B. Lee, *Effects of exogenous compatible solutes on growth of the hyperthermophilic archaeon Sulfolobus solfataricus.* Journal of bioscience and bioengineering, 2000. **89**(4): p. 318-322.
266. Feehily, C. and K.-A.G. Karatzas, *Role of glutamate metabolism in bacterial responses towards acid and other stresses.* Journal of applied microbiology, 2013. **114**(1): p. 11-24.
267. Diez-Gonzalez, F. and Y. Karaibrahimoglu, *Comparison of the glutamate-, arginine-and lysine-dependent acid resistance systems in Escherichia coli O157: H7.* Journal of applied microbiology, 2004. **96**(6): p. 1237-1244.
268. Bearson, B.L., I.S. Lee, and T.A. Casey, *Escherichia coli O157: H7 glutamate-and arginine-dependent acid-resistance systems protect against oxidative stress during extreme acid challenge.* Microbiology, 2009. **155**(3): p. 805-812.
269. Álvarez-Ordóñez, A., et al., *Arginine and lysine decarboxylases and the Acid Tolerance Response of Salmonella Typhimurium.* International journal of food microbiology, 2010. **136**(3): p. 278-282.
270. Consalv, V., et al., *Glutamate dehydrogenase from the thermoacidophilic archaebacterium Sulfolobus solfataricus.* European journal of biochemistry, 1991. **196**(2): p. 459-467.
271. Kim, H.-W., et al., *Purification and characterization of the first archaeal glutamate decarboxylase from Pyrococcus horikoshii.* Bioscience, biotechnology, and biochemistry, 2009. **73**(1): p. 224-227.
272. Hong, S.-J., et al., *Overexpression and characterization of recombinant glutamate decarboxylase from Thermococcus kodakaraensis KOD1.* Journal of the Korean Society for Applied Biological Chemistry, 2012. **55**(2): p. 213-218.

273. Lee, E.-S., et al., *Gene expression and characterization of thermostable glutamate decarboxylase from Pyrococcus furiosus.* Biotechnology and Bioprocess Engineering, 2013. **18**(2): p. 375-381.
274. Sanangelantoni, A.M., et al., *Cloning and nucleotide sequence of an archaebacterial glutamine synthetase gene: phylogenetic implications.* Molecular and General Genetics MGG, 1990. **221**(2): p. 187-194.
275. Yin, Z., et al., *The glutamine synthetase from the hyperthermoacidophilic crenarcheon Sulfolobus acidocaldarius: isolation, characterization and sequencing of the gene.* Biological Chemistry, 1998. **379**(11): p. 1349-1354.
276. Rich, P.R., *The molecular machinery of Keilin's respiratory chain.* Biochemical Society Transactions, 2003. **31**: p. 1095-1105.
277. Sugiyama, N., et al., *Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications.* Molecular & Cellular Proteomics, 2007. **6**(6): p. 1103-1109.
278. Aryal, U.K. and A.R.S. Ross, *Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry.* Rapid Communications in Mass Spectrometry, 2010. **24**(2): p. 219-231.
279. Dunn, J.D., G.E. Reid, and M.L. Bruening, *Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry.* Mass Spectrom Rev, 2010. **29**(1): p. 29-54.
280. Jia, C., et al., *High-definition de novo sequencing of crustacean hyperglycemic hormone (CHH)-family neuropeptides.* Molecular & Cellular Proteomics, 2012. **11**(12): p. 1951-1964.
281. Longworth, J., *Proteomics in microalgae: a postgenomic approach for improved biofuel production.* PhD thesis, 2013.
282. Ow, S.Y., et al., *Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation.* Proteomics, 2011. **11**(11): p. 2341-2346.
283. Wu, J., et al., *Global profiling of phosphopeptides by titania affinity enrichment.* J Proteome Res, 2007. **6**(12): p. 4684-4689.
284. Kettenbach, A.N. and S.A. Gerber, *Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments.* Analytical chemistry, 2011. **83**(20): p. 7635-7644.
285. Hunter, T., *Signaling—2000 and beyond.* Cell, 2000. **100**(1): p. 113-127.
286. Aivaliotis, M., et al., *Ser/Thr/Tyr protein phosphorylation in the archaeon Halobacterium salinarum—a representative of the third domain of life.* PLoS One, 2009. **4**(3): p. e4777.
287. Verstraeten, N., et al., *The universally conserved prokaryotic GTPases.* Microbiology and Molecular Biology Reviews, 2011. **75**(3): p. 507-542.
288. Leipe, D.D., et al., *Classification and evolution of P-loop GTPases and related ATPases.* Journal of molecular biology, 2002. **317**(1): p. 41-72.
289. Wu, H., et al., *Structure of the ribosome associating GTPase HflX.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(3): p. 705-713.
290. Fischer, J.J., et al., *The ribosome modulates the structural dynamics of the conserved GTPase HflX and triggers tight nucleotide binding.* Biochimie, 2012. **94**(8): p. 1647-1659.
291. Huang, B., et al., *Functional study on GTP hydrolysis by the GTP-binding protein from Sulfolobus solfataricus, a member of the HflX family.* Journal of biochemistry, 2010. **148**(1): p. 103-113.
292. Kaur, G., et al., *A novel MntR-independent mechanism of manganese homeostasis in Escherichia coli by the ribosome-associated protein HflX.* Journal of bacteriology, 2014: p. JB. 01717-14.

293. Greggio, E., et al., *The Parkinson's disease kinase LRRK2 autophosphorylates its GTPase domain at multiple sites.* Biochemical and biophysical research communications, 2009. **389**(3): p. 449-454.
294. Mackay, D.T., et al., *An acetylase with relaxed specificity catalyses protein N-terminal acetylation in Sulfolobus solfataricus.* Molecular microbiology, 2007. **64**(6): p. 1540-1548.
295. Pham, K.T. and P.C. Wright, *The Proteomic Response of Saccharomyces cerevisiae in Very High Glucose Conditions with Amino Acid Supplementation.* J. Proteome Res., 2008. **7**(11): p. 4766-4774.
296. Tautenhahn, R., et al., *XCMS Online: a web-based platform to process untargeted metabolomic data.* Analytical chemistry, 2012. **84**(11): p. 5035-5039.
297. Patti, G.J., et al., *A view from above: cloud plots to visualize global metabolomic data.* Analytical chemistry, 2012. **85**(2): p. 798-804.
298. Meyer, M.R., F.T. Peters, and H.H. Maurer, *Automated mass spectral deconvolution and identification system for GC-MS screening for drugs, poisons, and metabolites in urine.* Clinical chemistry, 2010. **56**(4): p. 575-584.
299. Bouatra, S., et al., *The human urine metabolome.* PloS one, 2013. **8**(9): p. e73076.
300. Ogata, H., et al., *KEGG: Kyoto encyclopedia of genes and genomes.* Nucleic acids research, 1999. **27**(1): p. 29-34.
301. Albers, S.-V., et al., *SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon Sulfolobus solfataricus under temperature variation.* 2009.
302. Pham, T.K., et al., *Quantitative proteomic analysis of Sulfolobus solfataricus membrane proteins.* J Proteome Res, 2010. **9**(2): p. 1165-1172.
303. Humbard, M.A., et al., *Phosphorylation and methylation of proteasomal proteins of the haloarcheon Haloferax volcanii.* Archaea, 2010. **2010**.
304. Kirkland, P.A., et al., *Shotgun proteomics of the haloarchaeon Haloferax volcanii.* Journal of proteome research, 2008. **7**(11): p. 5033-5039.
305. Humbard, M.A., S.M. Stevens, and J.A. Maupin-Furlow, *Posttranslational modification of the 20S proteasomal proteins of the archaeon Haloferax volcanii.* Journal of bacteriology, 2006. **188**(21): p. 7521-7530.
306. Potters, M.B., et al., *Phosphoprotein with phosphoglycerate mutase activity from the archaeon Sulfolobus solfataricus.* Journal of bacteriology, 2003. **185**(7): p. 2112-2121.
307. Lower, B.H., M.B. Potters, and P.J. Kennelly, *A phosphoprotein from the archaeon Sulfolobus solfataricus with protein-serine/threonine kinase activity.* Journal of bacteriology, 2004. **186**(2): p. 463-472.
308. Redbird, R.A., *Identification of a protein kinase substrate in Sulfolobus solfataricus P2. Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in*
partial fulfillment of the requirements for the degree of Doctor of philosophy in biochemistry. Blacksburg, United States. . PhD thesis, 2010: p. 86-107.
309. Haile, J.D. and P.J. Kennelly, *The activity of an ancient atypical protein kinase is stimulated by ADP-ribose< i> in vitro</i>.* Archives of biochemistry and biophysics, 2011. **511**(1): p. 56-63.
310. Reimann, J., et al., *Regulation of archaella expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in Sulfolobus acidocaldarius.* Molecular microbiology, 2012. **86**(1): p. 24-36.

Appendix A

Table

Table TS2.1 Phosphoproteins identified in archaea

P-protein	**P-sites**	**Archaea**	**Method**	**Reference**
Light-regulated (LR) protein LR1, LR2	Ser/Thr	*Halobacterium halobium*	^{32}P labeling	Spudich, *et al*., 1980
Chemotaxis protein CheA	His	*H. salinarum*	^{32}P labeling	Rudolph, and Oesterhelt, 1995
Chemotaxis protein CheY	Asp	*H. salinarum*	^{32}P labeling	Rudolph, *et al*., 1995
Methyltransferase-activating protein	Ser, Thr or Tyr	*Methanosarcina barkeri*	^{32}P labeling	Daas, *et al*., 1996
Phosphtyrosine protein	Tyr	*Haloferax volcanii*	antibody with specificity for P-tyr immunoreactions	Smith, *et al.*, 1997
	Tyr	*Methanosarcina thermophila TM-1*,	antibody	Smith, *et al.*, 1997
	Tyr	*S. sulfataricus* P1 ATCC 35091	antibody	Smith, *et al.*, 1997
Zinc-dependent Leucyl aminopeptidases	Ser or Thr	*S. sulfataricus* MT4	^{32}P labeling	Codon, *et al.*, 1998
SSO0207, phosphohexomutase	Ser309	*S. sulfataricus*	^{32}P labeling & LC-MS/MS	Solow, *et al*., 1998, Ray et al., 2005
Glycogen synthase	Ser or Thr	*S. acidocaldarius*	^{32}P labeling	Cardona, *et al*., 2001
Phenylalanyl-tRNA synthetase β-chain (FTS)	Tyr	*Thermococcus kodakaraensis KOD1*	anti P-tyr antibody	Jeon, *et al*., 2002
Phosphomannomutase (PMM)	Tyr	*Thermococcus kodakaraensis KOD2*	anti P-tyr antibody	Jeon, *et al*., 2002

RNA terminal phosphate cyclase operon B (RtcB)	Tyr	*Thermococcus kodakaraensis KOD3*	anti P-tyr antibody	Jeon, *et al*., 2002
archaeal initiation factor 2(aIF2)	Ser48	*Pyrococcus horikoshii* OT3	^{32}P labeling (*in vitro*)	Tahara, *et al*., 2004
D-gluconate dehydratase , Sso3198	Ser	*S. sulfataricus*	enzyme purify, characterize	Kim and Lee, 2005
Rio2, Serine protein kinase	Ser128	*Archaeoglobus fulgidus*	^{32}P labeling	LaRonde-LeBlanc and Wlodawer, 2004, LaRonde-LeBlanc *et al*., 2005
Rio1p, homolog of Ser/Thr kinase	Ser/Thr	*Haloferax volcanii*	^{32}P labeling and MS	[303]Humbard *et al*., 2010
ST1565, Ser/Thr protein kinase	Thr326,negative regulation site; Thr329 , one of the major activation site	*S. tokodaii*	^{32}P labeling	Wang *et al*., 2010
ST0829, Fork head domain-containing protein	Ser/Thr	*S. tokodaii*	^{32}P labeling	Duan and He, 2011
CBS domain protein	Asp118	*S. tokodaii*	^{32}P labeling	Ragunathan *et al*., 2008
SSO0771, Cdc6-1 (cell division control protein 6)	Thr/Ser	*S. solfataricus*	^{32}P labeling	De Felice *et al*., 2003
SSO2184, Cdc6-2	Ser	*S. solfataricus*	^{32}P labeling	De Felice et al., 2004
SSO0257, Cdc6-3	Ser	*S. solfataricus*	^{32}P labeling	De Felice et al., 2006
Cell division control protein 6(Cdc6)	Ser	*Methanobacterium thermoautotrophicum*	^{32}P labeling	Grabowski and Kelman, 2001, Shin *et al*., 2003, Kasiviswanathan *et al*., 2005
Cdc6	Ser	*Thermoplasma acidophilum*	^{32}P labeling	Haugland *et al*., 2008a
Cdc6	Ser	*Pyrobaculum aerophilum*	^{32}P labeling	Grabowski and Kelman, 2001, De Felice *et al*., 2004
Cdc6	Ser	*Aeropyrum pernix*	^{32}P labeling	Atanassova and Grainge, 2008
Minichromosome maintenance	Ser/Thr/Tyr	*Aeropyrum pernix*	^{32}P labeling	Atanassova and Grainge, 2008

RNA kinase $tRNA^{Ile}$-agm^2C synthetase（TiaS）	Thr^{18}	*Archaeoglobus fulgidus*	^{32}P labeling	Osawa *et al*., 2011; Terasaka *et al*., 2011
Hypothetical protein (HVO_C0059)	Ser136 and Thr139	*Haloferax volcanii*	MS	[304]Kirkland, *et al*., 2008
RNA kinase $tRNA^{Ile}$-agm^2C synthetase（TiaS）	Thr^{18}	*Archaeoglobus fulgidus*	^{32}P labeling	Osawa *et al*., 2011; Terasaka *et al*., 2011
α1,α2,PAN-A subunit of proteasome	α1 Thr147, α2 Thr 13, and PNA-A Ser 340	*Haloferax volcanii*	MS/MS and ^{32}P labeling	[303]Humbard *et al*., 2010
β-subunit of proteasome	Ser129	*Haloferax volcanii*	MS	[305]Humbard *et al*., 2006
SSO2387, SSOPK2,putative protein serine kinase	Ser	*S. sulfataricus*	^{32}P labeling & MS	[306]Lower and Kennelly, 2003
SSO0469, SSOPK3,putative protein serine kinase	Ser/Thr	*S. sulfataricus*	^{32}P- labeling & MS	[307]Lower, *et al.*, 2004
SSO0563, SSOPK4,putative protein serine kinase	Ser	*S. sulfataricus*	^{32}P labeling & MS	[308]Redbird, 2010
SSO0433, SSOPK5/ homolog of the piD261/Bud32 family of protein kinase in Eucarya	Auto-P- Se/Thr	*S. sulfataricus*	^{32}P labeling Site-directed mutagenesis, SDS-PAGE, Western blotting	[309]Haile and Kennelly, 2011
The FHA domain-containing protein ArnA and the vWA domain-containing protein ArnB	Ser/Thr	*S.acidocaldarius*	^{32}P labeling	[310]Reimann, et al., 2012
SSO0417, cofactor-independent phosphoglycerate mutase	Ser59	*S. sulfataricus*		Potters, et al., 2003
69 proteins	Ser/Thy/Tyr	*H. salinarum*	SCX & TiO2, LC- MS/MS	[286]

622 proteins	Ser/Thr/Tyr	*S. sulfataricus*	LC- MS/MS	[7]
801 proteins	Ser/Thr/Tyr	*S.acidocaldarius*	LC- MS/MS	[47]

Appendix B

Fig FS 3.1 SCX chromatography of peptides separation (4 mg for each) of *S.solfataricus* P2 grown on different carbon sources.

Table TS3.1 Specific growth rates and doubling times of *S.solfataricus* under different conditions

Table TS3.2 Relationship between cell numbers of *S. solfataricus* and OD_{650}

Table TS3.3 Cell dry weight (g/L) of S. solfataricus strains.

Table TS3.4 Identified proteins and peptides from SDS-PAGE digestion/SCX fractions.

Table TS3.5 Identified protein and peptides using phosphorylation search

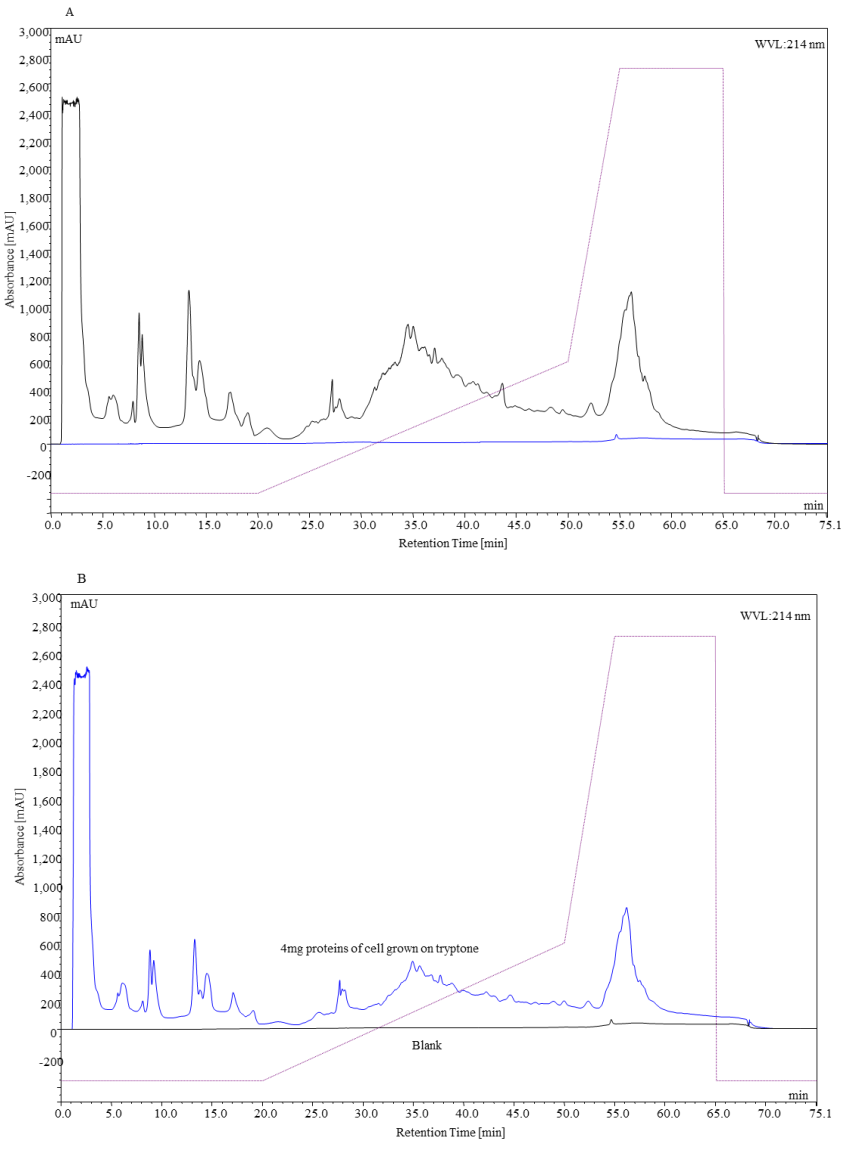

Fig FS 3.1 SCX chromatography of peptides separation of *S.solfataricus* P2 grown on glucose (A) and tryptone (B).

Table TS3.1 Specific growth rates and doubling times of *S.solfataricus* under different conditions

S. solfataricus PBL2025					*S. solfataricus*P BL2073				
Carbon	AVE µ	STD µ	AVE t	STD t	Carbon	AVE µ	STD µ	AVE t	STD t
G	0.0214	0.00045	32.5	0.69	G	0.0247	0.00114	28.1	1.27
T	0.0083	0.00064	83.5	6.66	T	0.0106	0.00064	65.4	4.08

µ: specific growth rate (h^{-1}), t: doubling time (h), t: Ln(2/t)

AVE: average from biological triplicates; STD: Standard Deviation from biological triplicates.

Table TS3.2 Relationship between cell numbers of *S. solfataricus* and OD_{650}

Carbon source	P2		PBL2025		PBL2073	
	OD	Cell number /ml ($x10^7$)	OD	Cell number /ml ($x10^7$)	OD	Cell number /ml ($x10^7$)
Glucose	0.202	0.234	0.198	0.269	0.221	0.280
	0.492	1.468	0.534	1.380	0.550	1.558
	0.795	2.275	0.830	2.760	0.769	2.764
	1.100	3.798	1.279	4.383	1.141	3.398
	1.386	4.721	1.398	5.287	1.328	4.252
Tryptone	0.199	0.113	0.203	0.227	0.189	0.235
	0.484	1.446	0.475	1.411	0.487	1.346
	0.549	1.805	0.657	2.137	0.780	2.160
	0.618	2.084	0.787	2.486	0.832	2.282

Number of cells ($X10^7$) = A * OD_{650} + B

y=3.8327x-0.536, y=4.1025x-0.6624, y=3.3913x-0.4059 for P2, PBL2025 and PBL2073 grown on glucose respectively; and y=4.7501x-0.836, y=3.9315x-0.5205 y=3.166x-0.3051

Table TS3.3 Cell dry weight (g/L) of S. solfataricus strains

Strains	0.4%Glucose	0.2%Tryptone
P2	0.81+ 0.01	0.41+ 0.04
PBL2025	0.64+ 0.06	0.43+ 0.05
PBL2073	0.75+ 0.03	0.48+ 0.07

Table TS3.4 Identified proteins and peptides from SDS-PAGE digestion/SCX fractions

Slice/Fractions	Unique protein	Unique Phosphoprotein	Unique peptide	Unique Phosphopeptide
S1	9	0	42	0
S2	21	1	62	1
S3	28	0	126	0
S4	26	1	140	1
S5	21	0	124	0
S6	26	2	81	2
S7	30	1	97	1
S8	32	2	76	2
S9	36	4	84	4
S10	15	0	32	0
SUM	177	4	412	4

Table TS3.5 Identified protein and peptides using phosphorylation search

Fractions	Unique protein	Unique Phosphoprotein	Unique peptide	Unique Phosphopeptide
F14	0	0	0	0
F16	0	0	0	0
F18	1	0	1	0
F20	18	0	24	0
F22	69	3	83	3
F24	33	1	40	1
F26	12	1	12	1
F28	15	1	16	1
F30	18	1	19	1
F32	10	0	12	0
SUM	137	7	221	7

Appendix C

Figures Fig FS 4.1 SDS-PAGE gel for iTRAQ sets.

Fig FS 4.2 Cell pellets from different strains.

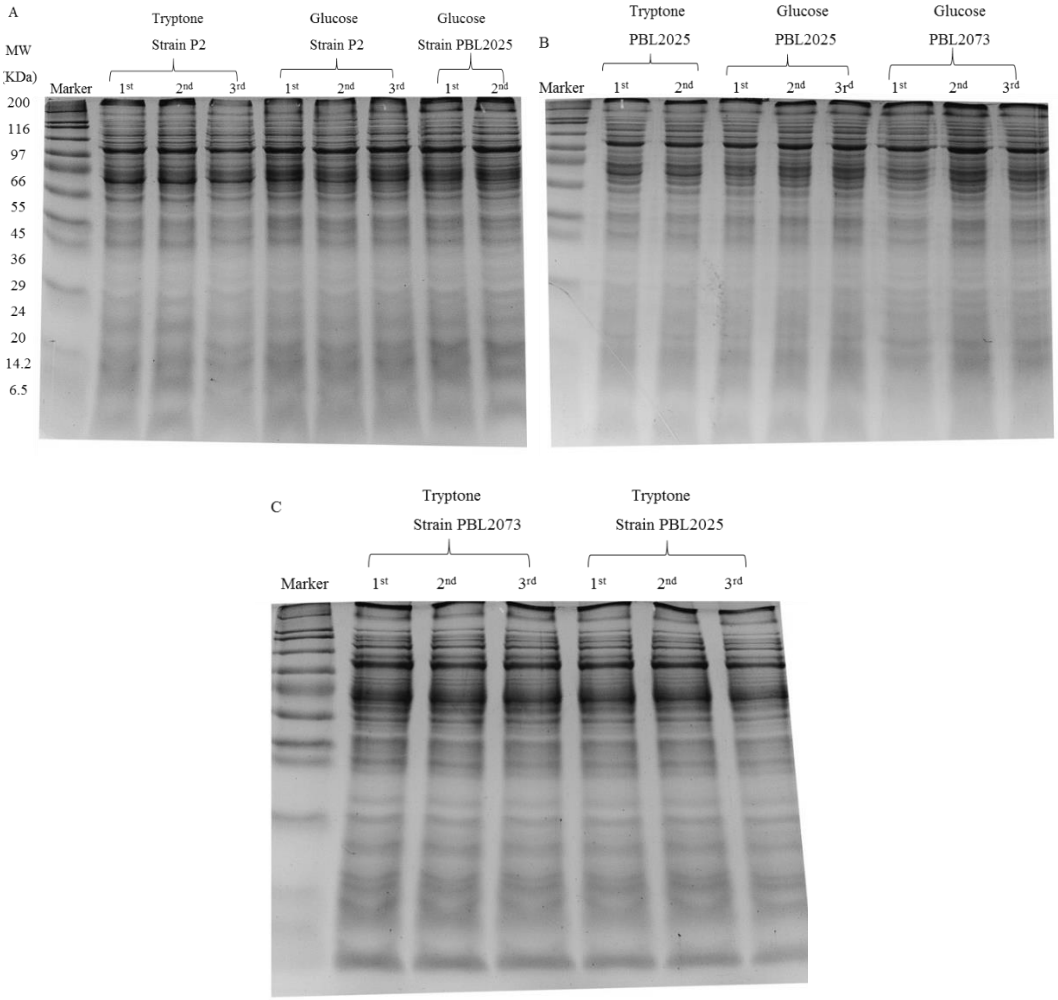

Fig FS 4.1 SDS-PAGE gel of proteins for iTRAQ sets: the 1st set (A); the 2nd set (B) and the 3rd set (C).

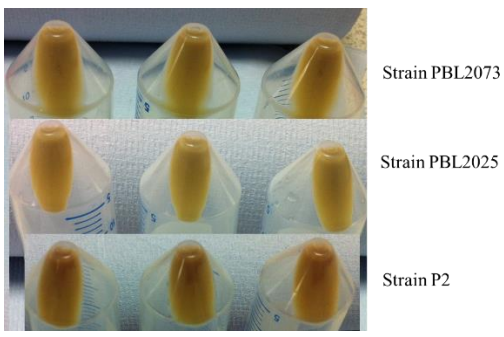

Fig FS 4.2 Cell pellets collected in late-exponenial growth phase of different *S. solfataricus* strains.

Appendix D

Table SP 4.3 The number of proteins from predicted database and iTRAQ data.

	Predicted	T/G-P2	T/G-2025	2025/P2-Glu
arCOG C	185	15	22	21
arCOG D	15	2	3	2
arCOG E	215	19	26	21
arCOG F	64	4	3	4
arCOG G	154	12	12	7
arCOG H	94	5	7	7
arCOG I	92	7	9	11
arCOG J	169	22	24	15
arCOG K	130	15	16	11
arCOG L	460	3	4	1
arCOG M	78	2	4	3
arCOG N	19	0	0	0
arCOG O	74	11	9	7
arCOG P	65	2	6	1
arCOG Q	57	1	2	2
arCOG R	348	17	31	16
arCOG S	460	14	24	14
arCOG T	15	1	1	1
arCOG U	12	1	1	1
arCOG V	125	6	4	13
arCOG X	91	0	0	0
Sum	2922	159	208	158

T: tryptone. G: glucose.

Appendix E

Table TS5.1 Buffers used for TiO_2 /IMAC enrichment

Table TS5.2-5.4, 5.6-5.9: Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr, and peptides with mono-/multiple- P-sites detected from different enrichment strategies:

Table TS5.5 Percentage of acidic amino acids in detected peptides from different acids of beads

Table TS5.10 Percentage of triplicate charge peptides in P-peptides and total detected peptides.

Tables Each supplemented table supports Fig 5.2-5.4 and Fig 5.6-5.9.

Table TS5.1 Buffers used for TiO_2 /IMAC enrichment

	Loading/binding buffer	Wash buffer (two times)	Elute buffer	Ref
1	1M glycolic acid in 80% ACN, 5%TFA	80% ACN, 1%TFA	5% NH_4OH, pH 12	GE Healthcare TiO_2 Mag Sepharose
2	2 M lactic acid in 80%ACN, 1%TFA	80% ACN, 5%TFA	3% NH_4OH, pH 10.5	[284] with modification
3	Citric acid (50 mM, 100 mM, 200 mM, 500 mM, 1M) in 80%ACN, 1%TFA	60% ACN, 0.1%TFA and 0.1%TFA	4% NH_4OH in 60% ACN	[176]
4	Glutamic acid (about 0.14 M) in 65%ACN, 2%TFA	65% ACN, 0.5%TFA and 65% ACN, 0.1%TFA	300 mM NH_4OH in 50% ACN, 500 mM NH_4OH in 60% ACN	[283]
5	Lactic acid (0.25 M) in 70%ACN, 3%TFA	70% ACN, 3%TFA and 70% ACN, 3%TFA, 125 mM of asparagine,125 mM of glutamine	500 mM KH_2PO_4, pH 7	[178]

Note: ammonium hydroxide (NH_4OH).

GE-Healthcare TiO_2 Mag Sepharose

(http://www.gelifesciences.com/webapp/wcs/stores/servlet/productById/zh/GELifeSciences/28944010)

Table TS5.2 Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr, and peptides with mono-/multiple- P-sites detected from IMAC elution buffer and supernatant flow.

	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
IMAC -IME	26	24	7	10	17	19	7
IMAC -ACN	26	25	10	7	19	19	7
IMAC-SP	35	35	12	7	30	22	13

Table TS5.3 Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected from different TiO_2 formats.

	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
Beads	146	140	68	42	96	96	50
Spin tip	24	23	10	6	18	16	8
Tio tip	29	27	22	5	2	22	7

Table TS5.4 Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected when various acids were used.

Different acids	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤2 Pi	Multi-P ≥ 3Pi
1M CA	11	11	11	3	3	10	1
500m MCA	12	12	11	5	4	4	1
200m MCA	5	5	5	4	3	3	2
100m MCA	11	11	12	3	2	10	1
50m MCA	14	14	10	10	5	12	2
Glycolic acid	18	18	14	12	6	14	4
Lactic acid	18	18	17	9	6	15	3
NQ-Glu	5	5	6	2	2	3	2
NQ-Asp	3	3	3	0	1	3	0
Glu acid	4	4	5	1	0	4	0
Control	2	2	1	1	2	2	0

Table TS5.5 Percentage of acidic amino acids in detected peptides from different acids of beads.

	Percentage of Asp (D)			Percentage of Glu (E)		
	Phospho-peptides	Non-P-peptides	Total peptides	Phospho-peptides	Non-P-peptides	Total peptides
1M CA	7.9	4.9	5.3	7.4	7.3	7.4
500mM CA	4.7	6.3	6.3	6.3	9.5	9.4
200mM CA	3.7	6.1	6.0	0.0	8.7	8.4
100mM CA	3.6	5.9	5.7	6.3	8.3	8.2
50mM CA	5.2	5.2	5.2	7.0	7.1	7.1
Glycolic acid	3.7	5.8	5.7	6.8	8.7	8.6
Lactic acid	4.2	5.6	5.5	6.6	9.2	8.9
Glutamic acid	7.4	6.1	6.3	5.6	9.6	9.5
NQ-Glu	2.5	6.8	6.7	3.8	8.6	8.5
NQ-Asp	5.5	7.4	7.5	12.7	9.2	9.6
Ctrl	5.1	6.5	6.4	5.7	9.1	8.9

Table TS5.6 Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected from different peptide-to-beads ratios.

Peptide-to-TiO_2 ratio	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
Ratio-80	99	92	58	25	59	62	37
Ratio-40	112	104	62	42	59	70	42
Ratio-20	146	140	68	42	96	96	50
Ratio-10	107	99	57	43	53	67	40

Table TS5.7 Number of p- peptides/proteins and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected from consecutive incubations of TiO_2 beads.

Peptide-to-TiO_2 ratio	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
F1	57	57	23	16	39	38	19
F2	47	44	20	9	35	32	14
F3	49	47	25	14	33	30	19
F4	43	40	26	16	23	25	18

Table TS5.8 Number of P-peptides/proteins and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected from different elution of SIMAC experiments.

Peptide-to-TiO_2 ratio	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
IMAC elute	28	26	16	5	13	22	6
TFA-TiO_2	37	36	23	9	30	20	17
TFA-TiO_2-2^{nd} cycle	17	17	10	5	6	13	4
SP-TiO_2	25	23	7	3	6	14	11
SP-TiO_2-2^{nd} cycle	12	13	11	14	14	8	4

Table TS5.9 Number of P-peptides/protein and P-sites on Ser/Thr/Tyr and peptides with mono-/multiple- P-sites detected from IMAC, TiO_2 and SIMAC.

Peptide-to-TiO_2 ratio	P-pep	P-pro	P-Ser	P-Thr	P-Tyr	Mono-P ≤1 Pi	Multi-P ≥ 2Pi
IMAC	26	25	10	7	19	19	7
SIMAC	95	92	51	32	57	61	34
TiO_2	146	140	68	42	96	96	50
Control	27	26	13	11	13	19	8

Table TS5.10 Percentage of triplicate charge peptides in P-peptides and total detected peptides.

	All enrichment	Ratio 20/1	SIMAC	IMAC	Ctrl
P-peptides	48.3	54.1	23.2	57.7	18.5
Total peptides	29.9	30.2	17.6	27.0	14.6

Note: Calculation is based on peptides with charge.

Appendix F

Table

Table TS7.1 Abundance ratios of different comparisons.

Table TS7.1. Ratios of different comparisons.

		Number of metabolites with	
		Ratio <1	Ratio >1
Comparison 1 Tryprone *vs* glucose	P2	24	116
	PBL2025	47	86
	PBL2073	30	110
Comparison 2 Strains grown on glucose	PBL2025/P2	10	122
	PBL2073/PBL2025	121	12
	PBL2073/P2	65	64
Comparison 3 Strains grown on tryptone	PBL2025/P2	35	95
	PBL2073/PBL2025	91	34
	PBL2073/P2	45	96

Supplementary information

Table SP 2.1 PTMs in archaea and human.

Table SP 2.1A PTMs in archaea

Key words Archaea and	methylation	glycosylation	phosphorylation	acetylation	proteolysis
2014	17	5	20	3	16
2013	26	21	34	3	33
2012	27	40	42	5	25
2011	16	19	19	3	9
2010	20	20	12	4	13
2009	12	5	10	4	6
2008	14	10	20	5	6
2007	16	7	17	4	10
2006	17	7	13	6	9
2005	26	13	23	3	5
2004	10	6	18	1	5

Table 2.1 B PTMs in human

Key words Human and	methylation	glycosylation	phosphorylation	acetylation	proteolysis
2014	4369	1529	9870	1209	1718
2013	7462	2812	17127	2085	3219
2012	7543	3144	18074	2135	2722
2011	3187	1381	7601	875	882
2010	2846	1252	7613	888	1419
2009	2391	1232	7225	824	616
2008	2341	1211	6964	732	657
2007	2152	1112	6458	773	597
2006	1838	1060	6109	667	636
2005	1666	1097	5961	596	597
2004	1471	1007	5635	526	648

Other supplementary tables for protein/peptide and P-protein/peptide were listed in the attached disk.

Tables from Chapter 4

Table SP4.1A 1st set iTRAQ Quant protein.

Table SP4.1A 1nd set iTRAQ Quant protein.

Table SP4.1A 3rd set iTRAQ Quant protein.

Quant indicates quantitative proteins with ≥2 unique peptides.

Table SP 4.1 D and E Sum on classification of quant and regulated protein.

Table SP4.2 Changed transporters of *S. solfataricus* grown on different carbon sources.

Table SP4.3A PBL2025 vs P2 both on Glu regulated proteins

Table SP4.3B Try vs Glu of P2 regulated proteins

Table SP4.3C Try vs Glu of PBL2025 regulated proteins

Tables from Chapter 5

Table SP5.1 P-peptides from enrichment techniques using Phenyx

Table SP5.2 P-peptides from enrichment techniques using Mascot Daemon

Table from Chapter 6

Table SP6.1A Quantified P-proteins and their classification based on arCOG functional category.

Table from Chapter 7

Table SP7.1A Metabolites with KEGG C number.